Appendix

A. Proofs

Proof of Proposition 1

This proof consists of several steps. We proof each of them individually:

Claim 1: Reviews converge, and the more polarizing product 2 is relatively overrated: $\mathcal{B}_\infty < 0$ or, equivalently since $\Delta(Q) = 0$, $\mathbb{E}_\infty(\mathcal{R}_2) > \mathbb{E}_\infty(\mathcal{R}_1)$

Proof of Claim 1:

We start by showing that reviews – and thus, *a fortiori*, their difference $\mathcal{B}_t$ – converge.

Because consumers are initially uniformed about quality differences, it is easy to see that $\mathbb{E}_0(\mathcal{R}) = Q_1 + \mathbb{E}(\theta_1|\theta_1 > \theta_2)$ and $\mathbb{E}_0(\mathcal{R}) = Q_2 + \mathbb{E}(\theta_2|\theta_2 > \theta_1)$. Because $\Delta(Q) = 0$ and s_1 is less polarizing than s_2, we have $\Delta_0(\mathcal{R}) < 0$.

Thus,

$$\Delta_1(\mathcal{R}) = \mathbb{E}(\theta_1|\theta_1 > \theta_2 - \Delta_0(\mathcal{R})) - \mathbb{E}(\theta_2|\theta_2 > \theta_1 + \Delta_0(\mathcal{R})).$$

Clearly, $\Delta_0(\mathcal{R}) < 0 \Rightarrow \Delta_1(\mathcal{R}) > \Delta_0(\mathcal{R})$.

Moreover, we have that

$$\begin{aligned}
\Delta_1(\mathcal{R}) &= \mathbb{E}\Big(\theta_1|\theta_1 > \theta_2 - \Delta_0(\mathcal{R})\Big) - \mathbb{E}\Big(\theta_2|\theta_2 > \theta_1 + \Delta_0(\mathcal{R})\Big) \\
&< \mathbb{E}(\theta_1|\theta_1 > \theta_2) - \frac{\Delta_0(\mathcal{R})}{2} - \Big(\mathbb{E}(\theta_2|\theta_2 > \theta_1) + \frac{\Delta_0(\mathcal{R})}{2}\Big) \\
&= \Big(\mathbb{E}(\theta_1|\theta_1 > \theta_2) - \mathbb{E}(\theta_2|\theta_2 > \theta_1)\Big) - \frac{\Delta_0(\mathcal{R})}{2} - \frac{\Delta_0(\mathcal{R})}{2} \\
&= \Delta_0(\mathcal{R}) - \frac{\Delta_0(\mathcal{R})}{2} - \frac{\Delta_0(\mathcal{R})}{2} \\
&= 0
\end{aligned}$$

where the inequality follows from Assumption 1, which implies that $\mathbb{E}(\theta_1|\theta_1 > \theta_2 - \Delta_0(\mathcal{R})) \leq \mathbb{E}(\theta_1|\theta_1 > \theta_2) - \frac{\Delta_0(\mathcal{R})}{2}$ and $\mathbb{E}(\theta_2|\theta_2 > \theta_1 + \Delta_0(\mathcal{R})) \geq \mathbb{E}(\theta_2|\theta_2 > \theta_1) + \frac{\Delta_0(\mathcal{R})}{2}$.

Thus, $\Delta_1(\mathcal{R}) \in (\Delta_0(\mathcal{R}), 0)$. Because $\Delta_{t+1}(\mathcal{R})$ is monotonically decreasing in $\Delta_t(\mathcal{R})$, this implies that $\Delta_2(\mathcal{R}) \in (\Delta_0(\mathcal{R}), \Delta_1(\mathcal{R}))$.

By induction, it is easy to see that

$$\Delta_{2k} \in (\Delta_{2k-2}(\mathcal{R}), \Delta_{2k-1}(\mathcal{R})), \quad \Delta_{2k+1} \in (\Delta_{2k}(\mathcal{R}), \Delta_{2k-1}(\mathcal{R})) \quad \forall k \geq 0. \tag{9}$$

Notice how this implies that the sequences $\{\Delta_{2k}(\mathcal{R})\}_{k=0}^{\infty}$ is increasing, while $\{\Delta_{2k+1}(\mathcal{R})\}_{k=0}^{\infty}$ is decreasing.

Moreover, Equation 9 implies that

$$\Delta_{2k}(\mathcal{R}) - \Delta_{2k+1}(\mathcal{R}) < \Delta_{2k-2}(\mathcal{R}) - \Delta_{2k-1}(\mathcal{R}).$$

Thus, $\{\Delta_{2k}(\mathcal{R}) - \Delta_{2k+1}(\mathcal{R})\}_{k=0}^{\infty} \nearrow 0$, proving that reviews to converge to $\Delta_\infty(\mathcal{R})$, which is defined by the unique solution of Equation 6. Because the right hand side of $\mathbb{E}_{t+1}(\mathcal{R}_1)$ and $\mathbb{E}_{t+1}(\mathcal{R}_2)$ solely depends on $\Delta_t(\mathcal{R})$, by continuity this implies that $\mathbb{E}_t(\mathcal{R}_1)$ and $\mathbb{E}_t(\mathcal{R}_2)$ also converge.

Subtracting the second line in Eq. 6 from the first we get:

$$\begin{aligned}\mathbb{E}_\infty(\mathcal{R}_1) - \mathbb{E}_\infty(\mathcal{R}_2) = (Q_1 - Q_2) &+ \mathbb{E}(\theta_{1j} \mid \mathbb{E}_\infty(\mathcal{R}_1) + \theta_{1j} > \mathbb{E}_\infty(\mathcal{R}_2) + \theta_{2j})) \\ &- \mathbb{E}(\theta_{2j} \mid \mathbb{E}_\infty(\mathcal{R}_2) + \theta_{2j} > \mathbb{E}_\infty(\mathcal{R}_1) + \theta_{1j})).\end{aligned} \tag{10}$$

Using the definition of

$$\mathcal{B}_\infty = (\mathbb{E}_\infty(\mathcal{R}_1) - \mathbb{E}_\infty(\mathcal{R}_2)) - (Q_1 - Q_2),$$

we can simplify this expression to

$$\mathcal{B}_\infty = \mathbb{E}(\theta_{1j} \mid \mathbb{E}_\infty(\mathcal{R}_1) + \theta_{1j} > \mathbb{E}_\infty(\mathcal{R}_2) + \theta_{2j})) - \mathbb{E}(\theta_{2j} \mid \mathbb{E}_\infty(\mathcal{R}_2) + \theta_{2j} > \mathbb{E}_\infty(\mathcal{R}_1) + \theta_{1j})).$$

Now note that $Q_1 = Q_2$ by assumption, and thus $\mathcal{B}_\infty = (\mathbb{E}_\infty(\mathcal{R}_1) - \mathbb{E}_\infty(\mathcal{R}_2))$.

Therefore, the above Eq. can be rewritten as

$$\mathcal{B}_\infty = \mathbb{E}(\theta_{1j} \mid \theta_{1j} > \theta_{2j} - \mathcal{B}_\infty)) - \mathbb{E}(\theta_{2j} \mid \theta_{2j} > \theta_{1j} + \mathcal{B}_\infty)). \tag{11}$$

We now have one Eq. in one variable, $\mathcal{B}_\infty$. To show that a solution exists and that it is unique, first notice that the LHS of Eq. 11 is (trivially) increasing in $\mathcal{B}_\infty$. The RHS, on the other hand, is decreasing in $\mathcal{B}_\infty$: this follows from the fact that $\mathbb{E}(\theta_{1j} \mid \theta_{1j} > \theta_{2j} - \mathcal{B}_\infty)$ is decreasing in $\mathcal{B}_\infty$, due to basic properties of conditional expectations, while the opposite is true for $\mathbb{E}(\theta_{2j} \mid \theta_{2j} > \theta_{1j} + \mathcal{B}_\infty)$.

