Online Appendix for

Understanding Markets with Socially Responsible Consumers

The Quarterly Journal of Economics

Marc Kaufmann
Peter Andre
Botond Kőszegi

A Background on Competitive Equilibrium

A.1 Graphical Illustration of Competitive Equilibrium

Figure 3 shows how to think about equilibrium determination using modifications of a standard supply-demand diagram, and illustrates the possibility of multiple equilibria. With slight abuse of notation, we write demand and supply curves as functions of the quantity rather than of the price (as we do in the rest of the paper). We can start by drawing the exogenous and linear supply curve $S(q)$. To proceed, we start from the demand curve $D_0(q)$ that would obtain if consumers were selfish ($k = 0$). This is a standard demand curve given by $D_0(q) = u'(q)$. For simplicity in making our points, we assume that this demand curve is two-piece linear and kinked. As indicated on the figure, the intersection of the supply curve and the selfish demand curve gives the unique selfish competitive equilibrium.

Now assume $k > 0$. Based on Definition 1, the equilibrium quantity q and price p must satisfy $u'(q) - p - kq_c = 0$, or $p = u'(q) - ks/(s - 1/u''(q))$. Using that $u'(q) = D_0(q)$ and therefore $u''(q) = D_0'(q)$, we define a "virtual demand curve"

$$D_k(q) = D_0(q) - k \cdot \frac{s}{s - 1/D_0'(q)}.$$

It is easy to see that intersections of this virtual demand curve with the supply curve correspond to competitive equilibria. We refer to $D_k(q)$ as virtual because it cannot be easily observed in its entirety. For instance, a common way of identifying the demand curve, looking at changes in the equilibrium due to exogenous shifts in supply, does not in general trace out any virtual demand curve as the virtual demand curve depends on the supply curve.

Multiple competitive equilibria can occur because even for a downward-sloping selfish demand curve $D_0(q)$, the virtual demand curve $D_k(q)$ can increase. Indeed, for our kinked $D_0(q)$, $D_k(q)$ jumps up. Intuitively, at the kink the price sensitivity of demand and hence dampening jump up, so a consumer's willingness to mitigate jumps down. This results in the two competitive equilibria identified in the figure. Both feature lower consumption than the selfish competitive equilibrium, but the consumption levels are quite different.

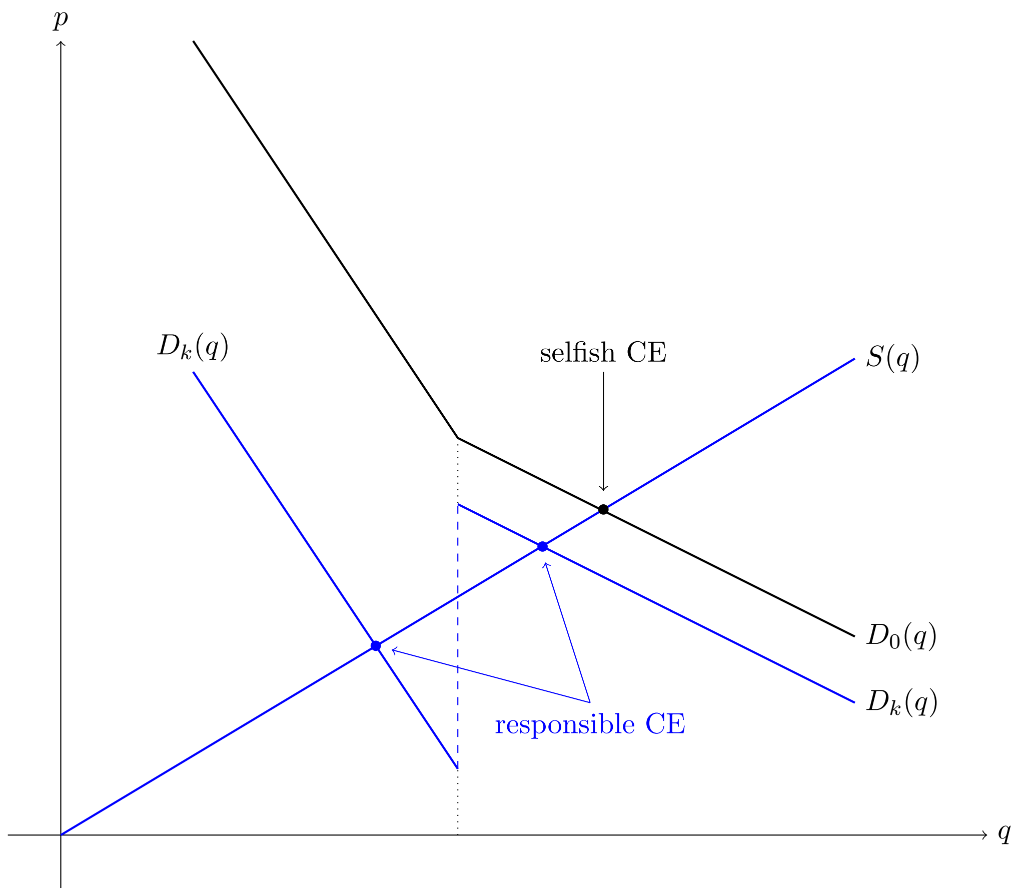


Figure 3: Illustration of Competitive Equilibrium and Multiple Equilibria

A.2 Microfoundation I

We consider an economy with $I+1$ identical consumers and I identical suppliers. Consumer i's utility is

$$U = Ac_i - \frac{1}{2}Bc_i^2 - pc_i - k\sum_j c_j,$$

where c_i denotes consumer i's consumption and $A, B > 0$. The supply of each seller is $S(p) = sp$ with $s > 0$. In addition, there is a shock s_0 to total supply, where s_0 is a non-degenerate random variable whose support includes 0. Hence, total supply is $s_0 + Isp$.

The timing of the game is as follows. First, each consumer submits a (weakly) decreasing demand schedule $c_i(p)$. Then, given all submitted schedules, the market-clearing price is determined. Since supply is linear with a strictly positive slope and demand is weakly decreasing, the market-clearing price exists and is unique. Finally, outcomes are determined and utilities are realized.

We look for linear symmetric Nash equilibrium in which all consumers submit the same linear demand schedule $c_i(p) = a - bp$ with a, $b \geq 0$.

Proposition 9. *There is a unique linear symmetric Nash equilibrium. In this equilibrium, each consumer chooses the schedule $c_i(p) = a - bp$, where*

$$b = \frac{-(IBs - I + 1) + \sqrt{\Delta}}{2IB}, \text{ with } \Delta = (IBs - I + 1)^2 + 4BI^2 s, \text{ and}$$
$$a = \left(A - \frac{ks}{b+s}\right) \cdot \frac{1}{B + \frac{1}{I(b+s)}}.$$

We now use Proposition 9 to provide a microfoundation for Definition 1. Notice that in Proposition 9, the parameters describing the equilibrium, a and b, are functions of I, so that we can write them as $a(I)$ and $b(I)$. We define:

Definition 2. A limiting Kyle equilibrium is a strategy profile in which a consumer chooses $c_i(p) = a_\infty - b_\infty p$, where $a_\infty = \lim_{I\to\infty} a(I)$ and $b_\infty = \lim_{I\to\infty} b(I)$.

We also say that A and B quadratically approximate the utility function $u(\cdot)$ around c if $B = -u''(c)$ and $A - Bc = u'(c)$.

Proposition 10. *The following are equivalent.*

I. The quantity $q^* > 0$, price $p^* > 0$, and consumer price responsiveness $q_p^* \in \mathbb{R}$ are part of a competitive equilibrium.

II. The pair q^*, p^* constitutes the realized outcome in the limiting Kyle equilibrium of the economy in which consumption utility $u(\cdot)$ is quadratically approximated around q^*, and $s_0 = 0$ (i.e., when $s_0 = 0$, the realized quantity and price are q^* and p^*). In this limiting Kyle equilibrium, consumers' price responsiveness equals $c_i'(p) = -b_\infty = q_p^*$.

Proposition 10 says that there is a one-to-one correspondence between competitive equilibria as we defined in the text for an economy with vanishingly small consumers and Nash equilibria of our game in this section with a diverging number of players. To map the economy with a general $u(\cdot)$ to a game with quadratic utility, we quadratically approximate $u(\cdot)$ around the equilibrium consumption level. The one-to-one correspondence means that the quantity, price, and consumer price responsiveness are the same in the competitive equilibrium of the economy and the limiting Nash equilibrium of the game. We do not explicitly include the market responsiveness q_c^* because it is a function of the supply responsiveness s and q_p^*.

A.3 Microfoundation II

We consider the game as in Section A.2 with the following modifications. Consumer i's utility is $u(c_i)$ — as in the text, i.e., not necessarily quadratic. At the same time, a consumer can only submit a linear demand schedule, i.e., she submits $c_i(p) = a_i - b_i p$, where she can specify $a_i \in \mathbb{R}$ and $b_i > 0$. We consider symmetric pure-strategy equilibria when $s_0 = 0$, but within that class impose a kind of robustness requirement with respect to shocks to supply. This is defined in the following way. Let $\bar{s}_0$ be a random variable with mean 0 that is continuously distributed with support $[-1, 1]$. When optimizing with respect to the strategies of other consumers, consumer i assumes that $s_0 = \varepsilon \bar{s}_0$, where $\varepsilon > 0$ and $E(s_0) = 0$. Consumer i's strategy must be the limit of optimal responses as $\varepsilon \to 0$.

Proposition 11. *The pair a, b, price p, and resulting consumption level $q = a - bp$ constitute a robust equilibrium if and only if market clearing is satisfied with q and p,*

$$\begin{aligned} 0 &= u'(q) - p - \frac{1}{I(b+s)} q - k \frac{s}{b+s}, \text{ and} \\ b &= \frac{1}{(-u''(q)) + 1/(I(b+s))}. \end{aligned}$$

Notice that as $I \to \infty$, the conditions of a robust equilibrium approach those of a competitive equilibrium where the price responsiveness of demand is $-b$ and the market responsiveness is $s/(s+b)$. Hence, competitive equilibria are in a sense close to robust equilibria with large I. Our final proposition makes this relationship precise.

Proposition 12. *The following are equivalent.*

I. The quantity $q^* > 0$, price $p^* > 0$, and consumer price responsiveness $q_p^* \in \mathbb{R}$ are part of a competitive equilibrium.

II. There is a sequence of thrice differentiable utility functions u_I and a sequence of robust equilibria $a(I)$, $b(I)$, $p(I)$ for u_I such that as $I \to \infty$, u_I and its derivatives converge uniformly on bounded intervals to u, $p(I) \to p^*$, $[a(I) - b(I)p(I)] \to q^*$, and $[-b(I_n)] \to q_p^*$.

Proposition 12 says that there is a one-to-one correspondence between competitive equilibria and limits of robust equilibria with large I and utility functions close to u.

B Proofs

Proof of Proposition 1. Let us prove that $p(c) \to p^*$ as $I \to \infty$. The rest follows from the derivations in the text. We know that $p(c)$ is determined by $c + ID(p(c)) = IS(p(c))$, while p^* satisfies $D(p^*) = S(p^*)$. We know that $D'(p) < 0$ and $S'(p) > 0$ everywhere, and $\lim_{p\to\infty}(S(p) - D(p)) = \infty$. Now fix some c, $S(p(c)) - D(p(c)) = c/I \to 0$ as $I \to \infty$, hence by continuity $p(c) \to p^*$.

This proves the proposition.

Proof of Observation 1. Define the following:

$$q(p) = S(p)$$
$$q_p(p) = 1/u''(q(p))$$
$$q_c(p) = s/(s - q_p(p))$$

Thus all the equilibrium conditions hold by construction for p, $q(p)$, $q_p(p)$, and $q_c(p)$ except for the consumer's utility maximization. Note that at $p = 0$, we have $u'(0) > k$ (by assumption), so $u'(0) > 0 + kq_c(0)$, since $q_c(p) \leq 1$ for any p. Clearly, as $p \to \infty$, we have that $q(p) \to \infty$, and that $u'(q(p))$ is eventually strictly less than p. Hence for sufficiently large p, we have $u'(q(p)) < p + kq_c(p)$. Therefore by continuity, there is some $p^* > 0$ s.t. $u'(q(p^*)) = p^* + kq_c(p^*)$. Hence $p^* > 0$, $q^* = q(p^*) > 0$, $q_c^* = q_c(p^*) = s/(s - q_p^*) > 0$, $q_p^* = q_p(p^*) = 1/u''(q^*) < 0$ is an equilibrium.

Further note, that the quantity consumed must be strictly less than that under a selfish equilibrium. Suppose, by contradiction, that it is not, and we have a selfish equlbrium with equilibrium price and quantity p_0^* and q_0^*. So we assume $q_0^* \leq q^*$. Then supply must be larger, which requires that prices must be larger, so that $p_0^* \geq p^*$. But then from the first order condition we have $u'(q^*) = p^* + kq_c^* > p^* \geq p_0^* = u'(q_0^*)$, which implies that $q^* < q_0^*$, a contradiction.

This proves the observation.

Proof of Proposition 2. The social welfare is:

$$u(q) - \int_0^q S^{-1}(x)dx - Kq = u(q) - \int_0^q \frac{x}{s}dx - Kq$$

since $S(p) = s \cdot p$. The socially optimal quantity q^{FB} satisfies the following first order condition:

$$u'(q^{FB}) = \frac{q^{FB}}{s} + K$$

On the other hand, from the definition of competitive equilibrium, the consumer's FOC is

$$u'(q^*) = p + kq_c^* = \frac{q^*}{s} + kq_c^*$$

with $q_c^* = s/(s - 1/u''(q^*))$. Since $u''(\cdot) \leq 0$, q_c^* is in the range $[0, 1)$: it is equal to 0 when $s = 0$ (fixed supply), and converges to 1 as s goes to infinity — for perfectly elastic supply, as would be the case when the marginal cost of production is constant.

We now show that $q^* > q^{FB}$. Suppose by contradiction that $q^* \leq q^{FB}$. Then $u'(q^*) \geq u'(q^{FB}) = \frac{q^{FB}}{s} + K \geq \frac{q^*}{s} + K \geq \frac{q^*}{s} + k > \frac{q^*}{s} + kq_c^*$, where the last inequality holds since $q_c^* \in [0, 1)$.

Proof of Observation 2. Choosing the first-best consumption level is obviously a dominant uniquely optimal strategy.

Proof of Proposition 3. Substituting the equilibrium conditions 1, 3, and 4 into condition 2, we obtain:

$$u'(sp) - p = k\frac{su''(sp)}{su''(sp) - 1} \tag{10}$$

Let $g(p) \equiv u'(sp) - p$ and $h(p) \equiv \frac{su''(sp)}{su''(sp)-1}$. Then a necessary and sufficient condition for an (interior) equilibrium is $g(p) = k \cdot h(p)$.

Note the following derivatives:

$$\begin{aligned} h'(p) &= \frac{d}{dp}\left(1 + \frac{1}{su''(p) - 1}\right) = -\frac{s^2u'''(sp)}{(su''(sp) - 1)^2} \\ g'(p) &= su''(sp) - 1. \end{aligned} \tag{11}$$

Then the following Lemma holds:

Lemma 1. *Fixing some $k > 0$ and $s > 0$, with $u'(0) > k$, if there is an equilibrium at $p^* > 0$ s.t. $k \cdot h'(p^*) < g'(p^*)$, then there is (i) an equilibrium at p^+ with $p^+ > p_*$ and (ii) an equilibrium at p^- with $p^- \in (0, p^*)$.*

Proof. Since $u''(c) < 0$, $h(p) > 0$ and $h(p) \in (0,1)$ for all p. Second, $g'(p) = su''(sp) - 1 < -1$ for all p, so that $g(p) \le g(0) - p \to -\infty$ as $p \to \infty$. Hence, $h(p) > g(p)$ and therefore $k \cdot h(p) > g(p)$ for sufficiently large p.

Part i). Take the equilibrium at price p^*, so that $k \cdot h(p^*) = g(p^*)$. Since $k \cdot h'(p) < g'(p)$, we have $k \cdot h(p^* + \varepsilon) < g(p^* + \varepsilon)$ for sufficiently small $\varepsilon > 0$. Since for sufficiently large p, we have $k \cdot h(p) > g(p)$, this implies that there is some $p^+ > p^* + \varepsilon$ s.t. $k \cdot h(p^+) = g(p^+)$ – and hence there is another equilibrium at $p^+ > p^*$.

Part ii). Since $u'(0) > k$, we have that $g(0) = u'(0) > k \ge kh_0$, where $h_0 = \lim_{x \to 0^+} h(x) \in [0, 1]$. Therefore $g(p) > k \cdot h(p)$ for p sufficiently close to 0. Since $k \cdot h'(p) < g'(p)$, we have $k \cdot h(p - \varepsilon) > g(p - \varepsilon)$ for sufficiently small $\varepsilon > 0$. So by continuity there is some $p^- \in (0, p^*)$ s.t. $k \cdot h(p^-) = g(p^-)$, and there is an equilibrium at p^-. $\square$

Now suppose that we have a given k (with $k > 0$) and s with a competitive equilibrium (q^*, p^*, q_p^*, q_c^*). Since this is an equilibrium, it must satisfy equation 10, which puts constraints on $u'(q^*)$ and $u''(q^*)$ but leaves $u(\cdot)$ otherwise unconstrained.

Next let us note that if we can show that $u'''(q^*)$ is sufficiently large implies that $k \cdot h'(p^*) < g'(p^*)$, then we are done. In that case, by Lemma 1 we know that there must be a strictly higher equilibrium at price $p^+ > p^*$ and a strictly lower equilibrium at $p^- < p^*$. At the strictly higher equilibrium, we have that $q^+ = sp^+ > sp^* = q^*$. This implies that the left-hand side of the equilibrium condition (10) is strictly lower in the equilibrium with price p^+ than the equilibrium with price p^*, and so is the right-hand side. Therefore $su''(q^+)/(su''(q^+-1) < su''(q^*)/(su''(q^*)-1)$, which implies that $|u''(q^*)| > |u''(q^+)|$. Thus $|q_p^+| = 1/|u''(q^+)| > 1/|u''(q^*)| = |q_p^*|$, and $q_c^+ = \frac{s}{s-q_p^+} = \frac{s}{s+|q_p^+|} < \frac{s}{s+|q_p^*|} = \frac{s}{s-q_p^*} = q_c^-$. The inequalities for the case $p^- < p^*$ are obtained similarly.