To show that the (only) solution $\mathcal{B}_\infty$ is negative, therefore, we have to show that *i*) if $\mathcal{B}_\infty = 0$, the RHS is negative and *ii*) if $\mathcal{B}_\infty$ becomes small (in a way that will be defined later in the proof), the RHS is larger than the LHS.

Let's start with *i*). Notice that whenever $\mathcal{B}_\infty = 0$, the RHS becomes

$$\mathbb{E}\big(\theta_{1j} \mid \theta_{1j} \geq \theta_{2j}\big) - \mathbb{E}\big(\theta_{2j} \mid \theta_{2j} \geq \theta_{1j}\big).$$

Therefore, we have to show the following:

Lemma 2. *Let the assumptions of Proposition 1 hold. Then*

$$\mathbb{E}\big(\theta_2 \mid \theta_2 \geq \theta_1\big) > \mathbb{E}(\mathbb{E}\big(\theta_1 \mid \theta_1 \geq \theta_2\big).$$

Proof:

We have that

$$\mathbb{E}(\theta_1|\theta_1 \geq \theta_2) = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\theta_1} x_1 f_1(\theta_1) f_2(\theta_2)\, d\theta_2\, d\theta_1}{\int_{-\infty}^{\infty}\int_{-\infty}^{\theta_1} f_1(\theta_1) f_2(\theta_2)\, d\theta_2\, d\theta_1} = \frac{\mathbb{E}_{F_{s_1}}(\theta F_{s_2}(\theta))}{1/2}$$

and similarly

$$\mathbb{E}(\theta_2|\theta_2 \geq \theta_1) = \frac{\mathbb{E}_{F_{s_2}}(\theta F_{s_1}(\theta))}{1/2},$$

where the denominators simplify to $1/2 = Prob(\theta_2 > \theta_1) = Prob(\theta_1 > \theta_2)$ given symmetry.

Thus, we are left with having to show that

$$\mathbb{E}_{F_{s_2}}(\theta F_{s_1}(\theta)) > \mathbb{E}_{F_{s_1}}(\theta F_{s_2}(\theta)).$$

Let us start by focusing on $\theta > 0$. Notice that symmetry implies that $F_{s_2}(\theta) < F_{s_1}(\theta)$ if and only if $\theta > 0$. In other words, for $\theta > 0$ we have that s_2 FOSD s_1. To show that the above inequality holds, we show that

$$\mathbb{E}_{F_{s_2}}(\theta F_{s_1}(\theta)) > \mathbb{E}_{F_{s_1}}(\theta F_{s_1}(\theta)) > \mathbb{E}_{F_{s_1}}(\theta F_{s_2}(\theta)).$$

The proof that $\mathbb{E}_{F_{s_1}}(\theta F_{s_1}(\theta)) > \mathbb{E}_{F_{s_1}}(\theta F_{s_2}(\theta))$ is immediate, since $F_{s_2}(\theta) < F_{s_1}(\theta)$ for every $\theta > 0$. To show that $\mathbb{E}_{F_{s_2}}(\theta F_{s_1}(\theta)) > \mathbb{E}_{F_{s_1}}(\theta F_{s_1}(\theta))$, notice that this is a property of FOSD (which holds for $\theta > 0$): for every non-negative function, the expected value under F_{s_2} is higher than under F_{s_1}.

To show the same for $\theta < 0$, notice that here $F_{s_2}(\theta) > F_{s_1}(\theta)$.

Once again, to show that the above inequality holds, we show that

$$\mathbb{E}_{F_{s_2}}(\theta F_{s_1}(\theta)) > \mathbb{E}_{F_{s_1}}(\theta F_{s_1}(\theta)) > \mathbb{E}_{F_{s_1}}(\theta F_{s_2}(\theta)).$$

The proof that $\mathbb{E}_{F_{s_1}}(\theta F_{s_1}(\theta)) > \mathbb{E}_{F_{s_1}}(\theta F_{s_2}(\theta))$ is immediate, since $F_{s_2}(\theta) > F_{s_1}(\theta)$ for every $\theta < 0$, which implies $\theta F_{s_2}(\theta) < \theta F_{s_1}(\theta)$. To show that $\mathbb{E}_{F_{s_2}}(\theta F_{s_1}(\theta)) > \mathbb{E}_{F_{s_1}}(\theta F_{s_1}(\theta))$, notice that this is a property of FOSD (which holds for $\theta < 0$): for every negative function, the expected value under F_{s_1} is higher (less negative) than under F_{s_2}. This concludes the proof of Lemma 5. ■

Conversely, denote by $\bar{\theta}_1$ and $\underline{\theta}_2$ the maximum possible value for θ_1 and minimum possible value for θ_2 respectively. Whenever $\mathcal{B}_\infty < \underline{\theta}_2 - \bar{\theta}_1$, we have $\mathcal{B}_\infty + \theta_1 \leq \underline{\theta}_2$ and thus

$$\mathbb{E}\big(\theta_2 \mid \theta_2 > \theta_1 + \mathcal{B}_\infty\big) \leq \mathbb{E}\big(\theta_2 \mid \theta_2 > \underline{\theta}_2\big) = \mathbb{E}\big(\theta_2\big) = 0.$$

On the other hand, it is straightforward to see that the first conditional expected value on the RHS is positive.

Therefore, we have that $\mathcal{B}_\infty < 0$, as desired. ■

Claim 2: Product 2 thus obtains a higher number of reviews: $\mathcal{N}_\infty(\mathcal{R}_2) > 1/2$.

Proof of Claim 2:
We have that

$$\mathcal{N}_\infty(\mathcal{R}_2) = Prob(\theta_{2j} > \theta_{1j} + \mathcal{B}_\infty).$$

Since $\mathcal{B}_\infty < 0$, we have $\mathcal{N}_\infty(\mathcal{R}_2) > 1/2$. ■

Claim 3: Nevertheless, some self-correction occurs, and both biases are less severe than in the short-run: $\mathcal{B}_0 < \mathcal{B}_\infty < 0$, $\mathcal{N}_0(\mathcal{R}_2) > \mathcal{N}_\infty(\mathcal{R}_2) > 1/2$.

Proof of Claim 3:
We have that

$$\mathbb{E}_0(\mathcal{R}_1) = Q_1 + \mathbb{E}(\theta_{1j}|\theta_{1j} > \theta_{2j})) \tag{12}$$

and similarly for product 2. Thus, $\mathcal{B}_0 = \mathbb{E}_1(\mathcal{R}_1) - \mathbb{E}_1(\mathcal{R}_2)$ implies

$$\mathcal{B}_0 = \mathbb{E}(\theta_{1j} \mid \theta_{1j} > \theta_{2j})) - \mathbb{E}(\theta_{2j} \mid \theta_{2j} > \theta_{1j})), \tag{13}$$

where we have simplified the RHS using the fact that $Q_1 = Q_2$ by assumption.
To show that $\mathcal{B}_0 < \mathcal{B}_\infty$, assume by contradiction $\mathcal{B}_0 = \mathcal{B}_\infty$. But then, we obtain

$$\mathcal{B}_0 = \mathbb{E}(\theta_{1j} \mid \theta_{1j} > \theta_{2j} - \mathcal{B}_0)) - \mathbb{E}(\theta_{2j} \mid \theta_{2j} > \theta_{1j} + \mathcal{B}_0)). \tag{14}$$