Therefore, we now show that that when $u'''(p^*)$ is sufficiently large, then $k \cdot h'(p^*) < g'(p^*)$. Using equations 11, we have:

$$k \cdot h'(p) < g'(p) \iff k\left(\frac{s^2 u'''(sp)}{(su''(sp)-1)^2}\right) < su''(sp) - 1 \iff u'''(sp) > -\frac{(su''(sp)-1)^3}{ks^2}$$

which completes the proof of Proposition 3.

Proofs of Propositions 4 and 5. First, we give the formal definition of the equilibrium with policy:

Definition 3. A competitive equilibrium with a permit-supply policy consists of a quantity $q^* > 0$, consumer price $p^* > 0$, permit fee $\tau^* > 0$, consumer price responsiveness $q_p^* \in \mathbb{R}$, and market responsiveness $q_c^* \in \mathbb{R}$ that satisfy the following conditions:

1. a. Supply equals q^*: $q^* = S(p^* - \tau^*)$.

   b. The market for permits clears: $\pi S(p^* - \tau^*) - (1-\pi)\tau^* + \pi_0 = 0$.

2. Demand equals q^*: $u'(q^*) = p^* + k \cdot q_c^*$.

3. Market responsiveness is consistent with the responsiveness of consumers and net supply: $q_c^* = s_{\text{net}}/(s_{\text{net}} - q_p^*)$, where $s_{\text{net}} \equiv \frac{(1-\pi)s}{(1-\pi)+\pi s}$ based on Equation (7).

4. Consumer price responsiveness is consistent with optimization: $q_p^* = 1/u''(q^*)$.

For a general policy, including our permit-supply policies, we can define an equilibrium with policy by p^*, q^*, q_c^*, q_p^*, τ^* that satisfy the equations in Definition 3, but subject to different policy constraints. Let us label the two policy types that we want to compare by π' and π.

To show that policy of type π is better than the policy of type π', for every equilibrium under policy of type π' satisfying $u'(0) > k + \tau^*$, we need to find an equilibrium under policy of type π with the same τ^* and strictly lower pollution. Consider an equilibrium under policy of type π', which satisfies

$$
\begin{aligned}
q^* &= S(p^* - \tau^*) \\
u'(q^*) &= p^* + kq_c^* \\
q_c^* &= \frac{s_{\text{net}}(\pi')}{s_{\text{net}}(\pi') - q_p^*} \\
q_p^* &= \frac{1}{u''(q^*)}
\end{aligned}
$$

where $s_{\text{net}}(\pi') = \left.\frac{dS_{\text{net}}(p-\tau^*)}{dp}\right|_{p=p^*}$, where the derivative is evaluated under policy π'. Let us define the following:

$$
\begin{aligned}
p(q) &= S^{-1}(q) + \tau^* \\
q_c(q) &= \frac{s_{\text{net}}(\pi)}{s_{\text{net}}(\pi) - q_p(q)} \\
q_p(q) &= \frac{1}{u''(q)}
\end{aligned}
$$

By construction, q, $p(q)$, $q_c(q)$, $q_p(q)$, and τ^* satisfy all the conditions for equilibrium except utility maximization and the policy constraint.

Lemma 2. *Suppose that $s_{net}(\pi) > s_{net}(\pi')$ holds for every τ^* and p^*. Then there are values q^-, p^-, q_c^-, and q_p^- that, together with τ^* satisfy $q^- < q^*$ and the equilibrium conditions, with the possible exception of the policy constraint.*

Proof. Since $s_{\text{net}}(\pi) > s_{\text{net}}(\pi')$, we have $q_c^* = \frac{s_{\text{net}}(\pi')}{s_{\text{net}}(\pi')-q_p^*} < \frac{s_{\text{net}}(\pi)}{s_{\text{net}}(\pi)-q_p^*} = q_c(q^*)$ and $p^* = p(q^*)$, so

$$u'(q^*) = p^* + kq_c^* = p(q^*) + kq_c^* < p(q^*) + kq_c(q^*).$$

Since $p(0) = \tau^*$, $q_c(0) \leq 1$, and $u'(0) > \tau^* + k$ all hold, we also have

$$u'(0) > \tau^* + k = p(0) + k \geq p(0) + kq_c(0).$$

By continuity of $u'(q)$, $p(q)$, and $q_c(q)$, there exists a $q \in (0, q^*)$ with

$$u'(q) = p(q) + kq_c(q)$$

so that q, $p(q)$, $q_c(q)$, $q_p(q)$, and τ^* satisfy the equilibrium conditions with $q < q^*$, with the possible exception of the policy constraint. □

To show that policies of type π are better than those of type π', it is enough to show that $s_{\text{net}}(\pi) > s_{\text{net}}(\pi')$ and that the policy constraint holds.

For policies satisfying equation (6), when $\pi' > \pi$, then we showed that $s_{\text{net}}(\pi) = \frac{(1-\pi)s}{(1-\pi)+\pi s}$, hence $s_{\text{net}}(\pi) > s_{\text{net}}(\pi')$. Since for every π we can pick π_0 such that the policy constraint holds, this proves Proposition 4.

Similarly for Proposition 5, we have that $\tau^* = \tau_0 + \tau_1 p^*$. Fixing τ^*, for any p^* and τ_1, we can pick τ_0 such that $\tau^* = \tau_0 + \tau_1 p^*$ such that the policy constraint holds. And since $s_{\text{net}}(\tau_1) = \frac{d}{dp}S(p - \tau_0 - \tau_1 p) = s \cdot (1 - \tau_1)$ is strictly decreasing in τ_1 we have $s_{\text{net}}(\tau_1) < s_{\text{net}}(0)$ when $\tau_1 < 0$; and $s_{\text{net}}(\tau_1) > s_{\text{net}}(0)$ when $\tau_1 > 0$. This proves the Proposition, showing that decreasing taxes are better than fixed ones which are better than increasing ones. This proves Proposition 5.

Proof of Observation 3. Obvious.

Proof of Proposition 6. We first define the competitive equilibrium with policy when there is trade. To adapt Definition 3 for this situation, we distinguish between the market responsivenesses of home- and foreign-supplied quantities, q_c^h and q_c^f. These can be calculated similarly to q_c in the

Definitions 1 and 3. Denoting the price responsiveness of demand by q_p, they are

$$\begin{array}{lcc} & q_c^h & q_c^f \\ \text{tax} & \dfrac{s^h}{s^h + s^f - q_p} & \dfrac{s^f}{s^h + s^f - q_p} \\ \text{cap} & 0 & \dfrac{s^f}{s^f - q_p} \end{array} \tag{12}$$

A consumer's effect on the externality is then $q_c^h e^h + q_c^f e^f$, and this is what she takes into account when choosing her consumption. This leads to the following definition of competitive equilibrium with trade:

Definition 4. A competitive equilibrium with trade and permit-supply policy consists of home and foreign quantities $q^{h*}, q^{f*} > 0$, consumer price $p^* > 0$, tax τ^*, a consumer price responsiveness $q_p^* \in \mathbb{R}$, and home and foreign market responsivenesses $q_c^{h*}, q_c^{f*} \in \mathbb{R}$ such that

1. a. $q^{h*} = S^h(p^* - \tau^*)$ and $q^{f*} = S^f(p^*)$.

   b. $\pi S^h(p^* - \tau^*) - (1 - \pi)\tau^* + \pi_0 = 0$.

2. $u'(q^{h*} + q^{f*}) = p^* + k \cdot (e^h q_c^{h*} + e^f q_c^{f*})$.

3. The responsivenesses q_c^{h*} and q_c^{f*} are given by Equation (12).

4. $q_p^* = 1/u''(q^{h*} + q^{f*})$.

From the producer side, Definition 4 can be seen as a mixture between Definitions 1 and 3, where the former applies to the foreign market and the latter to the home market. Hence, foreign suppliers receive the consumer price p^*, while domestic suppliers receive only $p^* - \tau^*$. In addition, the market for permits at home must clear. From the consumer side, equilibrium accounts for the fact that supply comes from two sources. Hence, consumers' total consumption equals total production, and a consumer takes into account her impact on both sources of supply.

We can now prove Proposition 6. Consider a policy-maker who compares a tax and a cap. First we show that, for every τ^* that satisfies $u'(S^f(\tau^*)) > ke^f + \tau^*$, a unique equilibrium exists for both a tax and a cap that leads to such a τ^*; then we compare these equilibria. Let $q^f(p) = S^f(p)$, $q^h(p) = S^h(p - \tau^*)$. Note that due to the quadratic utility, the price responsiveness is constant and equal to $1/u_{cc}$, while the consumer responsiveness in the home and foreign market depend on the type of policy, but are constant with respect to the price p.

At a price $p = \tau^*$, where $q^h(p) = 0$, we have $u'(q^f(p) + q^h(p)) = u'(q^f(\tau^*)) = u'(S^f(\tau^*)) > \tau^* + ke^f = p + ke^f \geq p + k(e^h q_c^{h*} + e^f q_c^{f*})$. Letting p increase also increases the quantity provided and decreases the marginal utility, until a point where that quantity satiates the consumer: $u'(q^f(p) + q^h(p)) = 0$. Denote this price by $\bar{p}$. At the same time, $p + k(e^h q_c^{h*} + e^f q_c^{f*})$ strictly increases, since all the terms except the first are constant in p. So there is a unique $p^* \in (\tau^*, \bar{p})$ such that the $u'(q^f(p^* + q^h(p^*) = p^* + k(e^h q_c^{h*} + e^f q_c^{f*})$ holds. At this price, all other equilibrium conditions hold, so we have a unique equilibrium for each policy with strictly positive supply in the home and foreign market.

By market clearing, we have $q^{h*} = S^h(p^* - \tau^*)$ and $q^{f*} = S^f(p^*)$ for both tax and cap. We have that q^{h*} and q^{f*}, and hence $q^* = q^{h*} + q^{f*}$ and $g^* = e^h q^{h*} + e^f q^{f*}$, are all strictly increasing in p^*. Since τ^* is the same for both cap and tax, the total pollution is therefore lower for the policy that generates the lower p^* and q^*.

From the FOC for utility maximization we have $u'(q^*) - p^* = k(q_c^h e^h + q_c^f e^f)$. A cap is strictly better than a tax if and only if $q^*_{\text{cap}} < q^*_{\text{tax}}$ and $p^*_{\text{cap}} < p^*_{\text{tax}}$, which from the FOC is equivalent to $q^h_{c,\text{cap}} e^h + q^f_{c,\text{cap}} e^f > q^h_{c,\text{tax}} e^h + q^f_{c,\text{tax}} e_f$. Plugging in the values from equation (12), we get that a cap is strictly better than a tax iff:

$$\begin{aligned}
& e^h q^h_{c,\text{tax}} + e^f q^f_{c,\text{tax}} - e^f q^f_{c,\text{cap}} < 0 \\
\iff & \frac{e^h s^h + e^f s^f}{s^h + s^f - q_p} < \frac{e^f s^f}{s^f - q_p} \\
\iff & (e^h s^h + e^f s^f)(s^f - q_p) < e^f s^f (s^h + s^f - q_p) \\
\iff & - q_p e^h s^h + e^h s^h s^f < e^f s^f s^h \\
\iff & - q_p < s^f \frac{e^f - e^h}{e^h} \\
\iff & - \frac{1}{u_{cc}} < s^f \frac{e^f - e^h}{e^h}
\end{aligned}$$

and a tax is strictly better if the same inequality holds strictly in the opposite direction.

This proves the proposition.

Notation and Setup for Propositions 7. Let p^c and p^d be the consumer prices for the clean and dirty product respectively. With some abuse of notation, we denote by p the difference between the consumer prices of the two products: $p = p^c - p^d$. We now show that for any such relative price p, there exist unique consumer prices $p^c(p)$ and $p^d(p)$ such that $p = p^c(p) - p^d(p)$, and market clearing holds on the supply side: $S^c(p^c(p)) + S^d(p^d(p)) = 1$.

Fix p, and consider p^d and p^c s.t. $p^c - p^d = p$. As $p^d \to -\infty$, we also have $p^c \to -\infty$ and thus $S^c(p^c) + S^d(p^d) \to -\infty$. Similarly, when $p^d \to \infty$, then $p^c \to \infty$ and hence $S^c(p^c) + S^d(p^d) \to \infty$. Next, as p^d increases, p^c strictly increases. Moreover, when $S^c(p^c) + S^d(p^d) > 0$, either $S^c(p^c)$ or $S^d(p^d)$ is strictly increasing in p^d. Therefore, by continuity, there is a unique p^d solving $S^c(p^c) + S^d(p^d) = 1$, so that $p^c(p)$ and $p^d(p)$ are well-defined.

Market clearing implies that $S^c(p^c) + S^d(p^d) = 1$, so that $s^c p^c + s^d p^d = 1$. Since $p^c - p^d = p$, we have $p^c = p + p^d$, so that $s^c(p + p^d) + s^d p^d = 1$. Taking derivatives with respect to p, we get that $s^c(1 + p^{d\prime}) + s^d p^{d\prime} = 0$, so that $p^{d\prime} = -\frac{s^c}{s^c + s^d}$.

Again abusing notation somewhat, we define $S(p)$ as the net *relative* supply of the clean product relative to the dirty product that is consistent with market clearing and p: $S(p) \equiv S^c(p^c(p)) -$

$S^d(p^d(p))$.

For ease of exposition, we drop the explicit dependence of $p^c(p)$ and $p^d(p)$ on p and write p^c and p^d.

We now derive $S(p)$. Note that $S(p)$ is continuous in p, since the consumer and production prices are continuous in p, and the supply functions are continuous in the consumer prices. Moreover, p^c and p^d are differentiable in p, hence so is $S(p)$.

We know that $p^{d\prime}(p) = -\frac{s^c}{s^c+s^d}$. Moreover, $S(p) - 1 = S^c(p^c) - S^d(p^d) - 1 = S^c(p^c) - S^d(p^d) - (S^c(p^c) + S^d(p^d)) = -2S^d(p^d)$, so $S'(p) = -2s^d p^d$, i.e.

$$S'(p) = 2\frac{s^d s^c}{s^c + s^d} \equiv s. \tag{13}$$

Hence we have that $S(p)$ is linear with slope s.

Definition of Competitive Equilibrium. Since each consumer has unit demand, aggregate demand is uniquely determined by knowing which consumers buy the clean good and which buy the dirty good. Without loss of generality, we describe demand by an indifferent consumer $k^* \in (\underline{k}, \overline{k})$ such that every consumer with $k > k^*$ buys the clean good and every consumer with $k < k^*$ buys the dirty good. To describe equilibrium, we follow the steps outlined in the text.

In Step (a), we introduce the market responsiveness dQ, which denotes by how much the total quantity of a good increases when a consumer switches to it.

In Step (b), we identify an optimality condition for demand given dQ. Since the consumer with k^* is indifferent between the products, the gain in money from switching from clean to dirty must equal the costs from increased externalities. The monetary gain is the equilibrium price difference $p^* = p^c - p^d$, while the cost is $k^*(e^d - e^c)dQ$. So k^* must satisfy

$$p^* = k^*(e^d - e^c)dQ. \tag{14}$$

In Step (c), we solve for dQ in terms of price responsiveness of demand and supply. We denote by $\mathrm{d}k^* \in [0, \infty]$ the price responsiveness of the cutoff consumer k^*. Note that $\mathrm{d}k^* = \infty$ captures the possibility that consumers respond perfectly elastically to price changes. This is relevant because the products are perfect substitutes.

Applying the formula for the quantity effect from Equation (2), we have that $dQ = S'(p^*)/(S'(p^*) - D'(p^*))$, where $S(\cdot)$ is the relative supply of the clean good and $D(\cdot)$ its relative demand. We have that $S'(p^*) = s$, and we know that $D(p^*) = 1 - 2F(k^*)$, so that $D'(p^*) = 2f(k^*)dk^*$. Putting these together, we get the market responsiveness:

$$dQ = \frac{s}{s + 2f(k^*)dk^*}.$$

Finally, in Step (d), we solve for the price responsiveness of demand captured in dk^*. For this, we totally differentiate the indifference equation (14) with respect to p^*, assuming (analogously to Definition 1) that over the infinitesimal range in question, the market responsiveness $\mathrm{d}Q$ is constant. This gives $\mathrm{d}k^* = 1/((e^d - e^c)\mathrm{d}Q)$.

We can now put the above together with the market clearing condition to define competitive equilibrium. To do so, let F denote the distribution function corresponding to f. We define a competitive equilibrium with substitutes as a relative price $p^* \geq 0$, an indifferent consumer $k^* \in (\underline{k}, \overline{k})$, and a consumer price responsiveness $\mathrm{d}k^* \in [0, \infty]$ such that:

1. The market clears: $S(p^*) = 1 - 2F(k^*)$.
2. The consumer $k = k^*$ is indifferent between the two products: $p^* = k^*(e^d - e^c)\mathrm{d}Q$, with

$$\mathrm{d}Q = \frac{s}{s + 2f(k^*)\mathrm{d}k^*}.$$

3. Consumer price responsiveness is consistent with consumer k^*'s indifference:

$$\mathrm{d}k^* = \frac{1}{(e^d - e^c)\mathrm{d}Q}.$$

Lemmas and Preliminary results for Proposition 7. First, we characterize equilibria with $dk^* = \infty$.

Lemma 3. *Suppose that* $S(0) \in (-1, 1)$. *Then:*

i) there is an equilibrium with $dk^* = \infty$;

ii) in any equilibrium with $dk^* = \infty$, *we have* $p^* = 0$, $dQ = 0$, *and* $S(0) = 1 - 2F(k^*)$, *and* $k^* \in (\underline{k}, \overline{k})$.