Therefore,

$$\begin{aligned}\mathcal{B}_0 &= \mathbb{E}(\theta_{1j} \mid \theta_{1j} > \theta_{2j})) - \mathbb{E}(\theta_{2j} \mid \theta_{2j} > \theta_{1j})) \\ &< \mathbb{E}(\theta_{1j} \mid \theta_{1j} > \theta_{2j} - \mathcal{B}_\infty)) - \mathbb{E}(\theta_{2j} \mid \theta_{2j} > \theta_{1j} + \mathcal{B}_\infty)) \\ &= \mathcal{B}_\infty,\end{aligned}$$

where the inequality follows from the fact that $\mathcal{B}_\infty < 0$, as established in the Proof of **Claim 1**. The conclusions hold *a fortiori* if $\mathcal{B}_0 > \mathcal{B}_\infty$. This proves that $\mathcal{B}_0 < \mathcal{B}_\infty$. The fact that $\mathcal{N}_1(\mathcal{R}_2) > \mathcal{N}_\infty(\mathcal{R}_2)$ follows straightforwardly from $\mathcal{B}_0 < \mathcal{B}_\infty$ using the same argument as in the Proof of **Claim 2**. ■

Proof of Proposition 2

Claim 1: Reviews converge, and the higher quality product 1 has higher reviews: $\mathbb{E}_\infty(\mathcal{R}_1) > \mathbb{E}_\infty(\mathcal{R}_2)$,

Proof of Claim 1:
We start by showing that reviews – and thus, *a fortiori*, $\mathcal{B}_t$ – converge.
Because consumers are initially uniformed about quality differences, it is easy to see that $\mathbb{E}_0(\mathcal{R}) = Q_1 + \mathbb{E}(\theta_1|\theta_1 > \theta_2)$ and $\mathbb{E}_0(\mathcal{R}) = Q_2 + \mathbb{E}(\theta_2|\theta_2 > \theta_1)$. Because $\Delta(Q) > 0$ and $s_1 = s_2$, we have $\Delta_0(\mathcal{R}) = \Delta(Q)$. Thus,

$$\Delta_1(\mathcal{R}) = \Delta(Q) + \mathbb{E}(\theta_1|\theta_1 > \theta_2 - \Delta_0(\mathcal{R})) - \mathbb{E}(\theta_2|\theta_2 > \theta_1 + \Delta_0(\mathcal{R})).$$

Clearly, $\Delta_0(\mathcal{R}) > 0 \Rightarrow \Delta_1(\mathcal{R}) < \Delta_0(\mathcal{R})$.

Moreover, we have that

$$\begin{aligned}
\Delta_1(\mathcal{R}) &= \Delta(Q) + \mathbb{E}(\theta_1|\theta_1 > \theta_2 - \Delta_0(\mathcal{R})) - \mathbb{E}(\theta_2|\theta_2 > \theta_1 + \Delta_0(\mathcal{R})) \\
&= \Delta(Q) + \mathbb{E}(\theta_1|\theta_1 > \theta_2 - \Delta(Q)) - \mathbb{E}(\theta_2|\theta_2 > \theta_1 + \Delta(Q)) \\
&> \Delta(Q) + \mathbb{E}(\theta_1|\theta_1 > \theta_2) - \frac{\Delta(Q)}{2} - \left(\mathbb{E}(\theta_2|\theta_2 > \theta_1) + \frac{\Delta(Q)}{2}\right) \\
&= 0,
\end{aligned}$$

where the last equality follows from the fact that $s_1 = s_2$ (and, therefore, $\mathbb{E}(\theta_1|\theta_1 > \theta_2) = \mathbb{E}(\theta_2|\theta_2 > \theta_1)$, while the inequality follows from our regularity assumption **??**, which implies that $\mathbb{E}(\theta_1|\theta_1 > \theta_2 - \Delta(Q)) \geq \mathbb{E}(\theta_1|\theta_1 > \theta_2) - \frac{\Delta(Q)}{2}$ and $\mathbb{E}(\theta_2|\theta_2 > \theta_1 + \Delta(Q)) \leq \mathbb{E}(\theta_2|\theta_2 > \theta_1) + \frac{\Delta(Q)}{2}$.

Thus, $\Delta_1(\mathcal{R}) \in (\Delta_0(\mathcal{R}), 0)$. Because $\Delta_{t+1}(\mathcal{R})$ is monotonically decreasing in $\Delta_t(\mathcal{R})$, this implies that $\Delta_2(\mathcal{R}) \in (\Delta_0(\mathcal{R}), \Delta_1(\mathcal{R}))$.

By induction, it is easy to see that

$$\Delta_{2k} \in (\Delta_{2k-2}(\mathcal{R}), \Delta_{2k-1}(\mathcal{R})), \quad \Delta_{2k+1} \in (\Delta_{2k}(\mathcal{R}), \Delta_{2k-1}(\mathcal{R})) \quad \forall k \geq 0. \tag{15}$$

Notice how this implies that the sequences $\{\Delta_{2k}(\mathcal{R})\}_{k=0}^{\infty}$ is increasing, while $\{\Delta_{2k+1}(\mathcal{R})\}_{k=0}^{\infty}$ is decreasing.

Moreover, Equation 15 implies that

$$\Delta_{2k+1}(\mathcal{R}) - \Delta_{2k}(\mathcal{R}) < \Delta_{2k-1}(\mathcal{R}) - \Delta_{2k-2}(\mathcal{R}).$$

Thus, $\{\Delta_{2k+1}(\mathcal{R}) - \Delta_{2k}(\mathcal{R})\}_{k=0}^{\infty} \searrow 0$, proving that reviews converge to $\Delta_\infty(\mathcal{R})$, which is defined by the unique solution of Equation 6. Because the right hand side of $\mathbb{E}_{t+1}(\mathcal{R}_1)$ and $\mathbb{E}_{t+1}(\mathcal{R}_2)$ solely depends on $\Delta_t(\mathcal{R})$, by continuity this implies that $\mathbb{E}_t(\mathcal{R}_1)$ and $\mathbb{E}_t(\mathcal{R}_2)$ also converge.

Denote $\mathbb{E}_\infty(\mathcal{R}_1) - \mathbb{E}_\infty(\mathcal{R}_2)$ by $\Delta(\mathcal{R})$. Notice that, by definition of $\mathcal{B}_\infty$, the claim is equivalent to $\mathcal{B}_\infty > -(Q_1 - Q_2)$.

To show that this is indeed the case, suppose by contradiction that $\mathbb{E}_\infty(\mathcal{R}_1) - \mathbb{E}_\infty(\mathcal{R}_2) < 0$ and, thus, $\mathcal{B}_\infty < -(Q_1 - Q_2)(< 0)$.