Proof. First, suppose $S(0) \in (-1, 1)$. Then let us show that $dk^* = \infty$, $p^* = 0$, and k^* s.t. $1 - 2F(k^*) = S(0) \in (-1, 1)$ is an equilibrium. Since $f(k) > 0$ for all $k \in (\underline{k}, \overline{k})$, $k^* \in (\underline{k}, \overline{k})$ is uniquely determined. Market clearing holds, since $S(p^*) = S(0) = 1 - 2F(k^*)$. We have that $f(k^*) > 0$ and $dQ = \frac{S'(p^*)}{S'(p^*)+2f(k^*)dk^*}$, so that $dQ = 0$. Hence $p^* = k^*(e^d - e^c)dQ$ holds, since $p^* = 0$ and $dQ = 0$. This proves part i).

Suppose that we have an equilibrium with $dk^* = \infty$. Then by consumer price responsiveness, we have $dk^* = \frac{1}{(e^d - e^c)dQ}$, hence $dQ = 0$. For the indifferent consumer, we have $p^* = k^*(e^d - e^c)dQ$, so $p^* = 0$. By market clearing, we must have $S(p^*) = S(0) = 1 - 2F(k^*) \in (-1, 1)$, which also pins down $k^* \in (\underline{k}, \overline{k})$. This proves part ii). $\square$

Next we look at equilibria that satisfy $dk^* < \infty$.

Lemma 4. *Let* $\underline{f} \equiv \inf_{[\underline{k},\overline{k}]} f(k)$ *and* $\overline{f} \equiv \sup_{[\underline{k},\overline{k}]} f(k)$. *Let* $\overline{p} = 1/s^c$.

i) When $s(e^d - e^c) < 2\underline{f}$, *there is no equilibrium with* $k^* \in (\underline{k}, \overline{k})$ *and* $dk^* < \infty$.

ii) When $s(e^d - e^c) > 2\overline{f}$, *and* $\overline{p} > \underline{k}(e^d - e^c)$, *there is at least one equilibrium with* $k^* \in (\underline{k}, \overline{k})$ *and* $dk^* < \infty$. *In such an equilibrium,* $p^* > 0$.

iii) For any equilibrium with $k^* \in (\underline{k}, \overline{k})$ *and* $dk^* < \infty$, *we have that* $dQ \leq dQ(\underline{f}) \equiv \frac{s(e^d - e^c) - 2\underline{f}}{s(e^d - e^c)} < 1$.

iv) If $\overline{p}s < \underline{k}s(e^d - e^c) - 2f(\underline{k})$, *then there is no equilibrium with* $dk^* < \infty$ *and* $k^* \in (\underline{k}, \overline{k})$.

Proof. Since we are looking for equilibria with $k^* \in (\underline{k}, \overline{k})$, there has to be some supply of each product and hence $f(k^*) > 0$. Combining the equilibrium conditions 2 and 3, we get that:

$$\frac{1}{(e^d - e^c)\mathrm{d}k^*} = dQ = \frac{s}{s + 2f(k^*)\mathrm{d}k^*} \tag{15}$$

Multiplying out and rearranging (which relies on $dk^* < \infty$), we obtain:

$$(s \cdot (e^d - e^c) - 2f(k^*))dk^* = s \tag{16}$$

If $s \cdot (e^d - e^c) < 2\underline{f}$, then the left-hand side of (16) is negative, while the right-hand side is strictly positive, so the inequality cannot hold. This proves part i).

We can rearrange equation (16) to obtain

$$\frac{1}{dk^*} = \frac{1}{s}\Big(s(e^d - e^c) - 2f(k^*)\Big) \tag{17}$$

From condition 3 of the equilibrium, we know that $(e^d - e^c)dQ = 1/dk^*$, so we can replace $(e^d - e^c)dQ$ in $p^* = k^*(e^d - e^c)dQ$ to obtain $p^* = k^*/dk^*$, hence $1/dk^* = p^*/k^*$ which we can plug into equation (17):

$$p^* = \frac{k^*}{s}\Big(s \cdot (e^d - e^c) - 2f(k^*)\Big) \tag{18}$$

Notice that at price $\overline{p}$, the whole market is served by the clean good, since the prices for clean and dirty of $p^c = 1/s^c$ and $p^d = 0$ satisfy market clearing and $p^c - p^d = \overline{p}$. Hence $S(\overline{p}) = 1$, and for $p \leq \overline{p}$, we have $S(p) = S(\overline{p}) - s(\overline{p} - p) = 1 - s(\overline{p} - p)$, since we showed that $S'(p) = s$.

By market clearing we have $S(p^*) = 1-2F(k^*)$, so $1-s(\overline{p}-p^*) = 1-2F(k^*)$, i.e. $\overline{p}s - 2F(k^*) = sp^*$, where we can replace p^* by its value from equation (18):

$$\overline{p}s - 2F(k^*) = k^*(s(e^d - e^c) - 2f(k^*)) \tag{19}$$

Claim: $\overline{p} > \underline{k}(e^d - e^c)$ and $s(e^d - e^c) > 2\overline{f}$ implies that equation (19) has at least one solution with $k^* \in (\underline{k}, \overline{k})$.

Proof of claim: Let $l(k^*) \equiv \overline{p}s - 2F(k^*)$ and $r(k^*) \equiv k^*(s(e^d - e^c) - 2f(k^*))$. Then $l(0) = \overline{p}s > 0$ and $l(\overline{k}) = \overline{p}s - 2$. Since $\overline{p}s = \frac{s}{s^c} = \frac{2s^d}{s^c + s^d} < 2$, we have $l(\overline{k}) = \overline{p}s - 2 < 0$. Next, $r(0) = 0$ and since $s(e^d - e^c) > 2\overline{f} \geq 2f(k^*)$, we have $r(k) > 0$ for $k > 0$.

Now we can show that the curves of $r(k)$ and $l(k)$ must cross at least once in $(\underline{k}, \overline{k})$. Since $\overline{p} > \underline{k}(e^d - e^c)$, we have that $l(0) = \overline{p}s > k^*(e^d - e^c)s \geq r(k^*)$ for all $k^* \leq \underline{k}$. When $k^* < \underline{k}$,

$F(k^*) = 0$ so $l(k^*) = \overline{p}s - F(k^*) = \overline{p}s > r(k^*)$, so the two curves do not cross for any $k^* < \underline{k}$ (nor do they cross at $\underline{k}$, where $r(k^*)$ jumps discontinuously down). So $r(\underline{k}) < l(\underline{k})$. Therefore $r(\cdot)$ and $l(\cdot)$ cross at least once for some $k^* \in (\underline{k}, \overline{k})$, since $l(\cdot)$ is continuous and strictly decreasing in that range and ends below 0, while $r(\cdot)$ is continuous, and ends up strictly above 0.

This proves the claim.

Note that in the above argument, if we have $l(\underline{k}) < r(\underline{k})$, then the curve of r lies above that of l for all $k \geq \underline{k}$ and we have no such competitive equilibrium with $k \in (\underline{k}, \overline{k})$, but instead have one with $k < \underline{k}$ if we appropriately define dQ outside of the range $(\underline{k}, \overline{k})$. This proves condition iv).

Plugging the value of k^* where the curves cross into equations (18) and (17) yields equilibrium values for p^* and dk^* that are strictly positive, since $s(e^d - e^c) > 2\overline{f} \geq 2f(k^*)$. This proves part ii).

To prove part iii), start with equation (17) which must hold for an arbitrary interior equilibrium:

$$\frac{1}{dk^*} < \frac{s(e^d - e^c) - 2\underline{f}}{s} \tag{20}$$

Since $dQ = \frac{1}{(e^d - e^c)dk^*}$, this implies that

$$dQ = \frac{1}{(e^d - e^c)dk^*} \leq \frac{s(e^d - e^c) - 2\underline{f}}{s(e^d - e^c)} = dQ(\underline{f}) < 1 \tag{21}$$

$\square$

Proof of Proposition 7. For Part I, we have $S(0) \in (-1, 1)$, since some of both products sell when they go at the same price (since s^d, $s^c > 0$). Thus by Lemma 3, there is an equilibrium with $p^* = 0$, so $p^{c*} = p^{d*}$, and in this equilibrium $dQ = 0$, so consumers are indifferent between the two products.

For part II, we know by Lemma 4 that there is no equilibrium with $dk^* < \infty$ and $k^* \in (\underline{k}, \overline{k})$ if $s(e^d - e^c) < 2\underline{f}$ (which is strictly greater than 0), where $s = \frac{2s^c s^d}{s^c + s^d}$. So there is no other equilibrium in which both products are consumed if $e^d - e^c$ or s^c are sufficiently small.

For part III, by Lemma 4, there is an equilibrium with $k^* \in (\underline{k}, \overline{k})$, $dk^* < \infty$, and $p^* > 0$ if

$\underline{k}(e^d - e^c) < \overline{p}$ and $s(e^d - e^c) > 2\overline{f}$. Since $\underline{k} = 0$, the first condition always holds. Hence when $e^d - e^c$ is sufficiently large, we have an equilibrium of this kind. Since $dk^* < \infty$, we have $dQ > 0$, so consumers with $k > k^*$ strictly prefer the clean product, and since the price $p^* > 0$, this means that $p^{c*} > p^{d*}$.

We prove Part IV directly. Since in any equilibrium, everyone consumes one unit and the two goods are perfect consumption substitutes, consumption utility is the same in any equilibrium, so we can ignore consumption utility when comparing the welfare of equilibria. Social welfare is thus given by:

$$W(q) = -\int_0^q (S^c)^{-1}(x) - (S^d)^{-1}(1-x)dx + Kq(e^d - e^c) - Ke^d \tag{22}$$

where q is the amount of the clean product, so that $1 - q$ is the amount of the dirty product. So $W'(q) = -\left((S^c)^{-1}(q) - (S^d)^{-1}(1-q)\right) + K(e^d - e^c) = -(p^c(q) - p^d(q)) + K(e^d - e^c) = -p + K(e^d - e^c)$, which is strictly positive as long as $p(q) < K(e^d - e^c)$, and maximized at quantity q^{FB} satisfying $p(q^{FB}) = K(e^d - e^c)$.

Suppose we have two equilibria with k^* taking the values $k^*_H < k^*_L$, with $k^*_i \in (\underline{k}, \overline{k})$, so that the consumption of the clean good, and therefore its price premium, is higher for the equilibrium with k^*_H than the one with k^*_L — thus $p^*_H > p^*_L$.

Then the high equilibrium must be a non-selfish equilibrium with $dQ \in (0, 1)$, so that $p^*_H = k^*_H(e^d - e^c)dQ$. Therefore $p^*_H < k^*_H(e^d - e^c) < \overline{k}(e^d - e^c) \leq K(e^d - e^c)$, and hence social welfare is still increasing in the quantity of the clean product, therefore the equilibrium with higher consumption in the clean good has higher social welfare. This proves part IV and thus completes the proof of Proposition 7.

Proof of Proposition 8. The clean good is in fixed supply S^c, the dirty good has perfectly elastic supply at price P^d, and a share $\alpha \in [0, 1]$ of consumers are naive, which means that they perceive the externality from the clean and dirty good to be 0 and 1, respectively.

First notice that because $u'(S^c) > P^d + k$, even consumers who think that the externality from consuming the clean good is 1 will buy more than the clean good alone can provide. Therefore any equilibrium will feature some of both goods.

We look for equilibria in which rational consumers assume that substitution dampening is full: thus the impact in terms of externality is the same whether they buy a unit of the clean or the dirty good. Hence they always buy the cheapest good. They take into account quantity dampening, thinking that for each unit of consumption (no matter which product), they cause an increase of the externality of exactly e^d, which we will show later is equal to 1. The naive consumers on the other hand believe that the clean good has no externality and so strictly prefer it as long as the price premium compared to the dirty good is strictly below k.

This means that the price p of the clean good must be at least P^d. If not, then both the naive and the rational strictly prefer it and do not buy any of the dirty good, but this violates market clearing, since we showed that they consume more than the available supply of the clean good.

Similarly, we must have $p \leq P^d + k$: if $p > P^d + k$, not just rational but also naive consumers strictly prefer the dirty good, which means that no one would buy the clean good. This violates market clearing, given the fixed supply of the clean good.

To prove part I, let $\bar{c}$ be s.t. $u'(\bar{c}) = P^d$, i.e. the amount a naive person consumes of the clean good if it sells at price P^d. Let $\underline{\alpha}$ be s.t. $\underline{\alpha}\bar{c} = S^c$. Then for $\alpha < \underline{\alpha}$, in any equilibrium we must have that $p = P^d$. Suppose not, so that $p > P^d$. Then only the naive consumers prefer buying the clean good, and each of them buys strictly less than $\bar{c}$. But since $\alpha\bar{c} < S^c$, this means that demand is strictly less than supply, thus markets do not clear. We have a contradiction and thus must have $p = P^d$.

The fact that prices are constant and equal to P^d in any equilibrium also means that when one person buys one more unit, they cause the full externality from that unit, which means that $e^d = 1$. Therefore rational consumers consume $\underline{c}$ given by $u'(\underline{c}) = P^d + k$, whereas naive consumers consume $\bar{c}$ given by $u'(\bar{c}) = P^d$.

Social welfare is given by $W \equiv \alpha u(\bar{c}) + (1-\alpha)u(\underline{c}) - P^d q - Kq$, where $q = \alpha\bar{c} + (1-\alpha)\underline{c} - S^c$ is the amount produced that causes externalities (i.e. net of the supply of the clean good). Then social welfare changes with α as follows:

$$
\begin{aligned}
\frac{dW}{d\alpha} &= u(\bar{c}) - u(\underline{c}) - (P^d + K)\frac{dq}{d\alpha} \\
&= \int_{\underline{c}}^{\bar{c}} u'(x)dx - (P^d + K)(\bar{c} - \underline{c}) \\
&< u'(\underline{c})(\bar{c} - \underline{c}) - (P^d + K)(\bar{c} - \underline{c}) \\
&\leq (\bar{c} - \underline{c})(u'(\underline{c}) - (P^d + K)) \\
&= 0
\end{aligned}
$$

where the first inequality holds because u is strictly concave, and the last line holds since $u'(\underline{c}) = P^d + k \leq P^d + K$. Thus social welfare strictly decreases in α.

This proves the first part.

Now suppose that $\alpha > \underline{\alpha}$. This means that at the price of $p = P^d$, the naive consumers all want to buy $\bar{c}$ of the clean good, which exceeds its supply. Therefore any equilibrium will have to feature $p > P^d$. This means that rational consumers will strictly prefer the dirty good, which they perceive as causing the same externality at strictly lower monetary cost.

Let $\overline{\alpha}$ be s.t. $\overline{\alpha}\underline{c} = S^c$. Now consider $\alpha \in (\underline{\alpha}, \overline{\alpha})$. Then let $c(\alpha) = S^c/\alpha$, which is the amount each naive can consume for that α, based on market clearing and the fact that naive consumers all buy the clean good. We have that $c(\alpha) = S^c/\alpha > S^c/\overline{\alpha} = \underline{c}$. Then p must satisfy $u'(c(\alpha)) = p$: if it is larger, then the naive consumers consume too little, if it is smaller naive consumers consume too much; in both cases the market for the clean good doesn't clear. Note that since $c(\alpha) > \underline{c}$, we have that $p = u'(c(\alpha)) < u'(\underline{c}) = P^d + k$, thus the naive consumers strictly prefer to buy the clean good. Hence this is an equilibrium.

That this leads to a welfare improvement can be shown formally by computing $dW/d\alpha$, or as follows. Consider increasing α from α_L to $\alpha_H \in (\alpha_L, \overline{\alpha})$. Let us consider three groups of consumers: the naive consumers (those that are initially naive), the rational consumers (those rational before and after), and the switchers (those initially rational, later naive). The welfare depends only on the consumption of these three groups and total consumption.

Then note that the change from α_L to α_H can be considered as the naive consumers reducing their consumption from $c(\alpha_L)$ to $c(\alpha_H) < c(\alpha_H)$. Each unit that they reduce their consumption goes either towards a net reduction of total consumption or towards allowing a switcher to consume more. The marginal cost to the naive consumers is at most $u'(c(\alpha_H))$, while the marginal benefit to a switcher is at least $u'(c(\alpha_H))$, since they used to consume strictly less. And the marginal social benefit of reducing total consumption is $P^d + K$, but we know that $P^d + K = u'(\underline{c}) > u'(c(\alpha_H))$, hence this change leads to a strict net benefit.

This proves part II.

Finally, consider $\alpha > \overline{\alpha}$. Then the price of the clean good must be $P^d + k$. We showed already that it cannot be larger. If it was smaller, then all the naive consumers would strictly prefer the clean good, which would lead to overdemand. When $p = P^d + k$, the naive consumers are indifferent between the two goods, while the rational consumers still prefer the dirty good. The naive consumers now choose consumption to solve $u'(c) = p = P^d + k$, which is the same as the rational consumers, hence everyone consumes $\underline{c}$. Thus social welfare is the same as when $\alpha = 0$.

This proves part III, and thereby Proposition 8.

Proof of Proposition 9. Let us solve for consumer i's best response $c_i(p)$ (a schedule) given given other consumers $j \neq i$ are adopting schedule $c_j(p) = a - bp$. Let us solve for the hypothetical best response if the consumer knew the actual ex post realization s_0. It will turn out that, while the consumer does not know it ex ante, in equilibrium there is a one-to-one mapping between the equilibrium price p and s_0. Thus, since the consumer can condition their consumption on p via their schedule, they effectively can condition their consumption on s_0, hence they can achieve the outcome as if they knew s_0.

For now, suppose the consumer knows what s_0 will be. Then if they submit $c_i(p)$, they know that this will determine via market clearing their own consumption and the equilibrium price:

$$c_i(p(s_0)) + I(a - bp(s_0)) = s_0 + Isp(s_0) \implies c_i(p(s_0)) = s_0 + Ip(s_0)(s + b) - Ia \qquad (23)$$

Since we assume that the consumer knows what s_0 will be, they would obtain exactly the same outcome if they submitted a constant schedule equal to $c_i(p(s_0))$ everywhere. This leads to exactly the same consumption for consumer i, the same equilibrium price $p(s_0)$ and hence the same total consumption. Therefore it has the same utility for consumer i. By conditioning on s_0, we can therefore reframe the game as one in which the consumer submits their consumption level x_i.

Let us define $x_i \equiv c_i(p(s_0))$. Then the utility from submitting x_i is

$$U(x_i) = Ax_i - \frac{1}{2}Bx_i^2 - p(s_0)x_i - k\left(x_i + \sum_{j\neq i} c_j\right) \tag{24}$$

Taking derivatives with respect to x_i of equation (23) (noting that $c_i(p(s_0)) = x_i$) and rearranging, we get the price impact of x_i:

$$\frac{dp(s_0)}{dx_i} = \frac{1}{I(b+s)}. \tag{25}$$

Next, we can take the derivative of total consumption with respect to x_i, and use the price impact of x_i to obtain the quantity impact of x_i:

$$\frac{d}{dx_i}\left(x_i + \sum_{j\neq i} c_j\right) = 1 - \frac{d}{dx_i}I(a-bp) = 1 - \frac{Ib}{I(b+s)} = \frac{s}{b+s}.$$

Finally, using these impacts, we can compute the first order condition for the utility maximization:

$$\begin{aligned} A - Bx_i - p(s_0) - x_i\frac{1}{I(b+s)} - k\frac{s}{b+s} &= 0 \\ \Longrightarrow x_i = \left(A - \frac{ks}{b+s}\right)\cdot\frac{1}{B+\frac{1}{I(b+s)}} &- \frac{1}{B+\frac{1}{I(b+s)}}p(s_0) \end{aligned}$$

Thus we have $x_i = \alpha - \beta p(s_0)$ for some constants α and β, with $\beta > 0$ that do not depend on s_0. This provides the best possible utility if the consumer could know s_0 ex ante. It is easy to check — by plugging the value of $x_i = \alpha - \beta p(s_0)$ into the market clearing condition — that $p(s_0)$ is a strictly decreasing linear function of s_0. So $p(s_0)$ has a strictly decreasing (and linear) inverse, which we can write $s_0(p)$.

Now we can show that the consumer can achieve this outcome with the linear schedule $c_i(p) = \alpha - \beta p$. When the consumer ends up paying the equilibrium price p, then it must be the case that the realized s_0 is equal to $s_0(p)$. The realized s_0 must satisfy the following two conditions:

$$c_i(p) = s_0 + Ip(s+b) - Ia$$
$$c_i(p) = \alpha - \beta p$$

Since both equations are linear in p, this uniquely determines s_0 as a function of p, and we know that $s_0(p)$ satisfies both of these equations, so the realized s_0 equals $s_0(p)$. Hence this schedule achieves the same utility as a schedule when the consumer knows the ex post realization of s_0, where the consumer can maximize state by state, hence this schedule is the optimal schedule.

Imposing symmetry, the coefficient on p must equal the slope of other consumers' schedules, b:

$$b = \frac{1}{B + \frac{1}{I(b+s)}} \iff IBb^2 + (IBs - I + 1)b - Is = 0.$$

Notice that one root is positive and one root is negative, hence the positive one is chosen.

$$b = \frac{-(IBs - I + 1) + \sqrt{\Delta}}{2IB}, \text{ where } \Delta = (IBs - I + 1)^2 + 4BI^2 s.$$

This completes the proof.

Proof of Proposition 10. We prove $I \Rightarrow II$ and $II \Rightarrow I$ separately.

Step 1: $I \Rightarrow II$. Assume that the tuple p^*, q^*, q_c^*, q_p^* constitutes a competitive equilibrium.

Since p^*, q^* are part of a competitive equilibrium, condition #2 implies $u'(q^*) = p^* + k \cdot q_c^*$. We quadratically approximate consumer utility around q^*, which implies $B = -u''(q^*)$ and $A - Bq^* = u'(q^*) = p^* + k \cdot q_c^*$.

Condition #3 and condition #4 imply that $q_c^* = \frac{s}{s - q_p^*} = \frac{s}{s - \frac{1}{u''(q^*)}}$, so that $A - Bq^* = p^* + k \cdot \frac{s}{\frac{1}{B} + s}$. This is equivalent to

$$q^* = (A - \frac{ks}{\frac{1}{B} + s})\frac{1}{B} - \frac{1}{B}p^*.$$

From proposition 9, we know that optimal demand schedule for these values of A and B is:

$$\begin{aligned}
c_i(p) &= a(I) - b(I)p, \text{ where} \\
a(I) &= A - \frac{ks}{b(I)+s} \cdot \frac{1}{B + \frac{1}{I(b(I)+s)}} \\
b(I) &= \frac{-(IBs - I + 1) + \sqrt{\Delta}}{2IB} p \\
\Delta &= (IBs - I + 1)^2 + 4BI^2 s
\end{aligned}$$

Computing limits as $I \to \infty$, we find that $b(I) \to b_\infty = \frac{1}{B}$, and hence $a(I) \to \left(A - \frac{ks}{\frac{1}{B}+s}\right) \cdot \frac{1}{B}$. Therefore in the limiting Kyle equilibrium the schedule is given by

$$c_i(p) = \left(A - \frac{ks}{\frac{1}{B}+s}\right)\frac{1}{B} - \frac{1}{B}p. \tag{26}$$

Hence the pair (p^*, q^*) is on the demand schedule for the limiting Kyle schedule, i.e. $q^* = c_i(p^*)$. Finally, since market clearing holds in the competitive equilibrium, we have $q^* = S(p^*)$, hence $c_i(p^*) = S(p^*)$, which is the market clearing condition of the limiting Kyle equilibrium when the shock $s_0 = 0$ is realized.

In this limiting Kyle equilibrium, the consumers' price responsiveness is $c_i'(p) = -\frac{1}{B} = \frac{1}{u''(q^*)}$ by the quadratic approximation.

Step 2: $II \Rightarrow I$**.** Suppose the pair (p^*, q^*) constitutes the realized outcome in the limiting Kyle equilibrium of the economy in which consumer utility is quadratically approximated around q^*, and when $s_0 = 0$. Due to the quadratic approximation, we know that $u'(q^*) = A - Bc_i(p^*)$ and $B = -u''(q^*)$. Moreover, the consumers' price responsiveness is $c_i'(p^*) = -b_\infty = -\frac{1}{B} = \frac{1}{u''(q^*)}$. Using these equalities, we check that the pair (p^*, q^*) satisfies condition #1 – #4 one by one.

1. condition #1: the market clearing for the Kyle equilibrium with realization $s_0 = 0$ yields $c_i(p^*) = S(p^*) + s_0 = S(p^*)$, so market clearing in the competitive equilibrium, $q^* = S(p^*)$, holds with $q^* = c_i(p^*)$.

2. condition #2: Using the quadratic approximations and the schedule of the limiting Kyle equilibrium from equation (26), we get

$$u'(q^*) = A - Bc_i(p^*) = A - (A - \frac{ks}{\frac{1}{B} + s}) + p = p - \frac{ks}{s + \frac{1}{B}} = p - \frac{ks}{s - \frac{1}{u''(q^*)}} = p - k\frac{s}{s - q_q^*}$$

3. condition #4: the consumers' price responsiveness q_p^* is $c_i'(p) = b_\infty = -\frac{1}{B} = \frac{1}{u''(q^*)}$.

4. condition #3: note that setting $q_c^* = \frac{s}{s - q_p^*}$, this is consistent with condition #2 and #4.

This completes the proof.

Proof of Proposition 11. We consider consumer i's strategic situation when all other consumers are choosing the pair a, b. Given all consumers' strategies, the realized price is a function of s_0, which we denote by $p(s_0)$. As a reminder, $s_0 = \varepsilon \bar{s}_0$, where $\bar{s}_0$ is continuously distributed with support $[-1, 1]$ and with mean 0. Similarly, consumer i's consumption is a function $c_i(s_0)$. With some abuse of notation, we redefine a_i as consumer i's consumption level when $s_0 = 0$. (We cannot, and do not, at the same time redefine the a in other consumers' strategies.) This means that $c_i(s_0) = a_i - b_i \Delta p(s_0)$, where $\Delta p(s_0) = p(s_0) - p(0)$. Market clearing has the following implications:

$$p(0) = \frac{a_i + Ia}{I(b+s)}$$
$$\Delta p(s_0) = \frac{-s_0}{b_i + I(b+s)}$$

Furthermore, the total quantity in the market is $s_0 + Isp(s_0) = s_0 + Is(p(0) + \Delta p(s_0))$. For future reference,

$$-b_i \Delta p(s_0) = \frac{b_i s_0}{b_i + I(b+s)} = s_0 - \frac{s_0 I(b+s)}{b_i + I(b+s)}$$

We only consider $a_i, b_i \geq 0$. Define $A(a_i) \equiv u'(a_i) - \frac{2a_i + Ia}{I(b+s)} - k\frac{s}{b+s} = u'(a_i) - p(0) - \frac{1}{I(b+s)} a_i - k\frac{s}{b+s}$. Let $\bar{a}$ be the unique solution to $A(a_i) = 0$:

$$u'(\bar{a}) - \frac{2\bar{a} + Ia}{I(b+s)} - k\frac{s}{b+s} = 0 \tag{27}$$

This is unique since $A(a_i)$ is strictly decreasing at a rate of more than $2/(I(b+s))$, with $\bar{a} > 0$. Note that $A(a_i) < 0$ if $a_i > \bar{a}$ and $A(a_i) > 0$ if $a_i < \bar{a}$.

We want to show that the consumer's expected utility is maximized in some compact and strictly positive region $O = [a_0, a_1] \times [b_0, b_1]$, where a_0, $b_0 > 0$ and a_1, $b_1 < \infty$. For this, we replace the consumption utility $u(\cdot)$ by $w(\cdot)$, where $w(x) = u(x)$ for all $x \geq \delta$. But for $x < \delta$, we set $w''(x) = \max\{u''(\delta), u''(x)\}$, so that together with $w(\delta) = u(\delta)$, $w'(\delta) = u'(\delta)$, and $u''(\delta) = w''(\delta)$ we have a well-defined function w, s.t. w is twice continuously differentiable, and since $w''(x) \geq u''(x)$ (alongside the boundary conditions), we have that $w(x) \geq u(x)$.[33] Let EU and EW denote the expected utility under u and w respectively.

Lemma 5. *Let $\delta < \frac{1}{2}a_0$ and $\varepsilon < \frac{1}{2}a_0$, then if the expected utility with w achieves any global maximum in $O = [a_0, a_1] \times [b_0, b_1]$ with a_0, $b_0 > 0$ and a_1, $b_1 < \infty$, then the expected utility under u also achieves any global maximum in O.*

Proof. Suppose that w achieves its maximum on O. Since the optimal $(a_i, b_i) \in O$, we have that the utility is given by

$$EW = E_{s_0}[w(c_i(s_0)) - p(s_0)c_i(s_0) - kIsp(s_0)]$$

We have that $c_i(s_0) > \delta$ if and only if $a_i + \frac{b_i s_0}{b_i + I(b+s)} > \delta$, that is if and only if

$$s_0 > -(a_i - \delta)(\frac{b_i + I(b+s)}{b_i}) \iff |s_0| < |a_i - \delta|(\frac{b_i + I(b+s)}{b_i})$$

The right-hand side is strictly larger than $|a_i - \delta|$, which is strictly larger than $(a_0 - \delta) > \frac{1}{2}a_0$, while $|s_0| \leq \varepsilon$. Thus a sufficient condition for $c_i(s_0) > \delta$ for any schedule chosen from O is that $\varepsilon < \frac{1}{2}a_0$.

Therefore $EW = EU$ for any $(a_i, b_i) \in O$. Hence if some $(a_i, b_i) \in O$, maximizes EW, then it clearly maximizes also EU. Suppose this is not the case, so we have some $(a_i, b_i) \in O$ leading to schedule $c_i(\cdot)$ maximizing EW and some $(a_j, b_j) \notin O$ leading to schedule $c_j(\cdot)$ maximizing EU. But since $w(c) \geq u(c)$, we have $EW(c_j) \geq EU(c_j)$, which since it maximizes EU implies $EU(c_j) > EU(c_i)$, yet this latter equals $EW(c_i)$ since we just showed that these schedules have the

[33] When $\lim_{x \to x_0} u''(x) \to -\infty$ for some $x_0 \leq 0$ — such as for $u(x) = \log(x)$ — we treat $u''(x) = -\infty$ and thus $w''(x) = u''(\delta)$ for all $x \leq x_0$.

same expected utility. But then $EW(c_j) > EW(c_i)$, contradicting that c_i maximized EW. $\square$

Anticipating that we will use Lemma 5 later on, we now only consider utility functions that are twice differentiable with bounded first- and second-order derivatives. We will show that these have their global maxima inside of some region O, thus we can apply the Lemma to w defined above to show that the original utility function (which might have unbounded derivatives) also achieves its global maxima in that same region. Therefore the utility is well-defined also for negative consumption, which can in principle occur for some consumption schedules.

For steps 1 through 4, we write u for such a utility function with bounded derivatives.

For the remaining steps, assume that $\delta < \frac{1}{3}\bar{a}$.

Step 1: For any b_i, the a_i that maximizes EU lies in $[\bar{a}-\delta, \bar{a}+\delta]$ for $\varepsilon < \delta C$ for some constant C, with δ and C independent of b_i. The consumer's expected utility is

$$
\begin{aligned}
&E_{s_0}\left[u(c_i(s_0)) - p(s_0)c_i(s_0) - kIsp(s_0)\right] \\
=&E_{s_0}\left[u(a_i - b_i\Delta p(s_0)) - (p(0)+\Delta p(s_0))(a_i - b_i\Delta p(s_0)) - kIs(p(0)+\Delta p(s_0))\right],
\end{aligned}
$$

where E_{s_0} denotes expectation taken over the distribution of s_0.

The derivative of the consumer's utility with respect to a_i is

$$
\frac{\partial EU}{\partial a_i} = E_{s_0}\left[u'(c_i(s_0)) - (p(0)+\Delta p(s_0)) - \frac{1}{I(b+s)}c_i(s_0) - k\frac{s}{b+s}\right].
$$

Applying the intermediate value theorem, there is some $\delta(s_0) \in (0, s_0)$ s.t. this equals

$$
\begin{aligned}
&u'(c_i(0)) - p(0) - \frac{1}{I(b+s)}c_i(0) - k\frac{s}{b+s} + \frac{1}{b_i + I(s+b)}E_{s_0}\left[s_0\left(u''\Big(c_i\big(\delta(s_0)\big)\Big)b_i + 1 - \frac{b_i}{I(b+s)}\right)\right] \\
=&u'(a_i) - p(0) - \frac{a_i}{I(b+s)} - k\frac{s}{b+s} + \frac{1}{b_i + I(s+b)}E_{s_0}\left[s_0\left(u''\Big(c_i\big(\delta(s_0)\big)\Big)b_i + 1 - \frac{b_i}{I(b+s)}\right)\right]
\end{aligned}
\tag{28}
$$

Let $K = \sup_{x \geq 0} |u''(x)|$. Then:

$$
\begin{aligned}
|\frac{\partial EU(a_i, b_i)}{\partial a_i} - A(a_i)| &= \frac{1}{b_i + I(s+b)} |E_{s_0}\left[s_0\left(u''\left(c_i\big(\delta(s_0)\big)\right)b_i + 1 - \frac{b_i}{I(b+s)}\right)\right]| \\
&\leq \frac{1}{b_i + I(s+b)} E_{s_0}[|s_0|](Kb_i + 1 + \frac{b_i}{I(b+s)}) \\
&\leq \frac{b_i}{b_i + I(s+b)} E_{s_0}[|s_0|](K + \frac{1}{I(b+s)}) + \frac{1}{b_i + I(s+b)} E_{s_0}[|s_0|] \\
&\leq \varepsilon(K + \frac{2}{I(s+b)})
\end{aligned} \tag{29}
$$

Now, remembering that $A(a_i)$ is decreasing with slope at least $2/(I(b+s))$, we have that $A(\bar{a}-\delta) \geq A(\bar{a}) + 2/(I(b+s))\delta = 2\delta/(I(b+s))$, and similarly $A(\bar{a}+\delta) \leq -2\delta/(I(b+s))$. Now pick ε s.t.

$$
\varepsilon(K + 1 + \frac{1}{I(b+s)}) < \frac{2\delta}{I(b+s)}.
$$

Then the derivative of EU with respect to a_i is strictly positive for all $a_i < \bar{a} - \delta$, and it is strictly negative for all $a_i > \bar{a} + \delta$. Hence EU is maximized for some $a_i \in [\bar{a} - \delta, \bar{a} + \delta]$.

Step 2: Let $\varepsilon < \frac{1}{2}\bar{a} - \delta$. For any $a_i \in [\bar{a} - \delta, \bar{a} + \delta]$, the b_i that maximizes EU is achieved for some $b_i \leq \overline{B}$ independent of δ, ε, or a_i. Fix some $a_i \in I_A = [\bar{a} - \delta, \bar{a} + \delta]$. Consider only $\varepsilon < \frac{1}{2}\bar{a} - \delta$. Then $c_i(s_0)$ is strictly bounded away from 0:

$$
c_i(s_0) = a_i + \frac{b_i s_0}{b_i + I(b+s)} > \bar{a} - \delta - \varepsilon > \frac{1}{2}\bar{a}
$$

From a symmetric argument, we get an upper bound at $3/2\bar{a}$. Therefore $\bar{u}'' \equiv \max_{[\frac{1}{2}\bar{a}, \frac{3}{2}\bar{a}]} u''(x)$ and $\underline{u}'' \equiv \min_{[\frac{1}{2}\bar{a}, \frac{3}{2}\bar{a}]} u''(x)$ are both well-defined strictly negative numbers. They satisfy $\underline{u}'' \leq u''(c_i(s_0)) \leq \overline{u}''$ for all consumption levels that are possible given $a_i \in I_A$ and $\varepsilon < \frac{1}{2}\bar{a} - \delta$.