But then, we have

$$\begin{aligned}
\mathbb{E}_\infty(\mathcal{R}_1) - \mathbb{E}_\infty(\mathcal{R}_2) = (Q_1 - Q_2) &+ \mathbb{E}(\theta_{1j} \mid \mathbb{E}_\infty(\mathcal{R}_1) + \theta_{1j} > \mathbb{E}_\infty(\mathcal{R}_2) + \theta_{2j})) \\
&- \mathbb{E}(\theta_{2j} \mid \mathbb{E}_\infty(\mathcal{R}_2) + \theta_{2j} > \mathbb{E}_\infty(\mathcal{R}_1) + \theta_{1j})).
\end{aligned}$$

which, using the definition of

$$\mathcal{B}_\infty = (\mathbb{E}_\infty(\mathcal{R}_1) - \mathbb{E}_\infty(\mathcal{R}_2)) - (Q_1 - Q_2),$$

simplifies to

$$\begin{aligned}\mathcal{B}_\infty &= \mathbb{E}(\theta_{1j} \mid \mathbb{E}_\infty(\mathcal{R}_1) + \theta_{1j} > \mathbb{E}_\infty(\mathcal{R}_2) + \theta_{2j})) \\ &\quad - \mathbb{E}(\theta_{2j} \mid \mathbb{E}_\infty(\mathcal{R}_2) + \theta_{2j} > \mathbb{E}_\infty(\mathcal{R}_1) + \theta_{1j})).\end{aligned} \tag{16}$$

Since the LHS is negative, it is enough to show that the RHS is positive to reach a contradiction. To see that this is indeed the case, notice that

$$\begin{aligned}&\mathbb{E}(\theta_{1j} \mid \theta_{1j} > \theta_{2j} + \mathbb{E}_\infty(\mathcal{R}_2) - \mathbb{E}_\infty(\mathcal{R}_1)) \\ &- \mathbb{E}(\theta_{2j} \mid \theta_{2j} > \theta_{1j} + \mathbb{E}_\infty(\mathcal{R}_1) - \mathbb{E}_\infty(\mathcal{R}_2)) \\ &= \mathbb{E}(\theta_{1j} \mid \theta_{1j} > \theta_{2j} - \Delta(\mathcal{R}))) \\ &- \mathbb{E}(\theta_{2j} \mid \theta_{2j} > \theta_{1j} + \Delta(\mathcal{R}))) \\ &> 0\end{aligned} \tag{17}$$

Where the inequality follows from $s_1 = s_2$ ($\in \{s_L,\ s_H\}$) and the fact that $\Delta_\infty(\mathcal{R}) < 0$ by assumption. Thus, we have that $\mathbb{E}_\infty(\mathcal{R}_1)) \geq \mathbb{E}_\infty(\mathcal{R}_2))$.

To rule out equality, notice that if $\mathbb{E}_\infty(\mathcal{R}_1)) = \mathbb{E}_\infty(\mathcal{R}_2))$ the RHS is Eq. (16) is 0 by symmetry of designs, while the LHS is negative since $\mathcal{B}_\infty = -Q_1 + Q_2 < 0$. We have thus proved that $\mathbb{E}_\infty(\mathcal{R}_1) > \mathbb{E}_\infty(\mathcal{R}_2)$. ■

Claim 2: Despite being relatively underrated: $\mathcal{B}_\infty < 0$.

Proof of Claim 2:

Subtracting the second line in Eq. 6 from the first – just like we did in the Proof of Proposition 1 – we get:

$$\begin{aligned}\mathbb{E}_\infty(\mathcal{R}_1) - \mathbb{E}_\infty(\mathcal{R}_2) = (Q_1 - Q_2) &+ \mathbb{E}(\theta_{1j} \mid \mathbb{E}_\infty(\mathcal{R}_1) + \theta_{1j} > \mathbb{E}_\infty(\mathcal{R}_2) + \theta_{2j})) \\ &- \mathbb{E}(\theta_{2j} \mid \mathbb{E}_\infty(\mathcal{R}_2) + \theta_{2j} > \mathbb{E}_\infty(\mathcal{R}_1) + \theta_{1j})).\end{aligned}$$

Using the definition of

$$\mathcal{B}_\infty = (\mathbb{E}_\infty(\mathcal{R}_1) - \mathbb{E}_\infty(\mathcal{R}_2)) - (Q_1 - Q_2),$$

we can simplify this expression to

$$\begin{aligned}\mathcal{B}_\infty &= \mathbb{E}(\theta_{1j} \mid \mathbb{E}_\infty(\mathcal{R}_1) + \theta_{1j} > \mathbb{E}_\infty(\mathcal{R}_2) + \theta_{2j}) \\ &\quad - \mathbb{E}(\theta_{2j} \mid \mathbb{E}_\infty(\mathcal{R}_2) + \theta_{2j} > \mathbb{E}_\infty(\mathcal{R}_1) + \theta_{1j}).\end{aligned} \tag{18}$$

Now assume $\mathcal{B}_\infty = 0$. Then, Eq. (18) becomes

$$\mathbb{E}(\theta_{1j} \mid Q_1 + \theta_{1j} > Q_2 + \theta_{2j})) = \mathbb{E}(\theta_{2j} \mid Q_2 + \theta_{2j} > Q_1 + \theta_{1j})). \tag{19}$$

Denoting by $\Delta(Q) = Q_1 - Q_2$, and noticing that $\Delta(Q) > 0$ by assumption, this is equivalent to

$$\mathbb{E}(\theta_{1j} \mid \theta_{1j} > \theta_{2j} - \Delta(Q))) = \mathbb{E}(\theta_{2j} \mid \theta_{2j} > \theta_{1j} + \Delta(Q)))). \tag{20}$$

But this is a contradiction, since:

$$\begin{aligned}\mathbb{E}(\theta_{1j} \mid \theta_{1j} > \theta_{2j} - \Delta(Q))) &= \mathbb{E}(\theta_{2j} \mid \theta_{2j} > \theta_{1j} - \Delta(Q))) \\ &< \mathbb{E}(\theta_{2j} \mid \theta_{2j} > \theta_{1j})) \\ &< \mathbb{E}(\theta_{2j} \mid \theta_{2j} > \theta_{1j} + \Delta(Q)))).\end{aligned} \tag{21}$$

Where the equality follows from $s_1 = s_2 (\in \{s_L,\ s_H\})$ and the two inequalities follow from progressively increasing the lower bound of integration in the conditional expected value of θ_{2j}.

The case $\mathcal{B}_\infty > 0$ can be handled similarly. Therefore, in equilibrium we have $\mathcal{B}_\infty < 0$. ■

Claim 3: It thus obtains a higher number of reviews: $\mathcal{N}_\infty(\mathcal{R}_1) > 1/2$, but less than it would if reviews were unbiased.

Proof of Claim 3: The Proof is a natural consequence of Claim 1 and 2. We have that

$$\begin{aligned}\mathcal{N}_\infty(\mathcal{R}_1) &= Prob(\theta_{1j} + \mathbb{E}_\infty(\mathcal{R}_1) > \theta_{2j} + \mathbb{E}_\infty(\mathcal{R}_2)) \\ &> Prob(\theta_{1j} > \theta_{2j}) \\ &= \frac{1}{2},\end{aligned}$$

where the inequality follows from the fact that $\mathbb{E}_\infty(\mathcal{R}_1) > \mathbb{E}_\infty(\mathcal{R}_2)$ as shown in Claim 1, and the last equality follows from symmetry in designs. ■

Claim 4: Nevertheless, some self-correction occurs, and both biases are less severe than in the short-run: $\mathcal{B}_1 < \mathcal{B}_\infty < 0$, $\mathcal{N}_\infty(\mathcal{R}_1) > \mathcal{N}_2(\mathcal{R}_1)$.