Now define $\overline{v}(c) \equiv u(a_i) + u'(a_i)(c - a_i) + \overline{u}''\frac{(c-a_i)^2}{2}$. Note that $\overline{v}(a_i) = u(a_i)$, $\overline{v}'(a_i) = u'(a_i)$, and $\overline{v}''(c) = \overline{u}'' \geq u''(c)$, therefore:

$$
\overline{v}(c) - u(c) = \int_{a_i}^{c} (\overline{v}'(x) - u'(x))dx = \int_{a_i}^{c}\int_{a_i}^{x} (\overline{v}''(y) - u''(y))dydx \geq 0
$$

and thus $\overline{v}(c) \geq u(c)$, with $\overline{v}(a_i) = u(a_i)$.

We will next show that there is some $\overline{B}$ s.t. $\overline{v}(0) > \overline{v}(\overline{B})$. The consumer's expected utility $EV(b_i)$ when their consumption utility is $\overline{v}$ is:

$$E_{s_0}[\overline{v}(c_i(s_0)) - p(s_0)c_i(s_0) - kIsp(s_0)]$$

We expand this and, using $E_{s_0}[s_0] = 0$, all the terms linear in s_0 drop out, yielding:

$$\begin{aligned} EV(b_i) =& E_{s_0}\left[\overline{v}(c_i(s_0))\right] - p(0)c_i(0) + E_{s_0}[s_0^2]\frac{b_i}{(b_i + I(s+b))^2} - kIsp(0) \\ =& u(a_i) + \frac{\overline{u}''}{2}\frac{b_i^2}{(b_i + I(s+b))^2}E_{s_0}[s_0^2] - p(0)c_i(0) + E_{s_0}[s_0^2]\frac{b_i}{(b_i + I(s+b))^2} - kIsp(0) \\ =& u(a_i) + \left(\frac{\overline{u}''}{2}b_i + 1\right)\frac{b_i}{(b_i + I(s+b))^2}E_{s_0}[s_0^2] - p(0)c_i(0) - kIsp(0) \end{aligned}$$

Then, since the only terms depending on b_i are those multiplied by $E_{s_0}[s_0^2]$, $EV(b_i) - EV(0)$ equals

$$\left(\frac{\overline{u}''}{2}b_i + 1\right)\frac{b_i}{(b_i + I(s+b))^2}E_{s_0}[s_0^2]$$

Thus $EV(0) > EV(b_i)$ when $\frac{\overline{u}''}{2}b_i + 1 < 0$, or when $b_i > \frac{2}{-\overline{u}''}$. Thus letting $\overline{B} = \frac{2}{-\overline{u}''}$ is such that the global maximizer b_i for EV at this a_i has to satisfy $b_i \leq \overline{B}$. Since $EU(0) = EV(0) > EV(b_i) \geq EU(b_i)$, this shows that the maximum for EU (for a given a_i) is also achieved for $b_i \leq \overline{B}$. Since this bound is independent of $a_i \in I_A$, we have shown that for every $(a_i, b_i) \notin I_A \times [0, \overline{B}]$, there is some point in this region that achieves strictly higher expected utility.

Step 3: Let $\varepsilon < \frac{1}{2}\bar{a} - \delta$. For any $a_i \in [\bar{a} - \delta, \bar{a} + \delta]$, the b_i that maximizes EU is achieved for some $b_i \geq \underline{B}$ independent of δ, ε, or a_i. We now repeat the same argument as above, but this time bounding $u(\cdot)$ via $\underline{v}(c) \equiv u(a_i) + u'(a_i)(c - a_i) + \underline{u}''\frac{1}{2}(c - a_i)^2$. So now $\underline{v}''(c) \leq u''(c)$ for any consumption level that is possible for $a_i \in I_A$. We then find that $EV(b_i) - EV(0)$ is given by

$$\left(\frac{\underline{u}''}{2}b_i + 1\right)\frac{b_i}{(b_i + I(s+b))^2}E_{s_0}[s_0^2]$$

We want to show that the person prefers some strictly positive b_i, i.e. $EV(b_i) - EV(0) > 0$. This holds for every $b_i \in (0, \frac{2}{-\underline{u}''})$, with equality at the corners. Therefore EV achieves a maximum on the interior of $[0, \frac{2}{-\underline{u}''}]$. Denote this maximizer by M, which is independent of a_i and ε, so that $EV(M) > 0$. Notice that $\partial EU/\partial b_i \leq E_{s_0}[u'(c_i(s_0))s_0]\frac{I(b+s)}{(b_i+I(b+s))^2}$, since only the consumption utility term contributes positively (see equation (30) ahead). Since $c_i(s_0)$ is bounded, so is $u'(c_i(s_0))$, so there is some $\overline{u}' = \max_{c=c_i(s_0)} u'(c)$, which holds for all $a_i \in I_A$. So there is some $\lambda < \infty$ such that $\partial EU/\partial b_i \leq \lambda$. Since $EU(M) \geq EV(M) > EV(0) = EU(0)$, there is some minimal $x > 0$ s.t. $EU(x) = EV(M)$, and this x is at least $(EV(M) - EV(0))/\lambda \equiv \underline{B}$ given the function cannot increase too quickly. Hence for every $b_i < \underline{B}$, we have $EU(b_i) < EV(M) \leq EU(M)$, so there is some point $(a_i, b_i) \in I_A \times [\underline{B}, \overline{B}]$ that is larger.

Step 4: global maximum determined by FOCs. By steps 2 and 3, we have shown that EU has a global maximum, achieved on $I_A \times [\underline{B}, \overline{B}]$ for every ε sufficiently small. By continuity and differentiability of the functions involved, this global maximum satisfies the first order conditions.

Since w as defined in the beginning has bounded derivatives, this implies that it achieves its maximum in a region $O \in I_A \times [\underline{B}, \overline{B}]$, where $\overline{B}$ and $\underline{B}$ are both strictly positive and finite and independent of ε. Since EW achieves any global maximum on $O = I_A \times [\underline{B}, \overline{B}]$, by Lemma 5, so does EU. Moreover, any such maximum is characterized by the first order conditions, since u is continuous and differentiable on O

Step 5: taking limits From now on, having applied Lemma 5, we again write u for the original utility function, knowing that any maximum of EU is achieved inside of O. Since the global maximum is characterized by the first order conditions on the compact set $I_A \times [\underline{B}, \overline{B}]$, the best response in the limit is a limit of the best responses and inside this same region. Moreover, by continuity and differentiability, this limit best response satisfies the first order conditions in the limit, which we now compute.

As $\varepsilon \to 0$, we have $c_i(s_0) \to c_i(0) = a_i$, so we see from equation (28) that the first order condition converges to

$$u'(a_i) - p(0) - \frac{a_i}{I(b+s)} - k\frac{s}{b+s} = 0$$

which is the same as equation (27), so that $a_i \to \bar{a}$ as $\varepsilon \to 0$. (Notice that even if there are multiple solutions to the first order condition for $\varepsilon > 0$, they all converge to the same limit.)

Next let us compute the derivative of the consumer's utility with respect to b_i, when a_i satisfies the first order condition at ε:

$$E_{s_0}\left[u'(c_i(s_0))\frac{s_0 I(b+s)}{(b_i+I(b+s))^2} - \frac{s_0}{(b_i+I(b+s))^2}c_i(s_0) - p(s_0)\frac{s_0 I(b+s)}{(b_i+I(b+s))^2} - kIs\frac{s_0}{(b_i+I(b+s))^2}\right]. \tag{30}$$

Multiplying by $(b_i + I(b+s))^2/(I(b+s))$ gives

$$E_{s_0}\left[s_0\left(u'(c_i(s_0)) - \frac{1}{I(b+s)}c_i(s_0) - p(s_0) - k\frac{s}{b+s}\right)\right].$$

Subtracting the left-hand side of (27), which is zero, inside the expectation yields

$$E_{s_0}\left[s_0\left(\left(u'(c_i(s_0)) - u'(\bar{a})\right) - \frac{1}{I(b+s)}(-b_i\Delta p(s_0)) - \Delta p(s_0)\right)\right] + E_{s_0}\left[2s_0\frac{\bar{a}-a_i}{I(b+s)}\right]. \tag{31}$$

Let us focus on the first term in this expression and rewrite it. By the intermediate value theorem, there is some $\delta_1(s_0)$ lying between 0 and s_0 such that the first term in the previous equation equals

$$E_{s_0}\left[s_0\left(u''(c_i(\delta_1(s_0)))c_i'(\delta_1(s_0))s_0 - \frac{1}{I(b+s)}(-b_i\Delta p(s_0)) - \Delta p(s_0)\right)\right].$$

Using $c_i'(s) = \frac{b_i}{b_i+I(b+s)}$, so that $s_0 c_i'(s) = -b_i\Delta p(s_0)$, we can factor out $\Delta p(s_0)$ and substituting for it:

$$E_{s_0}\left[\frac{-s_0^2}{b_i+I(b+s)}\left(u''(c_i(\delta_1(s_0))(-b_i) - \frac{1}{I(b+s)}(-b_i) - 1\right)\right].$$

or $-\epsilon^2$ times

$$E_{\bar{s}_0}\left[\frac{\bar{s}_0^2}{b_i+I(b+s)}\left(u''(c_i(\delta(\epsilon\bar{s}_0)))(-b_i) - \frac{1}{I(b+s)}(-b_i) - 1\right)\right].$$

Note that $-\varepsilon^2$ times the expression in equation (31) equals zero at the optimum. Moreover, when $\epsilon \to 0$, we have $\delta_1(\epsilon\bar{s}_0) \to 0$ and that $a_i \to \bar{a}$. So for the consumer's first order condition with respect to b_i to be satisfied as $\varepsilon \to 0$, the whole expression must converge to 0. The second term converges to 0 since $a_i \to \bar{a}$, thus the first term must converge to 0 too. For this to hold, since the term inside the expectation is a product of a strictly positive number and a number that converges to a constant, this constant must be 0, yielding

$$b_i = \frac{1}{(-u''(a_i)) + 1/(I(b+s))}.$$

Notice that given (a, b), the solution of (a_i, b_i) that satisfies the limit first order conditions is unique. Since there is a global maximum satisfying the first order conditions, the unique solution to these first order conditions must be a local, and hence global, maximum.

This shows that we have a unique best response for given (a, b) characterized by these first order conditions. Finally we impose symmetry by requiring $a_i = a - bp(0)$ (since we transformed a_i to be the consumption amount consumed when the supply shock $s_0 = 0$, i.e. for the price $p(0)$) and $b_i = b$. Let us write q for the equilibrium consumption, which is linked to the equilibrium price p once via the schedules chosen by consumers, $q = a - bp$, and once via market clearing for shocks $s_0 = 0$, i.e. the market clearing condition for $p(0)$, $pI(b + s) = (I + 1)a$. We thus obtain that any linear symmetric Nash equilibrium is characterized by $q = a - bp$, market clearing, and the following first order conditions:

$$0 = u'(q) - p - \frac{q}{I(b+s)} - k\frac{s}{b+s}$$
$$b = \frac{1}{-u''(q) + \frac{1}{I(b+s)}}$$

This proves the proposition.

Proof of Proposition 12.

I $\Longrightarrow$ II Suppose we have a competitive equilibrium p^*, q^*, and q_p^*, which satisfy the following:

$$q^* = sp^*$$
$$u'(q^*) = p^* + k\frac{s}{s - q_p^*}$$
$$q_p^* = \frac{1}{u''(q^*)}$$

Then define $b(I) = -q_p^*$, and $p(I) = p^*$ independent of I. Further define the following

$$\begin{aligned}
q_I &\equiv \frac{I}{I+1} q^* \\
\lambda_I &\equiv \frac{u'(q^*)}{u'(q_I)} \\
\mu_I &\equiv \frac{u''(q^*) + \frac{1}{I(b(I)+s)}}{u''(q_I)} \\
u_I(x) &\equiv \mu_I u(x) + \left((\lambda_I - \mu_I) u'(q_I) + \frac{q_I}{I(b(I)+s)} \right) (x - q_I)
\end{aligned}$$

Then we can check that $a(I) = q_I + b(I)p(I)$, $b(I)$, $p(I)$ is a robust equilibrium. First, market clearing holds, since $(I+1)(a(I) - b(I)p(I)) = (I+1)q_I = Iq^*$, which by market clearing in the competitive equilibrium equals Isp^*. Thus market clearing holds.

Next, $u_I'(q_I) = (\mu_I + \lambda_I - \mu_I)u'(q_I) + \frac{q_I}{I(b(I)+s)} = \lambda_I u'(q_I) + \frac{q_I}{I(b(I)+s)} = u'(q^*) + \frac{q_I}{I(b(I)+s)}$. Using the first order condition of the competitive equilibrium, we can replace $u'(q^*)$ by $p^* + k\frac{s}{s-q_p^*}$, q_p^* by $-b(I)$, and p^* by $p(I)$:

$$u_I'(q_I) = p(I) + k\frac{s}{s+b(I)} + \frac{q_I}{I(b(I)+s)} \implies 0 = u_I'(q_I) - p(I) - \frac{q_I}{I(b(I)+s)} - k\frac{s}{s+b(I)}$$

which is exactly the first of the first order conditions.

Finally, we have that $u_I''(q_I) = \mu_I u''(q_I) = u''(q^*) + \frac{1}{I(b(I)+s)}$. Since $-b(I) = q_p^* = 1/u''(q^*)$, this implies:

$$u_I''(q_I) = -\frac{1}{b(I)} + \frac{1}{I(b(I)+s)} \implies b(I) = \frac{1}{-u_I''(q_I) + \frac{1}{I(b(I)+s)}}$$

which is the second of the first order conditions. Thus we have a robust equilibrium. Moreover, as $I \to \infty$, we have that $u_I(x) \to u(x)$ uniformly on any bounded interval, $p(I) = p^* \to p^*$, $a(I) - b(I)p(I) = q_I \to q^*$, and $-b(I) = q_p^* \to q_p^*$ as required.

II $\implies$ I Suppose that we have a sequence of robust equilibria as stated. By Proposition 11, we know that market clearing holds for each of the equilibria as well as the following equations:

$$q(I) = a(I) - b(I)p(I)$$
$$(I+1)q(I) = Isp(I)$$
$$0 = u_I'(q(I)) - p(I) - \frac{q(I)}{I(b(I)+s)} - k\frac{s}{b(I)+s}$$
$$b(I) = \frac{1}{-u_I''(q(I)) + \frac{1}{I(b(I)+s)}}$$

where $u_I \to u$ uniformly on bounded intervals. We have that $p(I) \to p^*$ and $-b(I) \to q_q^*$, and $q(I) \to q^*$. Hence market clearing holds, as $q^* = sp^*$ as $I \to \infty$. Similarly, by uniform convergence, we obtain the following limit first order conditions:

$$0 = u'(q^*) - p^* - \frac{s}{s - q_p^*}$$
$$q_p^* = \frac{1}{u''(q(I))}$$

which proves that we have a competitive equilibrium after defining $q_c^* = s/(s - q_p^*)$.

This proves the proposition. $\square$

C Additional material for the empirical study

C.1 Sample

Sampling We recruited respondents in October 2023 using the online survey company Prolific. We recruited respondents from different parts of the Prolific respondent pool in order to approximate the general US population in terms of gender, age, income, and region.

Final Sample Characteristics Table C.1 presents demographic summary statistics for our final sample and compares them to the demographic characteristics of the US adult population.

Exclusion Criteria All exclusion criteria are preregistered. The sample does not contain the following responses: incomplete responses, responses at both extreme 1% tails in the response duration, and duplicate respondents (very rare cases).

Attention Screener Only participants who pass an attention screener at the beginning of the survey can proceed to the main part of the survey.

Attrition A total of 2,358 respondents start the survey. One respondent does not confirm the consent form, and 93 respondents do not access the survey from a desktop computer, fail the attention screener, or do not complete the demographic questions. Hence, 2,264 respondents reach the main part of the survey. Of those, 2,092 (92%) complete the questions on beliefs about dampening, 2,043 (90%) complete the questions on the role of consequences, 2,042 (90%) complete the full survey, and 2,000 respondents (88%) are included in the final survey (see exclusion criteria). Conditional on reaching the main part of the survey and completing the first set of demographic questions, we find that high income and high education significantly predict completing the survey and being included in the final sample. However, both effects are small, our final sample is balanced in terms of income, and our results are robust to using post-stratification weights that correct for the imbalances in education (see sensitivity analyses in the next two subsections).

Survey Duration and Remuneration The survey is relatively short to avoid response fatigue and ensure that respondents are willing to respond carefully to the open questions. The median response duration is approximately 7.5 minutes and most respondents complete the survey within 5 and 12 min (20%-80% quantile range). The standard reward for survey completion is $1.75.

Table C.1: Comparison of the Sample to the American Community Survey (ACS)

Variable	**ACS (2022)**	**Sample**
Gender		
Female	50%	51%
Age		
18-34	29%	32%
35-54	32%	33%
55+	38%	35%
Household net income		
Below 50k	34%	34%
50k-100k	29%	31%
Above 100k	37%	35%
Education		
Bachelor's degree or more	33%	60%
Region		
Northeast	17%	18%
Midwest	21%	22%
South	39%	39%
West	24%	22%
Race and ethnicity		
White	73%	78%
Black or African American	13%	11%
Hispanic/Latino	17%	7%
Asian	7%	8%
Political affiliation*		
Democrat	31%	53%
Republican	29%	21%
Independent	39%	26%
Sample size	1,980,550	2,000

*Data on political affiliation is taken from Chinoy et al. (2023) and based on Gallup surveys from the year 2022 (https://news.gallup.com/poll/15370/party-affiliation.aspx).

Notes: This table presents summary statistics from our sample and compares them to the American Community Survey (ACS) 2022. Respondents can identify with multiple races or ethnicities. We report statistics for the adult US population (18 years and above).

Table C.2: Predictors of Attrition

	Respondent is part of final sample (binary dummy)
Female (binary dummy)	−0.006
	(0.013)
Age (continuous)	0.001
	(0.000)
Income: 50-100k (binary dummy)	−0.001
	(0.018)
Income: 100k+ (binary dummy)	0.029^{*}
	(0.017)
At least Bachelor's degree (binary dummy)	0.048^{***}
	(0.015)
Region: Midwest (binary dummy)	−0.004
	(0.021)
Region: South (binary dummy)	0.001
	(0.019)
Region: West (binary dummy)	0.001
	(0.021)
Constant	0.820^{***}
	(0.029)
Observations	2,264
R^2	0.011

Notes: Results from an OLS regression, robust standard errors in parentheses. The sample includes all respondents who reached the main part of the survey. The outcome variable is a binary indicator that takes the value of 1 for respondents who are included in the final sample of the study. The regressors include various respondent characteristics. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

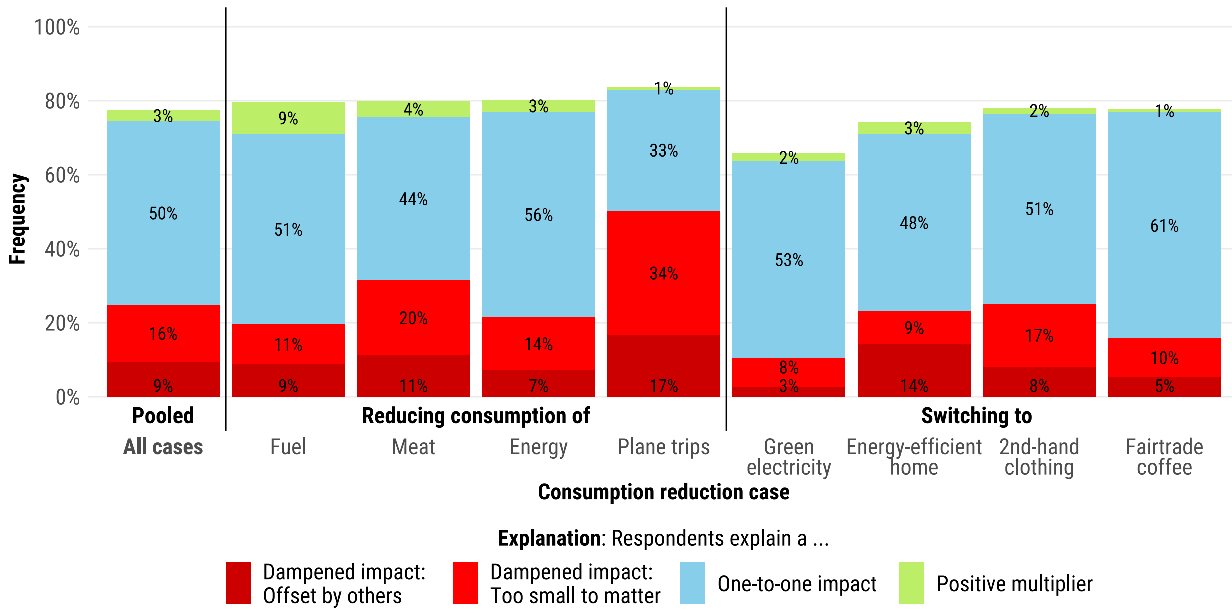
C.2 Additional Results: Belief in Dampened Effect

Table C.3: Heterogeneity: Who Predicts Dampened Effect?

	Belief in dampened effect (partial or full dampening; binary)	**Explanation of dampened effect** (binary)
	(1)	(2)
Female (binary)	−0.034	−0.008
	(0.022)	(0.019)
Age (continuous, in 10y)	−0.020***	−0.023***
	(0.007)	(0.006)
Income: 50-100k (binary)	0.022	0.035
	(0.027)	(0.024)
Income: 100k+ (binary)	−0.009	0.044*
	(0.028)	(0.024)
At least Bachelor's degree (binary)	−0.001	−0.030
	(0.023)	(0.021)
Politics: Independent (binary)	0.025	0.017
	(0.026)	(0.023)
Politics: Republican (binary)	0.064**	0.049**
	(0.028)	(0.025)
Consumes good (binary)	0.018	−0.008
	(0.027)	(0.023)
Constant	0.328***	0.290***
	(0.056)	(0.050)
Region FE	✓	✓
Case FE	✓	✓
Observations	2,000	2,000
R^2	0.060	0.081

Notes: Results from OLS regressions, robust standard errors in parentheses. The outcome variable is a binary indicator that takes the value of 1 for respondents who predict or explain that their own consumption reduction will lead to a partially or fully dampened reduction in aggregate consumption. The regressors include various respondent characteristics. The dummy "Consumes good" takes a value of 1 if the respondent reports that they regularly consume the good under consideration (fuel: drove 5,000 miles in last 12 months, meat: eat meat at least once per week, energy: annual energy bills of at least $500, flights: took at least one flight in last two years, electricity: annual electricity bill of at least $500, energy-efficient housing: annual energy bills of at least $500, clothing: purchase mostly new clothing, coffee: purchase mostly non-fairtrade coffee). The regressions contain census region and case (e.g., fuel, meat, flights ...) fixed effects. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Figure C.1: Reasoning about Dampening: Explanations Found in the Qualitative Text Data



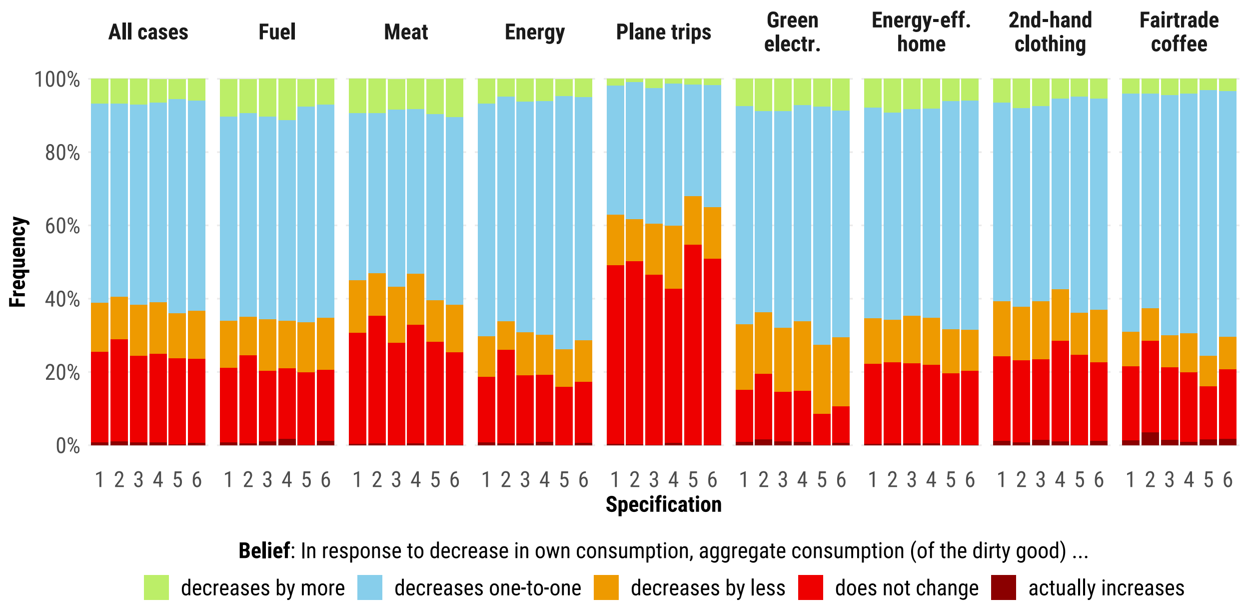
Notes: This figure displays the distributions of consumers' explanations for their beliefs about their own impact on aggregate consumption, as classified based on the open-ended text data. The first column displays results pooled across all eight cases, the other columns present the results for each of the eight cases. See Section C.4.1 for details on the coding scheme.

Sensitivity Tests Figure C.2 replicates the results for a variety of different specifications:

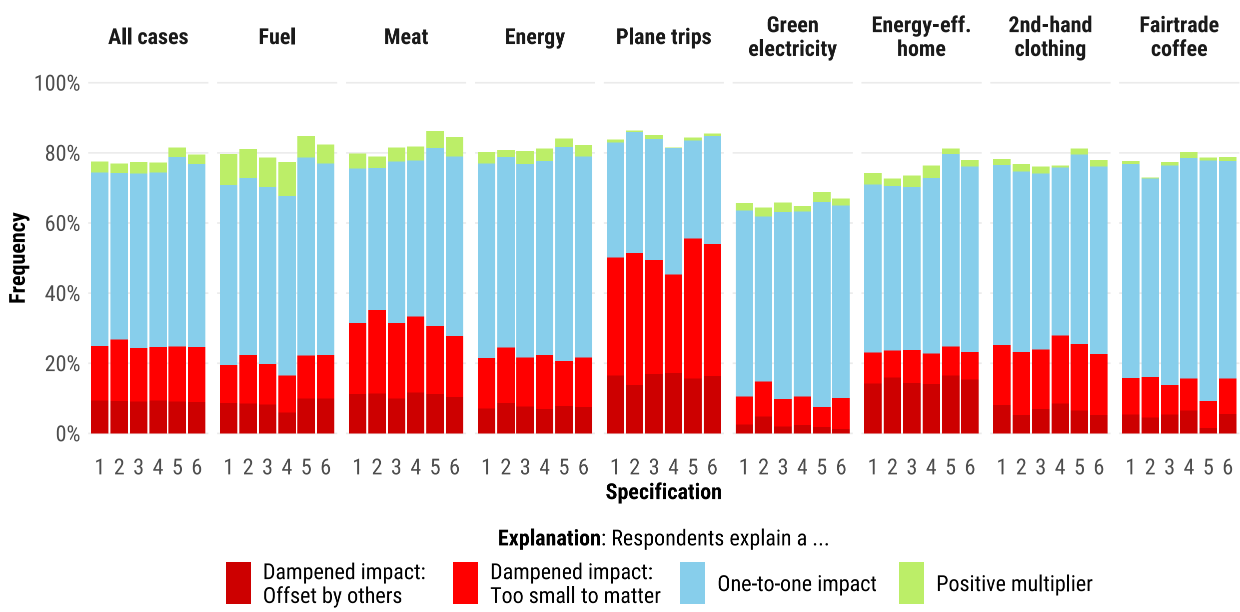
1. Main results.
2. Weighted sample. The weights correct for any imbalances in the characteristics reported in Table C.1. We follow the guidelines of the American National Election Study to calculate the survey weights using a raking procedure (Pasek et al., 2014).
3. We exclude the 20% of respondents with the shortest response duration, which potentially reflects that they paid comparatively less attention to the precise survey instructions.
4. We restrict the analysis to consumers who report regularly consuming the good under consideration.
5. We restrict the analysis to strict consequentialist consumers: valuation ratio = 0 (Figure 2).
6. We restrict the analysis to weak consequentialist consumers: valuation ratio < 1 (Figure 2).

Figure C.2: Robustness: Consumers' Beliefs and Reasoning about Dampening

(a) **Beliefs about Aggregate Impact of Own Consumption**



(b) **Explanations**



Notes: This figure displays the distributions of (a) consumers' beliefs and (b) explanations about their own impact on aggregate consumption for six different specifications. The specifications are described on the previous page. The first column displays results pooled across all eight cases, the other columns present the results for each of the eight cases.

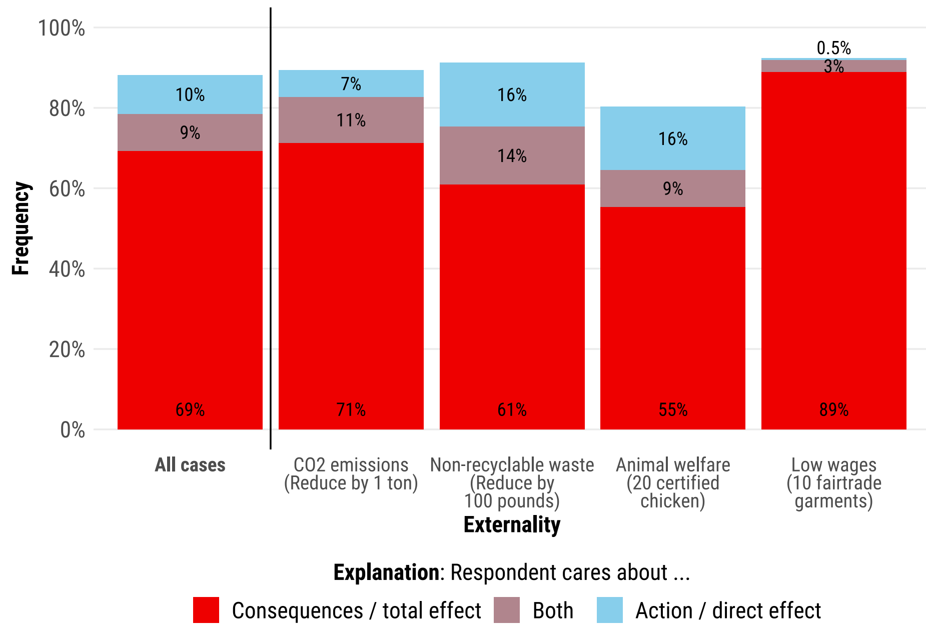
C.3 Additional Results: Nature of Social Concerns

Table C.4: Heterogeneity: Who Cares about Consequences?

	Cares at all?	**Valuation data**		**Explanation data**	
	Positive valuation	Strict conseq.	Weak conseq.	Strict conseq.	Weak conseq.
	(1)	(2)	(3)	(4)	(5)
Female	0.041***	0.026	0.011	0.015	0.035*
	(0.015)	(0.023)	(0.020)	(0.022)	(0.019)
Age (in 10y)	−0.014***	−0.028***	−0.036***	−0.020***	−0.024***
	(0.005)	(0.008)	(0.007)	(0.007)	(0.006)
Income: 50-100k	−0.001	0.000	0.005	0.054**	0.058**
	(0.019)	(0.029)	(0.025)	(0.028)	(0.025)
Income: 100k+	−0.004	0.003	0.010	0.064**	0.050**
	(0.020)	(0.030)	(0.025)	(0.028)	(0.025)
Bachelor's degree	0.013	0.015	0.022	0.004	0.010
	(0.017)	(0.025)	(0.021)	(0.024)	(0.021)
Politics: Independent	−0.113***	0.061**	0.030	0.070***	0.047**
	(0.020)	(0.028)	(0.024)	(0.026)	(0.023)
Politics: Republican	−0.132***	0.077**	0.031	0.016	−0.017
	(0.022)	(0.031)	(0.027)	(0.029)	(0.027)
Consumes good	−0.003	−0.015	0.002	−0.014	0.003
	(0.017)	(0.027)	(0.023)	(0.025)	(0.022)
Constant	1.012***	0.507***	0.750***	0.580***	0.693***
	(0.037)	(0.058)	(0.050)	(0.055)	(0.049)
Region FE	✓	✓	✓	✓	✓
Case FE	✓	✓	✓	✓	✓
Observations	2,000	1,714	1,714	1,702	1,702
R^2	0.057	0.088	0.089	0.091	0.081

Notes: Results from OLS regressions, robust standard errors in parentheses. The outcome variables are binary indicators. "Cares at all? Positive valuation" takes the value of 1 if the respondent positively values an effective reduction of the externality. The variables "Valuation data: Strict conseq." and "Weak conseq." take the value of 1 if the respondent has a valuation ratio of 0 or below 1, respectively (see Figure 2). "Explanation data: Strict conseq." and "Weak conseq." take the value of 1 if the respondent expresses (strict: only) consequentialist arguments in the open text data. The regressors include various respondent characteristics. The dummy "Consumes good" takes a value of 1 if the respondent reports that they regularly consume the good under consideration (CO_2: everyone, non-recyclable waste: everyone, animal welfare: eat meat at least once per week & eat mostly non-organic meat, low wages in textile industry: purchase mostly non-fairtrade garments). The regressions contain census region and case (e.g., CO_2, waste, ...) fixed effects. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Figure C.3: Social Concerns: Explanations Found in the Qualitative Text Data



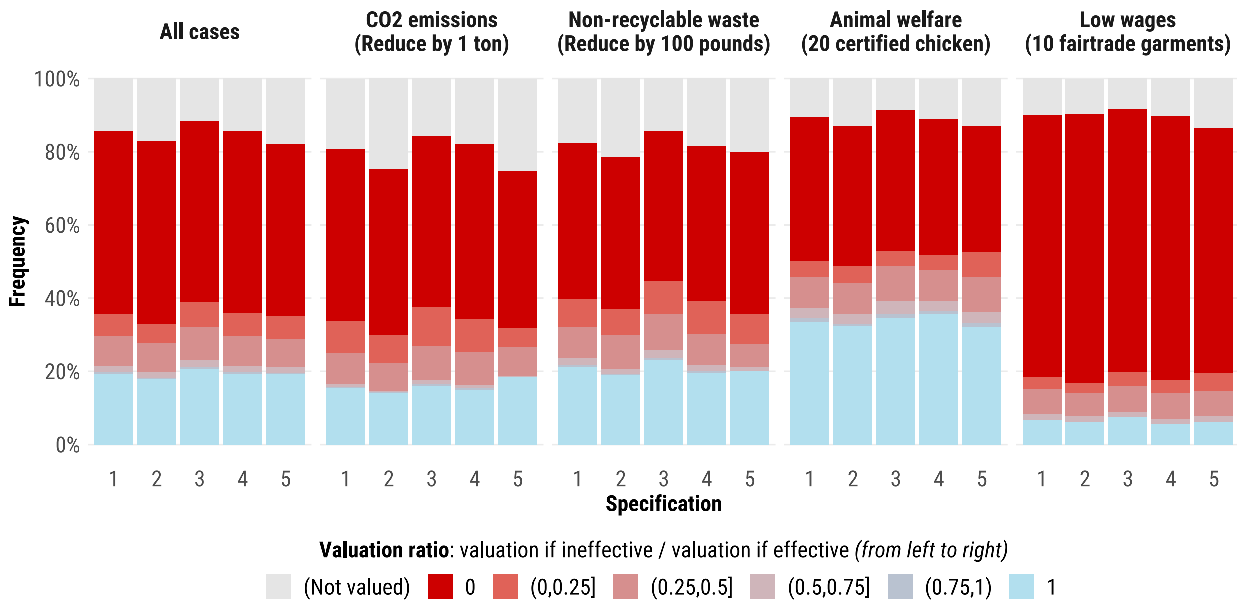
Notes: This figure displays the distributions of consumers' explanations for their valuations of effective and ineffective externality reductions, as classified based on the open-ended text data. Only consumers who positively value an effective reduction of the externality are asked to explain their responses. This figure presents results for these consumers. The first column displays results pooled across all four cases, the other columns present the results for each of the four cases. See Section C.4.2 for details on the coding scheme.