Proof of Claim 4:

We have already shown that $\mathcal{B}_0 = 0$, that is, $\mathbb{E}_0(\mathcal{R}_1) - \mathbb{E}_0(\mathcal{R}_2) = Q_1 - Q_2 > 0$. Thus, we have

$$\mathbb{E}_1(\mathcal{R}_1) = Q_1 + \mathbb{E}(\theta_1 \mid \theta_1 + \mathbb{E}_0(\mathcal{R}_1)) > \theta_2 + \mathbb{E}_0(\mathcal{R}_2)) = Q_1 + \mathbb{E}(\theta_1 \mid \theta_1 > \theta_2 - \Delta(Q))$$

and similarly

$$\mathbb{E}_1(\mathcal{R}_2) = Q_2 + \mathbb{E}(\theta_2 \mid \theta_2 + \mathbb{E}_0(\mathcal{R}_2) > \theta_1 + \mathbb{E}_0(\mathcal{R}_1)) = Q_2 + \mathbb{E}(\theta_2 \mid \theta_2 > \theta_1 + \Delta(Q))).$$

Jointly, these two imply that

$$\begin{aligned}\mathcal{B}_1 &= (Q_1 + \mathbb{E}(\theta_1 \mid \theta_1 > \theta_2 - \Delta(Q))) - (Q_2 + \mathbb{E}(\theta_2 \mid \theta_2 > \theta_1 + \Delta(Q))) - Q_1 + Q_2 \\ &= \mathbb{E}(\theta_1 \mid \theta_1 > \theta_2 - \Delta(Q))) - \mathbb{E}(\theta_2 \mid \theta_2 > \theta_1 + \Delta(Q))).\end{aligned} \tag{22}$$

But then, comparing the expressions in Eq. (18) and (54), we obtain $\mathcal{B}_1 < \mathcal{B}_\infty$ if and only if $\Delta(Q) > \mathbb{E}_\infty(\mathcal{R}_1) - \mathbb{E}_\infty(\mathcal{R}_2)$. But this is equivalent to $\mathcal{B}_\infty < 0$, which we have shown to be true in the Proof of **Claim 1**.

As in the previous cases, the fact that $\mathcal{N}_2(\mathcal{R}_1) < \mathcal{N}_\infty(\mathcal{R}_1)$ follows straightforwardly from $\mathcal{B}_1 < \mathcal{B}_\infty$. ■

Claim 5: Despite these distortions, reviews unambiguously increase consumer welfare.

Proof of Claim 5:

To show that reviews improve consumer welfare, define

$$\mathcal{J}_{BR} := \{j \mid \theta_{1j} > \theta_{2j} - \Delta(Q) \quad \& \quad \theta_{1j} < \theta_{2j} - \Delta(R)\}$$

and

$$\mathcal{J}_{NR} = \{j \mid \theta_{1j} > \theta_{2j} - \Delta(Q) \quad \& \quad \theta_{1j} < \theta_{2j}\}.$$

In words, $\mathcal{J}_{BR}$ is the set of consumers who would prefer product 1 but end up choosing their *subjectively* less preferred option given the presence of biased reviews (BR). That is, every consumer $j \in \mathcal{J}_{BR}$ prefers product 1 in light of the quality difference between the two product and their taste for each of the two products, but end up choosing product 2 because they are misled by reviews. Noice that $\Delta_\infty(R) < \Delta(Q)$ – or, equivalently, $\mathcal{B}_\infty$, which we have shown in **Claim 2** – implies this set is non-empty.

Similarly, $\mathcal{J}_{NR}$ is the set of consumers who would prefer product 1 but choose their *subjectively* less preferred option given no reviews (NR): these consumers choose purely based on taste, without accounting for the fact that $Q_1 > Q_2$.

We want to show that $\mathcal{J}_{BR} \subset \mathcal{J}_{NR}$. But this is immediate, since

$$\theta_{1j} < \theta_{2j} - \Delta(R) \quad \Rightarrow \quad \theta_{1j} < \theta_{2j}$$

given that $\Delta(R) > 0$ as we have shown in **Claim 1**. Therefore, $Prob(j \in \mathcal{J}_{NR}) > Prob(j \in \mathcal{J}_{BR})$.

It is immediate to realize that, in this context, welfare is proportional to the number of consumers who choose their subjectively optimal product. We have that this probability is 1 in the first best case, $1 - Prob(j \in \mathcal{J}_{BR})$ in the biased reviews case, and $1 - Prob(j \in \mathcal{J}_{NR})$ in the no reviews case.

Clearly, the above reasoning implies that

$$1 - Prob(j \in \mathcal{J}_{NR}) < 1 - Prob(j \in \mathcal{J}_{BR}) < 1, \tag{23}$$

which concludes the proof. ■

Proof of Proposition 3

Claim 1: Increasing the share of informed consumers, $1-\alpha$, worsens the amount of equilibrium bias, $\mathcal{B}_\infty(\alpha)$, in Propositions 1 and 2

Proof of Claim 1:

We have seen that in this case, Eq. (6) becomes

$$\begin{cases} \mathbb{E}_\infty(\mathcal{R}_1) = Q_1 + \alpha \cdot \mathbb{E}\big(\theta_{1j} \mid \mathbb{E}_\infty(\mathcal{R}_1) + \theta_{1j} \geq \mathbb{E}_\infty(\mathcal{R}_2) + \theta_{2j}\big) \\ \qquad\qquad\qquad +(1-\alpha) \cdot \mathbb{E}(\theta_{1j} \mid Q_1 + \theta_{1j} \geq Q_2 + \theta_{2j}) \\ \mathbb{E}_\infty(\mathcal{R}_2) = Q_2 + \alpha \cdot \mathbb{E}\big(\theta_{2j} \mid \mathbb{E}_\infty(\mathcal{R}_2) + \theta_{2j} \geq \mathbb{E}_\infty(\mathcal{R}_1) + \theta_{1j}\big) \\ \qquad\qquad\qquad +(1-\alpha) \cdot \mathbb{E}(\theta_{2j} \mid Q_2 + \theta_{2j} \geq Q_1 + \theta_{1j}\big) \end{cases} \tag{24}$$

Denote by $\mathcal{B}_\infty(\alpha)$ the amount of bias in reviews as a function of the fraction of naïve consumers.

The proof proceeds in two step. First, we show that $\mathcal{B}_\infty(0) < \mathcal{B}_\infty(1)$. Then, we show that $\mathcal{B}_\infty(\alpha)$ is monotonic in $[0,\ 1]$. Each of these two steps must be performed for both the case in Proposition 1 and that in Proposition 2.

Case 1: Extension of Proposition 1

Subtracting the second Eq. from the first in (25) and noting that here, $Q_1 = Q_2$ (and, thus, $B_\infty = \mathbb{E}_\infty(\mathcal{R}_1) - \mathbb{E}_\infty(\mathcal{R}_2) < 0$), we get:

$$\begin{aligned} \mathcal{B}_\infty = {} & \alpha \cdot \mathbb{E}\big(\theta_{1j} \mid \theta_{1j} \geq \theta_{2j} - \mathcal{B}_\infty\big) \\ & - \alpha \cdot \mathbb{E}(\theta_{2j} \mid \theta_{2j} \geq \theta_{1j} + \mathcal{B}_\infty\big) \\ & + (1-\alpha) \cdot \mathbb{E}(\theta_{1j} \mid \theta_{1j} \geq \theta_{2j}) \\ & - (1-\alpha) \cdot \mathbb{E}(\theta_{2j} \mid \theta_{2j} \geq \theta_{1j}). \end{aligned} \tag{25}$$

First, we want to show that $\mathcal{B}_\infty(1) < \mathcal{B}_\infty(0)$. We have

$$\begin{aligned} \mathcal{B}_\infty(1) = {} & \mathbb{E}\big(\theta_{1j} \mid \theta_{1j} \geq \theta_{2j} - \mathcal{B}_\infty(1)\big) \\ & - \mathbb{E}(\theta_{2j} \mid \theta_{2j} \geq \theta_{1j} + \mathcal{B}_\infty(1)\big). \end{aligned}$$

and

$$\begin{aligned} \mathcal{B}_\infty(0) = {} & \mathbb{E}(\theta_{1j} \mid \theta_{1j} \geq \theta_{2j}) \\ & - \mathbb{E}(\theta_{2j} \mid \theta_{2j} \geq \theta_{1j}). \end{aligned}$$

Recall that $\mathcal{B}_\infty(1) < 0$ (Proposition 1).