Sensitivity Tests Figure C.4 replicates the results for a variety of different specifications:

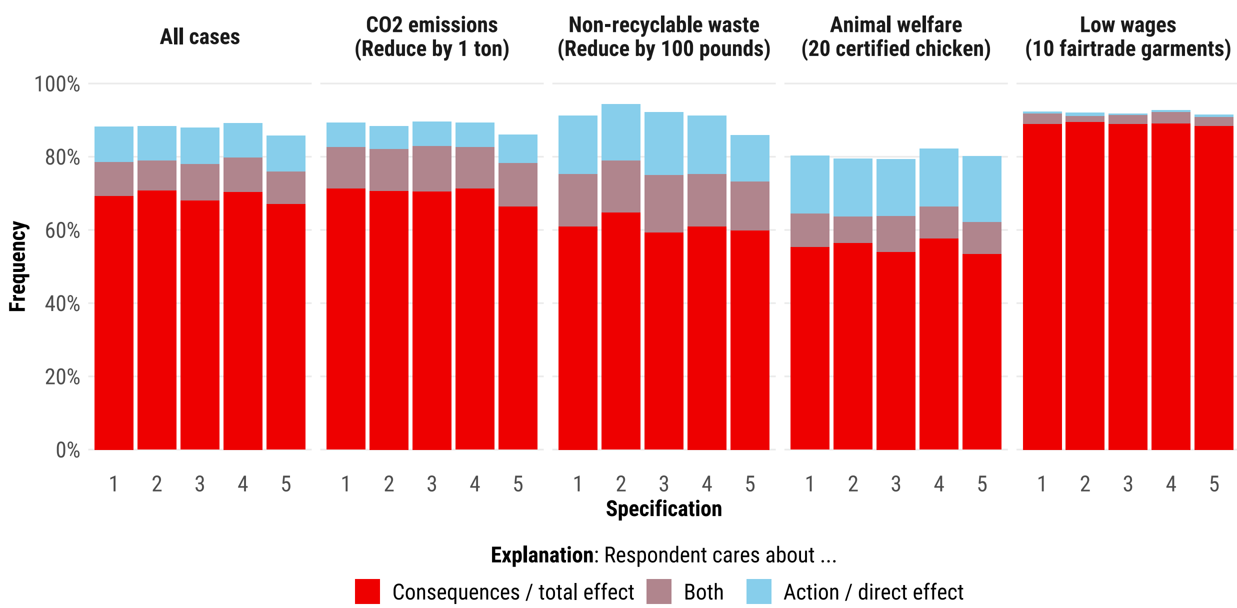
1. Main results.
2. Weighted sample. The weights correct for any imbalances in the characteristics reported in Table C.1. We follow the guidelines of the American National Election Study to calculate the survey weights using a raking procedure (Pasek et al., 2014).
3. We exclude the 20% of respondents with the shortest response duration, which potentially reflects that they paid comparatively less attention to the precise survey instructions.
4. We restrict the analysis to consumers who report regularly consuming the good under consideration.
5. We restrict the analysis the consumers who believe in full or partial dampening (Figure 1).

Figure C.4: Robustness: Consumers' Social Concerns

(a) **Valuation ratios**



(b) **Explanations**



Notes: This figure displays the distributions of (a) consumers' valuation ratios (consumers' valuation of the ineffective externality reduction divided by their valuation of the effective externality reduction) and (b) their explanations for their valuations for five different specifications. The different specifications are described on the previous page. The first column displays results pooled across all four cases, the other columns present the results for each of the four cases.

C.4 Categorization of the Qualitative Text-Data

C.4.1 Explanations of Beliefs in Dampened Effect

Coding Scheme Each open-text response is coded according to a detailed coding scheme that categorizes the line of reasoning that is expressed by the respondent. A summary of the three primary categories within our coding scheme, along with extra examples, can be found in Table C.5. The coding scheme was designed prior to the main data collection and was influenced by initial pilot interviews and the theoretical analysis in the paper. Responses that do not distinctly fall into one of the defined categories are classified into a residual category.

Ancillary Categories The coding scheme includes a few additional ancillary codes for the following type of responses.

1. Among responses assigned to the "dampened effect explanation" category, we distinguish between (i) respondents who refer to a consumption increase among other consumers and/or the price mechanism and (ii) respondents who argue that their individual impact is simply too small to shift aggregate quantities, without spelling out the precise dampening mechanism. 38% of all dampened effect explanations fall into the former category, and 62% fall into the latter category.

2. Responses that are assigned to the residual category and clearly reveal that the respondent did not even attempt to answer the question, e.g., copy-pasted their response from the scenario text or answered by writing about something completely unrelated to the question. This happens in less than 1% of all cases.

3. Responses that are assigned to the residual category and clearly reveal that the respondent misunderstood an important aspect of the question. For example, the response might reveal that the respondent believed that not only themselves but also others reduce their consumption or that the respondent predicted a general trend in global consumption rather than their own effect on global consumption. This applies to approximately 2% of all responses.

Coding Procedure Two research assistants trained in economics coded the text responses. They did not know the goals of our study. We use human coding because machine-based methods still struggle to detect the (often implicit) causal structure in human language.

To deal with the inherent subjectivity of human coding, we adopted two measures. First, we extensively trained the assistants. In the first training session, we taught the coding scheme and discussed many examples. Then, coders practiced on their own, and problematic cases were discussed, reviewed, and corrected in a second session — a process that we repeated one more time in a third and final training session. Second, each response was coded twice — independently by both reviewers. Whenever the two coders disagreed, we looked at the response and made a final decision. This approach ensures that close cases were reviewed a third time. It also allows us to assess the inter-rater reliability.

Inter-Rater Reliability We calculate how often the two independent reviewers assign the same code to a response. If we focus on our four key categories (dampened effect, one-to-one effect, positive multiplier, residual category), the coders agree in 87% of all cases. If we check consistency for the full coding scheme, including the ancillary codes, they agree in 82% of all cases. These numbers show that our coding scheme has a high degree of reliability.

Table C.5: Overview of the Coding Scheme for Reasoning about Dampening

Explanation	**Example**
Code: Dampened effect explanation	
Subcode: Own effect is offset by others Respondent mentions that others might increase their demand, and/or respondent explicitly refers to the mediating role of prices.	"My own consumption could be offset elsewhere so that it is less than 10 pounds." "I am but one person. Someone may choose to eat more, thus making my reduction less of an impact." "If I didn't buy the conventional coffee, someone else would. Maybe at a lower price, but it would be destroyed. Someone would offer the seller an acceptable price and the buyer would enjoy his coffee."
Subcode: Too small to matter Respondent argues that they are such a minuscule player on the global market that they have too little influence on aggregate consumption, and/or respondent argues that aggregate supply will not respond to their change in consumption.	"Just because I do not take a seat on the plane does not mean the plane will not fly." "Honestly, I don't think one person is going to impact global consumption of energy. The impact one person has for the entire global consumption is negligible."
Code: One-to-one effect explanation	
Respondent argues that their own consumption is part of the aggregate consumption so that a change in one variable implies a change in the other variable. Respondents might find this so obvious that they just refer to their own consumption reduction to explain the predicted global consumption reduction.	"It could only reduce the global consumption by the amount I saved." "If I reduce my own consumption it won't make others increase theirs. Therefore it will reduce the global consumption by the amount I reduced my own consumption." "Because my choosing to not buy 40 new garments would reduce the overall global consumption by 40 garments."
Code: Explanation for positive equilibrium multiplier	
Respondent explains why their own consumption reduction leads to an additional consumption reduction by others.	"Even though my reduced energy consumption would not be a significant reduction in the global consumption of energy my decrease would be an example for my family and friends." "I mean there's the 200 [gallons of fuel] we'd personally reduce and then less tankers and ships transporting it as well. So, I'd say more."

Notes: This table provides an overview of the three main categories in our coding scheme for respondents' explanation of their beliefs about dampening.

C.4.2 Explanations of Social Concerns

We follow an analogous procedure to classify the qualitative text data on consumers' reasoning about consequences.

Coding Scheme A summary of the two primary categories within our coding scheme, along with extra examples, can be found in Table C.6. Responses can also be assigned to both codes at the same time. Responses that do not distinctly fall into one of the defined categories are classified into a residual category.

Ancillary Categories The coding scheme includes a few additional ancillary codes for the following type of responses.

1. Responses that are assigned to the residual category and clearly reveal that the respondent did not even attempt to answer the question. This happens in less than 1% of all cases.

2. Responses that are assigned to the residual category and clearly reveal that the respondent misunderstood an important aspect of the question. This applies to less than 1% of all responses.

3. Responses that highlight potential positive consequences even in the scenario that states that respondents' contribution would be ineffective are an exception to this rule. These responses are common (11%) and signify strong consequentialist reasoning. In the main analysis, they are part of the consequentialist code.

4. Among responses assigned to the deontological / warm-glow code, we mark those who refer to the notion of personal responsibility (4% of all responses), choosing the morally right action (2%), or the desire to feel better about one's behavior (4%).

Inter-Rater Reliability We calculate how often the two independent reviewers assign the same code to a response. If we focus on our three key categories (consequentialist, deontological, residual category), the coders agree in 84% of all cases. If we check consistency for the full coding scheme, including the ancillary codes, they agree in 78% of all cases. These numbers show that our coding scheme has a high degree of reliability.

Table C.6: Overview of the Coding Scheme for Concerns for Consequences

Explanation	**Example**
Code: Consequentialist arguments	
The response reveals that consequences matter to the respondent.	"If it doesn't make a difference, then I don't see why I should pay anything to reduce my CO2 emissions." "I would pay extra to benefit workers only. I would not just pay extra and no one benefits."
Respondents explain a positive valuation for an ineffective reduction arguing that the action could eventually still lead to positive consequences.	"I would hope that the additional cost would maybe go to the organization's overhead and still benefit the workers in some way." "I still think it could have other positive effects. When people see you doing the right thing, they could be motivated to do the same."
Code: Deontological / warm-glow arguments	
The response reveals that the respondent cares about their action even if it has no net positive impact. For example, respondents might argue that they still want to do their own duty, follow a moral principle, or would feel better to at least try.	"It still seems like the right thing to do." "Because ethically it is the correct behavior. Just because the total impact is zero on the corporation's side, it still has impact personally since you are acting ethically. You do not get a pass to act unethically just because it has no effect on some other party." "I would be doing my bit and soothing my conscience. I would sleep better at night. This why I still recycle even though I've read creditable sources that my efforts are for naught [...]."

Notes: This table provides an overview of the two main categories in our coding scheme for respondents' explanation of their concerns about consequences.

C.5 Robustness Studies

In the design of our survey, we prioritize simplicity, a close relationship to relevant real-world settings, a clear mapping between evidence and theory, and we aim to give respondents a chance to share their reasoning. However, some of these design decisions also raise potential concerns that we address in this section.

All robustness studies are preregistered at www.doi.org/10.17605/osf.io/xw8mz.

C.5.1 Beliefs about Effects on Aggregate Production

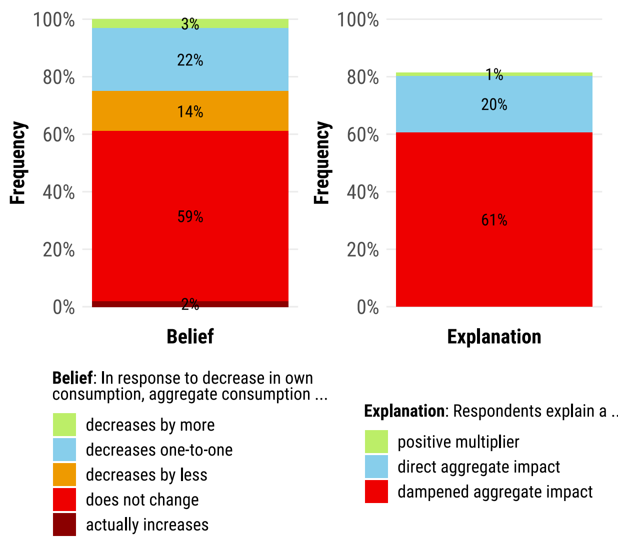
The main study elicits beliefs about aggregate consumption because the dampening mechanism in the model operates via the offsetting consumption increases of other agents. Of course, consumption and production are equivalent in the equilibrium of the model. Empirically, however, people's beliefs about dampening could be influenced by whether they think about the consumption or production side. How people reason about a problem often depends on which feature of the problem they pay attention to (Bordalo et al., 2023).

We explore this possibility in the **Production Robustness Study**. We conducted the study with the survey company Prolific and surveyed 259 US consumers in November 2023. Table C.7 describes the sample's demographic composition. The study builds on the "reducing your fuel consumption" scenario. Instead of asking respondents about their effect on the global total fuel *consumption*, we ask them about their effect on the global total fuel *production*.

The results are summarized in Figure C.5. When asked about their effect on production, even more consumers believe that they have a dampened aggregate effect. The share of consumers who predict a fully or partially dampened effect increases to 73%, and the share of consumers who provide explanations for a dampened effect increases to 61%. Many consumers argue that they do not affect aggregate production because producers will not even take notice of their individual consumption decrease.

The fact that consumers reason differently about consumption and production could indicate that many consumers believe that production can exceed consumption with the residual being wasted. As most externalities are created on the production, not the consumption side, this suggests

Figure C.5: Beliefs and Explanations in the Production Robustness Study



Notes: This figure presents results from the Production Robustness Study. It displays the distributions of consumers' beliefs about their own impact on aggregate consumption (left column) and their corresponding explanations, as classified based on the open-ended text data. See Section C.4.1 for details on the coding scheme of the open-text data. Each response is classified by one research assistant.

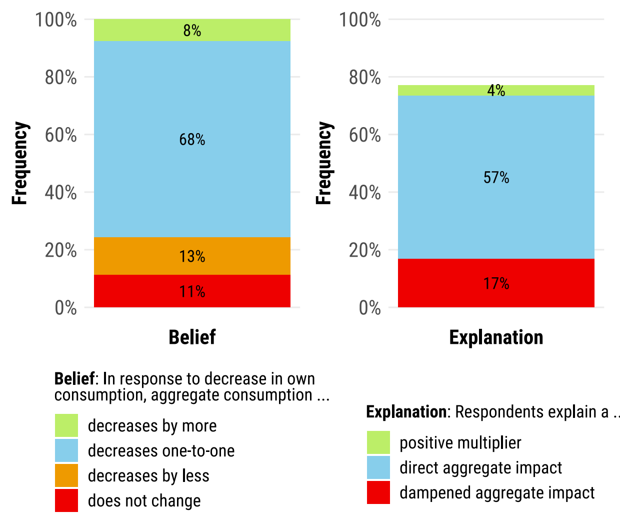
that our focus on consumption in the main study is conservative and tends to underestimate beliefs in a dampened impact.

C.5.2 Beliefs with Numeric Elicitation

The main study uses categorical response options. For example, in the scenario where respondents reduce their fuel consumption by 200 gallons, respondents predict whether global consumption would (i) decrease by more than 200 gallons, (ii) decrease by 200 gallons, (iii) decrease by less than 200 gallons, (iv) not change at all, or (v) actually increase. These categorical responses facilitate the subsequent measurement of open-ended explanations. We want respondents to explain why they think aggregate consumption falls by, say, less than 200 gallons (and not 200 gallons) rather than by 64 gallons (and not 89 gallons).

To ensure that this design choice does not have a strong effect on people's responses, we conduct the **Numeric Response Robustness Study**. We conducted the study with the survey company

Figure C.6: Beliefs and Explanations in the Numeric Response Robustness Study



Notes: This figure presents results from the Numeric Response Robustness Study. It displays the distributions of consumers' beliefs about their own impact on aggregate consumption (left column) and their corresponding explanations, as classified based on the open-ended text data. See Section C.4.1 for details on the coding scheme of the open-text data. Each response is classified by one research assistant.

Prolific and surveyed 250 US consumers in November 2023. Table C.7 describes the sample's demographic composition.

In this study, respondents predict the effect on aggregate consumption in an open numeric response box:

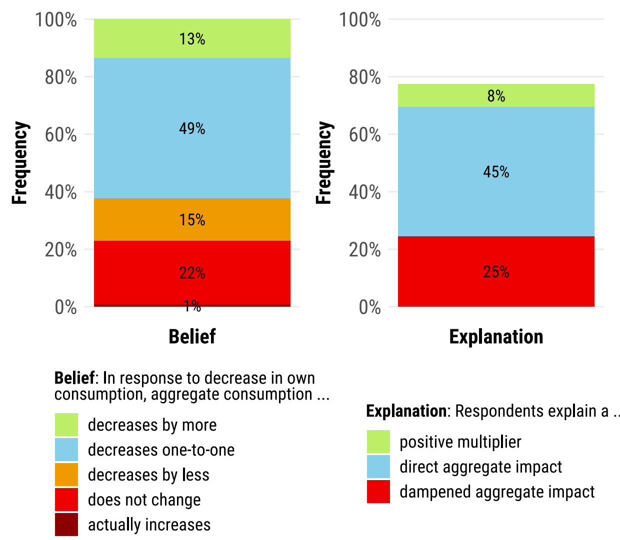
> You reduce your yearly fuel consumption by 200 gallons. This would reduce the yearly total global consumption of fuel by ...
>
> ___________ gallons

The results are summarized in Figure C.6 and qualitatively mirror the results of the main study.

C.5.3 Beliefs with Incentivization Scheme

Since we do not know the exact extent of dampening for the real-world markets and scenarios considered in the main study, we cannot incentivize beliefs in the main study. Fortunately, exist-

Figure C.7: Beliefs and Explanations in the Incentivized Beliefs Robustness Study



Notes: This figure presents results from the Incentivized Beliefs Robustness Study. It displays the distributions of consumers' beliefs about their own impact on aggregate consumption (left column) and their corresponding explanations, as classified based on the open-ended text data. See Section C.4.1 for details on the coding scheme of the open-text data. Each response is classified by one research assistant.

ing studies often find at most weak differences in the answers to incentivized and non-incentivized questions (Stantcheva, 2023). Nevertheless, we design an additional robustness study, the **Incentivized Beliefs Robustness Study**. We conducted the study with the survey company Prolific and surveyed 252 US consumers in November 2023. Table C.7 describes the sample's demographic composition.

Respondents make two predictions. First, they face the standard "reducing your fuel consumption" scenario. Then, they face an additional scenario that describes an introduction of a carbon tax in the United States, leading to a 1 billion ton reduction in US-wide CO_2 emissions. Respondents predict how this change would affect the yearly global emissions of CO_2. The scenario revolves around the issue of "carbon leakage", the concern that climate policies implemented in one country merely shift emissions to other countries instead of reducing them. Carbon leakage has frequently been studied by economists and their common conclusion is that carbon leakage is positive but not full (Dechezleprêtre and Sato, 2017, Grubb et al., 2022). In other words, we know that researchers' "best estimate" would be that the policy reduced the yearly global emissions of CO_2 by *less than 1 billion tons.*

The study employs a probabilistic incentivization scheme. Respondents are informed that

> "You can earn an additional bonus of \$2 if your predictions accord with recent research findings in economics. In particular, you will make two predictions. For one prediction, we reviewed the research literature in economics and determined a prediction that is plausible in light of recent research findings. If you make the same prediction, we will transfer a bonus of \$2 (or £1.70) to your Prolific account. However, you will not be told which of your two predictions will be tested, so please take both predictions seriously."

This approach allows us to truthfully incentivize both predictions, even though we do not know the correct answer to the first prediction. The procedure is akin to the approach taken by Bardsley (2000).

The results are summarized in Figure C.7 and closely mirror the results of the main study.[34]

[34] We find qualitatively and quantitatively very similar beliefs about dampened effects for a US-wide carbon tax. 30% predict a dampened effect of which, however, most respondents (77%) predict a partially dampened effect. This suggests that beliefs about dampening could also have consequences for the political support for climate policies.

Table C.7: Demographic Characteristics of the Samples in the Robustness Studies

Variable	**ACS (2022)**	**Production**	**Numeric Response**	**Incentivized Beliefs**
Gender				
Female	50%	51%	45%	47%
Age				
18-34	29%	38%	40%	45%
35-54	32%	51%	49%	40%
55+	38%	11%	11%	14%
Household net income				
Below 50k	34%	32%	31%	38%
50k-100k	29%	44%	42%	40%
Above 100k	37%	23%	26%	22%
Education				
Bachelor's degree or more	33%	56%	64%	55%
Region				
Northeast	17%	17%	22%	21%
Midwest	21%	19%	19%	21%
South	39%	48%	39%	40%
West	24%	17%	20%	18%
Race and ethnicity				
White	73%	75%	68%	71%
Black or Afric. American	13%	12%	15%	16%
Hispanic/Latino	17%	10%	12%	7%
Asian	7%	10%	7%	10%
Political affiliation*				
Democrat	31%	43%	48%	52%
Republican	29%	26%	21%	19%
Independent	39%	31%	31%	28%
Sample size	1,980,550	259	250	252

*Data on political affiliation is taken from Chinoy et al. (2023) and based on Gallup surveys from the year 2022 (https://news.gallup.com/poll/15370/party-affiliation.aspx).

Notes: This table presents summary statistics from our sample in the Robustness Studies and compares them to the American Community Survey (ACS) 2022. Respondents can identify with multiple races or ethnicities. We report statistics for the adult US population (18 years and above).

C.6 Instructions of Main Study

The complete instructions are available online at https://osf.io/u67wp. The survey begins with a participation information and informed consent form. Respondents who participate on a mobile device are screened out. Next, respondents have to pass an attention check. Subsequently, respondents fill out a block of demographic questions. Then, the main part of the survey begins.

Belief elicitation for the case of fuel consumption

[Respondents are randomized to one out of eight cases. The other cases are displayed below.]

Reducing your fuel consumption

Fuel consumption has a substantial impact on the climate. Transportation fuels account for a significant portion of global greenhouse gas emissions.

Your consumption of fuel is part of the **total global consumption of fuel**. We would like to know what you think would happen to the global consumption of fuel if you reduced your own consumption of fuel. Would it make a difference to the total consumption of fuel worldwide?

Consider these two scenarios:

Scenario 1: You consume 400 gallons of fuel every year.
Scenario 2: You consume 200 gallons of fuel every year.

In contrast to scenario 1, you would permanently reduce your yearly fuel consumption by 200 gallons in scenario 2.

What do you think would happen to the yearly total global consumption of fuel?

To repeat, you reduce your yearly fuel consumption by 200 gallons.
How would this affect the yearly total global consumption of fuel?

- It would **decrease** the yearly global consumption of fuel **by more than 200 gallons**.
- It would **decrease** the yearly global consumption of fuel **by 200 gallons**.
- It would **decrease** the yearly global consumption of fuel **by less than 200 gallons**.
- It would **not affect** the yearly global consumption of fuel.
- It would **actually increase** the yearly global consumption of fuel.

Please explain why you chose this response.

[Note: The order of response options is randomly reversed across respondents. If respondents select "decreases (...) by more", "decreases (...) by less", or "actually increases", a follow-up question appears on the same page.]

Follow-up question for "decreases (...) by more"

Please provide your best estimate.
You reduce your yearly fuel consumption by 200 gallons.
This decreases the yearly total global consumption of fuel by ...

- between 201 gallons to 300 gallons
- between 301 gallons to 400 gallons
- between 401 gallons to 500 gallons
- between 501 gallons to 600 gallons
- more than 600 gallons

Follow-up question for "decreases (...) by less"

(Order of response options is randomly reversed.)

Please provide your best estimate.
You reduce your yearly fuel consumption by 200 gallons.
This decreases the yearly total global consumption of fuel by ...

- between 150 gallons to 199 gallons
- between 100 gallons to 149 gallons
- between 50 gallons to 99 gallons
- between 1 gallons to 49 gallons

Follow-up question for "actually increases"

Please provide your best estimate.
You reduce your yearly fuel consumption by 200 gallons.
This increases the yearly total global consumption of fuel by ...

- between 1 gallons to 100 gallons
- between 101 gallons to 200 gallons
- between 201 gallons to 300 gallons
- between 301 gallons to 400 gallons
- more than 400 gallons

Additional consumption cases: Reducing consumption

Reducing your meat consumption

Meat consumption has a significant impact on the climate. According to a recent study, almost 60% of greenhouse gas emissions from food production come from meat alone. Moreover, meat often comes from industrial farming systems that not only subject animals to cramped, stressful conditions but also exact a heavy toll on the environment. On average, an American consumes about 200 pounds of meat per year.

Your consumption of meat is part of the **total global consumption of meat**. We would like to know what you think would happen to the global consumption of meat if you reduced your own meat consumption. Would it make a difference to the total consumption of meat worldwide?

Consider these two scenarios:

> Scenario 1: You eat 200 pounds of meat every year.
>
> Scenario 2: You eat 100 pounds of meat every year.

In contrast to scenario 1, you would permanently reduce your yearly meat consumption by 100 pounds in scenario 2.

Reducing your number of plane trips

Plane trips are a substantial contributor to climate change. They are particularly harmful for the climate because they emit large quantities of greenhouse gases, and they do so at high altitudes, where greenhouse gases have a more potent effect on the world's climate.

Your number of plane trips is part of the **total global number of plane trips**. We would like to know what you think would happen to the global number of plane trips if you reduced your own number of plane trips. Would it make a difference to the total number of plane trips worldwide?

Consider these two scenarios:

> Scenario 1: You take 8 plane trips every year.
>
> Scenario 2: You take 0 plane trips every year.

In contrast to scenario 1, you would permanently reduce your yearly number of plane trips by 8 trips in scenario 2.

Reducing your energy consumption

Household energy consumption is a significant contributor to climate change. In American homes, a variety of appliances and gadgets require energy to operate. About 50% of a household's yearly energy consumption is typically allocated to two main functions: heating and cooling. Energy consumption is measured in kilowatt-hours (kWh). It is not uncommon for US households to consume about 30,000 kWh of energy each year (electricity and gas).

Your energy consumption is part of the **total global consumption of energy**. We would like to know what you think would happen to the global consumption of energy if you reduced your own consumption of energy. Would it make a difference to the total consumption of energy worldwide?

Consider these two scenarios:

> Scenario 1: You consume 30,000 kWh of energy every year.
>
> Scenario 2: You consume 20,000 kWh of energy every year.

In contrast to scenario 1, you would permanently reduce your yearly energy consumption by 10,000 kWh in scenario 2.

Additional consumption cases: Reallocating consumption between close substitutes

Switching from brown to green electricity

You can choose between two types of electricity for your home: green or brown. Green electricity comes from clean sources like solar, wind, and water. Brown electricity comes from fossil fuels like coal and oil, which have a much higher carbon footprint and contribute to climate change. Electricity consumption is measured in kilowatt-hours (kWh). On average, a home in the US uses about 10,000 kWh of electricity each year.

Your consumption of brown electricity is part of the **total global consumption of brown electricity**. We would like to know what you think would happen to the global consumption of brown electricity if you fully switched from brown to green electricity. Would it make a difference to the total consumption of brown electricity worldwide?

Consider these two scenarios:

> Scenario 1: You only use brown electricity. Each year, you use 10,000 kWh of it.
>
> Scenario 2: You fully switch to green electricity. Each year, you use 10,000 kWh of it.

In contrast to scenario 1, you would permanently switch to green electricity in scenario 2.

Moving from an energy-inefficient to an energy-efficient home

Homes vary in their energy efficiency levels. An energy-efficient home is designed to use less energy due to [...]. On the other hand, an energy-inefficient home lacks these features and often wastes energy [...]. Energy consumption is measured in kilowatt-hours (kWh). It is not uncommon for US households to consume about 30,000 kWh of energy each year (electricity and gas).

Your consumption of energy is part of the **total global consumption of energy**. We would like to know what you think would happen to the global consumption of energy if you moved from an energy-inefficient home to an energy-efficient home. Would it make a difference to the total consumption of energy worldwide?

Consider these two scenarios:

> Scenario 1: You live in an energy-inefficient home. Each year, your household consumes 35,000 kWh of energy.
>
> Scenario 2: You move to an energy-efficient home. As a result, your household consumes 25,000 kWh of energy each year.

In contrast to scenario 1, you would permanently reside in an energy-efficient home and consume 10,000 kWh less each year in scenario 2.

Switching from new to second-hand clothing

You have a choice between two types of clothing: new clothing and second-hand clothing. Second-hand clothing refers to clothes that have been previously owned by someone else. By contrast, new clothing is freshly produced and has not been worn by other people before. The production of clothing is highly resource-intensive and clothes are often produced under poor conditions with low wages for laborers. It is not uncommon in the US to buy 40 garments per year.

Your consumption of new clothing is part of the **total global consumption of new clothing**. We would like to know what you think would happen to the global consumption of clothing if you fully switched from new clothing to second-hand clothing. Would it make a difference to the total consumption of new clothing worldwide?

Consider these two scenarios:

> Scenario 1: You only buy new clothing. Each year, you purchase 40 new garments.
>
> Scenario 2: You fully switch to second-hand clothing. Each year, you purchase 40 second-hand garments.

In contrast to scenario 1, you would permanently switch to second-hand clothing in scenario 2.

Switching from conventional to fairtrade coffee

You can choose between two types of coffee: fairtrade or conventional. Fairtrade coffee ensures that farmers receive a fair wage and work in safe conditions. The fairtrade system also encourages sustainable farming practices that are better for the environment. By contrast, conventional coffee often comes from large-scale industrial farming systems that may underpay farmers and usually does not prioritize sustainable farming methods. On average, an American consumes about 10 pounds of coffee beans per year.

Your consumption of conventional coffee is part of the **total global consumption of conventional coffee**. We would like to know what you think would happen to the global consumption of conventional coffee if you fully switched from conventional to fairtrade coffee. Would it make a difference to the total consumption of conventional coffee worldwide?

Consider these two scenarios:

> Scenario 1: You only drink conventional coffee. Each year, you consume 10 pounds of conventional coffee.
>
> Scenario 2: You fully switch to fairtrade coffee. Each year, you consume 10 pounds of fairtrade coffee.

In contrast to scenario 1, you would permanently switch to fairtrade coffee in scenario 2.

Questions on the nature of social concerns

[Respondents are randomized to one out of four cases. The other cases are displayed below.]

Carbon emissions

Now, on a different topic ...

The average US American emits about 15 tons of carbon dioxide (CO2) each year. Reducing CO2 emissions often comes at a financial cost. For example, the cost could arise because you buy eco-friendly products, invest in energy-saving solutions, transition to a renewable energy supplier, or acquire carbon offsets. We want to learn what the highest cost is that you would be willing to pay to reduce your personal CO2 emissions.

Please consider two different situations.

Situation 1: Your action has positive consequences

In situation 1, if you reduce your personal CO2 emissions by one ton, the **total global CO2 emissions also decrease by one ton**.

> **Your action**
> You reduce your personal CO2 emissions by one ton.
> **Consequence**
> The total global CO2 emissions decrease by one ton.

In situation 1, how much money would you be willing to pay to reduce your carbon emissions by 1 ton?

$ ______

Situation 2: Your action does not have any consequence

Now, please assume that, if you reduce your personal CO2 emissions by one ton, the **total global CO2 emissions do not change**. (For example, this could happen because your reduced emissions make it cheaper for others to emit, and consequently others increase their CO2 emissions by one ton, exactly offsetting your reduction.)

> **Your action**
> You reduce your personal CO2 emissions by one ton.
> **Consequence**
> The total global CO2 emissions do not change.

In situation 2, how much money would you be willing to pay to reduce your carbon emissions by 1 ton?

$ ______

[Note: Depending on the answers, open-ended text questions appear on the same page.]

Follow-up question if answers are identical

Please explain why you gave the same answer in the two situations.

Follow-up question if answers differ

Please explain why you gave different answers in the two situations.

Follow-up question if answer is positive in situation 2

Please explain why you would be willing to pay money in situation 2 where the total impact would be zero.

Additional cases (shortened)

Non-recyclable waste
Many households generate a significant amount of waste, a large proportion of which is not recyclable. Non-recyclable waste often results in more environmental pollution [...]

Please consider two different situations.

Situation 1: Your action has positive consequences
In this situation, if you reduce your non-recyclable waste by 100 pounds, the **total amount of non-recyclable waste in your community and all across the world also decreases by 100 pounds**.

Your action [...] / **Consequence** [...]

In situation 1, how much money would you be willing to pay to reduce your non-recyclable waste by 100 pounds?

Situation 2: Your action does not have any consequence
Now, please assume that, if you reduce your personal non-recyclable waste by 100 pounds, the **total non-recyclable waste in your community and all across the world does not change**. (For example, this could happen because your reduced waste makes it cheaper for others to dispose waste, and consequently others increase their non-recyclable waste by 100 pounds, exactly offsetting your reduction.)

Your action [...] / **Consequence** [...]

In situation 2, how much money would you be willing to pay to reduce your non-recyclable waste by 100 pounds?

Fairtrade wages
Workers in the textile industry in developing countries often earn wages that are below a decent living wage. Paying higher "fairtrade" wages (as defined by the Fairtrade Foundation) [...]

Please consider two different situations.

Situation 1: Your action has positive consequences
In this situation, for every additional dollar that you spend on fairtrade wages, **workers' total wages increase by that same dollar**.

Your action [...] / **Consequence** [...]

In situation 1, how much additional (beyond the standard price) would you be willing to spend on ten fairtrade garments?

Situation 2: Your action does not have any consequence
Now, please assume that if you choose to pay a premium to support fairtrade wages, **workers' total wages do not change**. (For example, this could happen because when you pay more for fairtrade products, you increase the demand for these products, making them more expensive for others, and they buy less of it.)

[...]

Animal welfare
On average, an American eats more than 20 chickens annually. Many of these chickens are raised in poor, cramped conditions [...]

Please consider two different situations.

Situation 1: Your action has positive consequences
In this situation, if you choose to pay a premium to support enhanced animal welfare standards for 20 chickens, the **overall number of chickens raised under better welfare standards increases by 20**.

Your action [...] / **Consequence** [...]

In situation 1, how much additional (beyond the standard price) would you be willing to spend on 20 animal-welfare-certified chickens?

Situation 2: Your action does not have any consequence
Now, please assume that if you choose to pay a premium to support enhanced animal welfare standards for 20 chickens, the **overall number of chickens raised under better welfare standards does not change**. (For example, this could happen because when you pay more for certified animal welfare products, you increase the demand for certified chicken, making it more expensive for others, and they buy less of it.)

Your action [...] / **Consequence** [...]

In situation 2, how much additional (beyond the standard price) would you be willing to spend on 20 animal-welfare-certified chickens?