But then,

$$\begin{aligned} \mathcal{B}_\infty(0) &= \mathbb{E}(\theta_{1j} \mid \theta_{1j} \geq \theta_{2j}) - \mathbb{E}(\theta_{2j} \mid \theta_{2j} \geq \theta_{1j}) \\ &\leq \mathbb{E}(\theta_{1j} \mid \theta_{1j} \geq \theta_{2j} - \mathcal{B}_\infty(1)) - \mathbb{E}(\theta_{2j} \mid \theta_{2j} \geq \theta_{1j} + \mathcal{B}_\infty(1)) \\ &= \mathcal{B}_\infty(1). \end{aligned}$$

Thus, $\mathcal{B}_\infty(0) < \mathcal{B}_\infty(1) < 0$: the equilibrium bias got *worse* if only Bayesian are present ($\alpha = 0$), compared to only naïves ($\alpha = 1$).

To show that we have monotonicity in $\alpha \in [0,\ 1]$, denote by

$$\begin{aligned} G(\alpha, \mathcal{B}) = \mathcal{B} &- \alpha \cdot \mathbb{E}(\theta_{1j} \mid \theta_{1j} \geq \theta_{2j} - \mathcal{B}) \\ &+ \alpha \cdot \mathbb{E}(\theta_{2j} \mid \theta_{2j} \geq \theta_{1j} + \mathcal{B}) \\ &- (1-\alpha) \cdot \mathbb{E}(\theta_{1j} \mid \theta_{1j} \geq \theta_{2j}\big) \\ &+ (1-\alpha) \cdot \mathbb{E}(\theta_{2j} \mid \theta_{2j} \geq \theta_{1j}). \end{aligned} \tag{26}$$

Then, for every $\alpha \in [0, 1]$, $\mathcal{B}_\infty(\alpha)$ solves $G(\alpha, \mathcal{B}) = 0$. By the Implicit Function Theorem, we have that

$$\frac{\partial \mathcal{B}(\alpha)}{\partial \alpha} = -\frac{\frac{\partial G(\alpha,\mathcal{B})}{\partial \alpha}}{\frac{\partial G(\alpha,\mathcal{B})}{\partial \mathcal{B}}} \tag{27}$$

The numerator is given by

$$\begin{aligned} \frac{\partial G(\alpha, \mathcal{B})}{\partial \alpha} = &- \mathbb{E}\big(\theta_{1j} \mid \theta_{1j} \geq \theta_{2j} - \mathcal{B}\big) \\ &+ \mathbb{E}(\theta_{2j} \mid \theta_{2j} \geq \theta_{1j} + \mathcal{B}) + \mathbb{E}(\theta_{1j} \mid \theta_{1j} \geq \theta_{2j}) - \mathbb{E}(\theta_{2j} \mid \theta_{2j} \geq \theta_{1j}), \end{aligned}$$

which is negative, because $-\mathbb{E}\big(\theta_{1j} \mid \theta_{1j} \geq \theta_{2j} - \mathcal{B}) + \mathbb{E}(\theta_{1j} \mid \theta_{1j} \geq \theta_{2j})$ is since $\mathcal{B} < 0$, and so is $\mathbb{E}(\theta_{2j} \mid \theta_{2j} \geq \theta_{1j} + \mathcal{B}) - \mathbb{E}((\theta_{2j} \mid \theta_{2j} \geq \theta_{1j}))$.

The denominator is given by

$$\frac{\partial G(\alpha, \mathcal{B})}{\partial \mathcal{B}} = 1 - \alpha \cdot \frac{\mathbb{E}\big(\theta_{1j} \mid \theta_{1j} \geq \theta_{2j} - \mathcal{B}\big)}{\partial \mathcal{B}} + \alpha \cdot \frac{\mathbb{E}\big(\theta_{2j} \mid \theta_{2j} \geq \theta_{1j} + \mathcal{B}\big)}{\partial \mathcal{B}}$$

which is positive because the first and third term also are, while the second one is negative.

So, overall we have that $\mathcal{B}_\infty(\alpha)$ is increasing for $\alpha \in [0,\ 1]$. An increase in the share of Bayesian consumers make the reviews more biased, that is, decreases $\mathcal{B}_\infty(\alpha)$ further away from 0.

Case 2: Extension of Proposition 2

The proof follows very similar steps to that of **Case 1: Extension of Proposition 1**. Nevertheless, there are some differences, so we also report this one in its entirety.

Subtracting the second Eq. from the first in (25), we get:

$$\begin{aligned} \mathcal{B}_\infty = \alpha \cdot \mathbb{E}\big(\theta_{1j} \mid \theta_{1j} \geq \theta_{2j} - \mathbb{E}(\mathcal{R}_1) + \mathbb{E}(\mathcal{R}_2) \\ - \alpha \cdot \mathbb{E}\big(\theta_{2j} \mid \theta_{2j} \geq \theta_{1j} + \mathbb{E}(\mathcal{R}_1) - \mathbb{E}(\mathcal{R}_2) \\ + (1-\alpha) \cdot \mathbb{E}(\theta_{1j} \mid \theta_{1j} \geq \theta_{2j} - Q_1 + Q_2) \\ - (1-\alpha) \cdot \mathbb{E}(\theta_{2j} \mid \theta_{2j} \geq \theta_{1j} + Q_1 - Q_2). \end{aligned} \tag{28}$$

First, we want to show that $\mathcal{B}_\infty(1) < \mathcal{B}_\infty(0)$. We have

$$\begin{aligned} \mathcal{B}_\infty(1) = \mathbb{E}\big(\theta_{1j} \mid \theta_{1j} \geq \theta_{2j} - \mathbb{E}(\mathcal{R}_1) + \mathbb{E}(\mathcal{R}_2)) \\ - \mathbb{E}(\theta_{2j} \mid \theta_{2j} \geq \theta_{1j} + \mathbb{E}(\mathcal{R}_1) - \mathbb{E}(\mathcal{R}_2)). \end{aligned}$$

and

$$\mathcal{B}_\infty(0) = \mathbb{E}(\theta_{1j} \mid \theta_{1j} \geq \theta_{2j} - Q_1 + Q_2) \\ - \mathbb{E}(\theta_{2j} \mid \theta_{2j} \geq \theta_{1j} + Q_1 - Q_2).$$

First notice that Proposition 2 implies that $\mathcal{B}_\infty(1) < 0$.
But then,

$$\begin{aligned}\mathcal{B}_\infty(0) &= \mathbb{E}(\theta_{1j} \mid \theta_{1j} \geq \theta_{2j} - Q_1 + Q_2) \\ &- \mathbb{E}(\theta_{2j} \mid \theta_{2j} \geq \theta_{1j} + Q_1 - Q_2) \\ &< \mathbb{E}(\theta_{1j} \mid \theta_{1j} \geq \theta_{2j} - \mathbb{E}(\mathcal{R}_1) + \mathbb{E}(\mathcal{R}_2)) \\ &- \mathbb{E}(\theta_{2j} \mid \theta_{2j} \geq \theta_{1j} + \mathbb{E}(\mathcal{R}_1) - \mathbb{E}(\mathcal{R}_2)) \\ &= \mathcal{B}_\infty(1).\end{aligned}$$

where the inequality uses the fact that $\mathbb{E}(\mathcal{R}_1) - \mathbb{E}(\mathcal{R}_2)) - Q_1 + Q_2 = \mathcal{B}_\infty(1)$ is negative. Thus, $\mathcal{B}_\infty(0) < \mathcal{B}_\infty(1) < 0$: the equilibrium bias got *worse* if only Bayesian are present, compared to only naïves.

To show that we have monotonicity in $\alpha \in [0,\ 1]$, denote by

$$\begin{aligned}G(\alpha, \mathcal{B}) = \mathcal{B} &- \alpha \cdot \mathbb{E}\big(\theta_{1j} \mid \theta_{1j} \geq \theta_{2j} - (\mathcal{B} + Q_1 - Q_2) \\ &+ \alpha \cdot \mathbb{E}\big(\theta_{2j} \mid \theta_{2j} \geq \theta_{1j} + \mathcal{B} + Q_1 - Q_2) \\ &+ (1-\alpha) \cdot \mathbb{E}(\theta_{1j} \mid \theta_{1j} \geq \theta_{2j} - Q_1 + Q_2) \\ &- (1-\alpha) \cdot \mathbb{E}(\theta_{2j} \mid \theta_{2j} \geq \theta_{1j} + Q_1 - Q_2).\end{aligned} \tag{29}$$

where we have substituted $\mathbb{E}(\mathcal{R}_1) - \mathbb{E}(\mathcal{R}_2)) = \mathcal{B} + Q_1 - Q_2$.

Then, for every $\alpha \in [0, 1]$, $\mathcal{B}_\infty(\alpha)$ solves $G(\alpha, \mathcal{B}) = 0$. By the Implicit Function Theorem, we have that

$$\frac{\partial \mathcal{B}(\alpha)}{\partial \alpha} = -\frac{\frac{\partial G(\alpha,\mathcal{B})}{\partial \alpha}}{\frac{\partial G(\alpha,\mathcal{B})}{\partial \mathcal{B}}} \tag{30}$$

The numerator is given by

$$\begin{aligned}\frac{\partial G(\alpha, \mathcal{B})}{\partial \alpha} = &- \mathbb{E}\big(\theta_{1j} \mid \theta_{1j} \geq \theta_{2j} - (\mathcal{B} + Q_1 - Q_2) \\ &- \mathbb{E}(\theta_{2j} \mid \theta_{2j} \geq \theta_{1j} + \mathcal{B} + Q_1 - Q_2) \\ &+ \mathbb{E}(\theta_{1j} \mid \theta_{1j} \geq \theta_{2j} - Q_1 + Q_2) \\ &- \mathbb{E}(\theta_{2j} \mid \theta_{2j} \geq \theta_{1j} + Q_1 - Q_2),\end{aligned}$$

which is negative, because the difference between its 1st and 3rd terms is, and the same is true for the 2nd and 4th.

The denominator is given by

$$\frac{\partial G(\alpha, \mathcal{B})}{\partial \mathcal{B}} = 1 - \alpha \cdot \frac{\partial \mathbb{E}\big(\theta_{1j} \mid \theta_{1j} \geq \theta_{2j} - \mathcal{B} - Q_1 + Q_2)}{\partial \mathcal{B}} + \alpha \cdot \frac{\partial \mathbb{E}\big(\theta_{2j} \mid \theta_{2j} \geq \theta_{1j} + \mathcal{B} + Q_1 - Q_2)}{\partial \mathcal{B}}$$

which is positive.

So, overall we have that $\mathcal{B}_\infty(\alpha)$ is decreasing for $\alpha \in [0,\ 1]$, as desired. An increase in the share of Bayesian consumers worsens the bias in reviews, that is, decreases $\mathcal{B}_\infty(\alpha)$ further away from 0. ■

Claim 2: Thus making naïve consumers strictly worse off.

Proof of Claim 2: Our welfare results in Proposition 5 show that welfare and bias go hand in hand: the more biased the reviews, the larger the welfare losses for naïve consumers. This result follows straightforwardly. ■

Proof of Proposition 4

We want to show that

$$Var\big(\theta_L \mid \theta_L > \theta_H + \Delta(Q)\big) < Var\big(\theta_H \mid \theta_H > \theta_H - \Delta(Q)\big), \quad \forall \Delta(Q) > \Delta^*(Q).$$

First notice that, when $\Delta(Q)$ approaches $\bar{\theta} - \underline{\theta}$, we have

$$Var\big(\theta_H \mid \theta_H > \theta_H - \Delta(Q)\big) \to Var(\theta_H).$$

On the other hand, the fact that $\theta_L > \theta_H + \Delta(Q)$ implies $\theta_L \in (\underline{\theta} + \Delta(Q),\ \bar{\theta})$. Therefore, Popoviciu's Inequality (Popoviciu, 1935) implies that

$$Var\big(\theta_L \mid \theta_L > \theta_H + \Delta(Q)\big) \leq \frac{1}{4}\big(\bar{\theta} - \underline{\theta} - \Delta(Q)\big)^2.$$

Notice that the right hand side gets arbitrarily small as $\Delta(Q) \to \bar{\theta} - \underline{\theta}$, implying the existence of a $\Delta^*(Q)$ such that $Var\big(\theta_L | \theta_L > \theta_H + \Delta(Q)\big) < Var(\theta_H)$ for every $\Delta(Q) > \Delta^*(Q)$. ■

B. Extensions

B.1. Duopoly with Outside Option

Assume now that consumers are not only decide *what*, but also *if* to buy. That is, they also have an outside option, of quality c. Without loss of generality, we assume that the outside option is non trivial, that is, it is chosen by at least some consumers. In particular, given that each consumer's taste shocks for the two products are independent, this is equivalent to requiring that a consumer with the lowest possible taste for both product 1 and product 2 would choose the outside option instead. In other words, this requires

$$c \geq \max(Q_1 + \underline{\theta}_1, Q_2 + \underline{\theta}_2).$$

For example, when $s_1 = s_2$ and $Q_1 > Q_2$, this amounts to requiring that $c > Q_2 + \underline{\theta}$. This condition is trivially satisfied for unbounded distribution (that is, in our leading example, the outside option is chosen by some consumers, even for very low values of c).

As before, upon choosing a product, each buyer reviews it honestly, but *subjectively*, by reporting their own experienced utility. Thus, in this case, Eq. (3) becomes

$$\mathcal{R}_{ij} := \begin{cases} \mathcal{U}_{ij} = Q_i + \theta_{ij} - P_i & \text{if } \mathbb{E}(\mathcal{U}_{ij}) \geq \max(\mathbb{E}(\mathcal{U}_{-ij}), c), \\ \varnothing & \text{otherwise.} \end{cases}$$

Denote by $\mathcal{J}_1^c$ and $\mathcal{J}_2^c$ the sets of buyers of product 1 and 2 respectively (we are omitting the t subscript for notational simplicity). That is,

$$\mathcal{J}_1^c = \{j \in \mathcal{J} \mid \mathbb{E}(Q_1) + \theta_{1j} - P_1 \geq \max(\mathbb{E}(Q_2) + \theta_{2j} - P_2, c)\},$$

and similarly for $\mathcal{J}_2^c$. It is immediate to see that, given the $\mathcal{J}_1$ and $\mathcal{J}_2$ defined in Eq. (4), we have $\mathcal{J}_1^c \subseteq \mathcal{J}_1$, $\mathcal{J}_2^c \subseteq \mathcal{J}_2$, with the inclusion being strict whenever the value of the outside option, c, is non-trivial. Moreover, for $c' > c$, we have $\mathcal{J}_1^{c'} \subseteq \mathcal{J}_1^c$ and $\mathcal{J}_2^{c'} \subseteq \mathcal{J}_2^c$. These simple observations form the basis for our first result: the presence of an outside option increases the reviews of each product, and decreases the number of reviews for each product.

We are interested in studying the robustness of our main results – in particular, Proposition 1, 2 and 4 – to the inclusion of an outside option.

We start from Proposition 1. We have the following result:

Proposition 5 (More Polarizing Products Are Relatively Overrated) *Let the quality of the outside option be c, and let the two products differ only in their design:* $Q_1 = Q_2$, $s_1 = H$, $s_2 = L$. *Assume that* s_1 *and* s_2 *are symmetric.*[39] *Then, in the long-run* ($t = \infty$)*:*

39. Symmetry is a sufficient, but not necessary, condition.

- *Both products' reviews are higher than they would be absent an outside option, and increasing in the outside option quality, c.*
- *The more polarizing product 2 is relatively overrated:* $\mathcal{B}_\infty < 0$*,*
- *and thus captures a higher number of reviews:* $\mathcal{N}_\infty(\mathcal{R}_2) > 1/2$*.*
- *Nevertheless, some self-correction occurs, and both biases are less severe than in the short-run:* $\mathcal{B}_0 < \mathcal{B}_\infty < 0$*,* $\mathcal{N}_1(\mathcal{R}_2) > \mathcal{N}_\infty(\mathcal{R}_2) > 1/2$*.*

Proposition 6 (High Quality Products Are Relatively Underrated.) *Let the quality of the outside option be c, and let the two products differ only in their qualities:* $Q_1 > Q_2$*,* $s_1 = s_2$*. Then, in equilibrium (*$t = \infty$*):*

- *The higher quality product 1 has higher reviews:* $\mathbb{E}_\infty(\mathcal{R}_1) > \mathbb{E}_\infty(\mathcal{R}_2)$*,*
- *Despite being relatively underrated:* $\mathcal{B}_\infty < 0$*.*
- *It thus obtains a higher number of reviews:* $\mathcal{N}_\infty(\mathcal{R}_1) > 1/2$*, but less than it would if consumers were fully informed.*
- *Nevertheless, some self-correction occurs, and both biases are less severe than in the short-run:* $\mathcal{B}_1 < \mathcal{B}_\infty < 0$*,* $\mathcal{N}_\infty(\mathcal{R}_1) > \mathcal{N}_2(\mathcal{R}_1)$*.*
- *Despite these distortions, reviews unambiguously increase consumer welfare.*

Proposition 7 (The Variance of Reviews Needs Not Proxy Product Design) *Let the quality of the outside option be c, and let the support for* θ_1 *and* θ_2 *be bounded above. Then, there exists a quality gap* $Q := Q_1 - Q_2 > 0$ *such that* $Var_\infty(\mathcal{R}_1) > Var_\infty(\mathcal{R}_2)$ *for all* s_1 *and* s_2*.*

Now that we have established robustness of our main results to the introduction of an outside option, we turn to a different question: quantitatively, does the outside option mitigate or worsen the biases? We show, directly, that the answer to this question depends on the outside option's quality, c. That is, one can find examples such that an outside option of quality c increases $\mathcal{B}_\infty$, while an outside option of quality c' decreases $\mathcal{B}_\infty$.

One such example can be found in Figure 4. Here, we can clearly see that low quality outside options worsen the bias in reviews, while higher quality ones mitigate (but fail to erase) it. Of course, this is a byproduct of the normal distribution. It is plausible that other distributions would cause the dependence of $\mathcal{B}^c_\infty$ on c to change. However, what the last three Propositions show is that the qualitative nature of our main results is robust to the introduction of an outside option.

B.2. Learning from Cumulative Reviews

Throughout our paper, period t consumers learn from the reviews of their predecessors, generation $t-1$. We now turn to studying the (empirically realistic) case of consumers learning from a (possibly weighted) average of reviews up to time t. To this end, we first formalize the average of cumulative reviews as follows. For $\beta \in [0,1]$, denote by $\mathbb{E}_t(\mathcal{R}^c)$ the average of cumulative reviews at time t:

$$\mathbb{E}_t(\mathcal{R}^C) = \frac{\sum_{\tau=0}^{\tau=t} \beta^{t-\tau} \mathcal{N}_\tau(\mathcal{R}) \mathbb{E}_\tau(\mathcal{R})}{\sum_{\tau=0}^{\tau=t} \beta^{t-\tau} \mathcal{N}_\tau(\mathcal{R})}. \tag{31}$$

As the formula makes apparent, $\mathbb{E}_t(\mathcal{R}^C)$ depends on both the average and the number of reviews that the product received in each period $\tau = 0, 1, ..., t$. The (backward) discount factor $\beta^{t-\tau}$ measures how much the platform underweights past reviews compared to more recent ones.[40] If $\beta = 0$, $\mathbb{E}_t(\mathcal{R}^C) = \mathbb{E}_t(\mathcal{R})$, and we recover the case studied throughout this paper. Conversely, if $\beta = 1$,

$$\mathbb{E}_t(\mathcal{R}^C) = \frac{\sum_{\tau=0}^{\tau=t} \mathcal{N}_\tau(\mathcal{R}) \mathbb{E}_\tau(\mathcal{R})}{\sum_{\tau=0}^{\tau=t} \mathcal{N}_\tau(\mathcal{R})},$$

so that $\mathbb{E}_t(\mathcal{R}^C)$ is simply the weighted average of all reviews, with the weights being given solely by the number of reviews in each period. Building on this definition, we also define the cumulative reviews advantage of product 1 as $\Delta_t^C(\mathcal{R}) = \mathbb{E}_t(\mathcal{R}_1^C) - \mathbb{E}_t(\mathcal{R}_2^C)$.

Clearly, our definition of cumulative reviews implies that, for every β, $\mathbb{E}_0(\mathcal{R}^C) = \mathbb{E}_0(\mathcal{R})$. The two definitions start to differ in period 1. For example, if reviews for product i decrease in period 1, $(\mathbb{E}_1(\mathcal{R}_i) < \mathbb{E}_0(\mathcal{R}_i))$, then naturally cumulative reviews also do $(\mathbb{E}_1(\mathcal{R}_i^C) < \mathbb{E}_0(\mathcal{R}_i^C))$, but less so than period by period ones $(\mathbb{E}_1(\mathcal{R}_i^C) > \mathbb{E}_1(\mathcal{R}_i))$. As a result, period 3 consumers will form different beliefs and make different choices, resulting in different period 3 reviews and, *a fortiori*, a difference between $\mathbb{E}_2(\mathcal{R}^C)$ and $\mathbb{E}_2(\mathcal{R})$.

Therefore, the dynamics of consumers' self-selection patterns, and thus reviews, are different depending on whether $\beta = 0$ (as is the case in our main body of the paper) or $\beta > 0$. We emphasize that not only there is a difference between $\mathbb{E}_t(\mathcal{R}^C)$ and $\mathbb{E}_t(\mathcal{R})$, but also that period by period reviews are different in the two cases, since they result from different learning dynamics. Figure 5, for instance, shows that cumulative reviews are smoother than period by period ones and, as a natural consequence of this fact, that when consumers learn from cumulative reviews, their period by period reviews are also smoother.

Given the totally different learning dynamics in the two cases, it is then natural to ask whether the reviews' convergence in the long-run, as well as the long-run bias (if any), depend on β. The next Proposition offers a striking answer to this question: despite completely different learning and review dynamics, both sequences of reviews converge, and moreover, their long-run limit is unchanged. Thus, the assumption

40. Overweighting (or overemphasizing) recent reviews is a widespread practice on many online platforms. In our context, quality is fixed over time, so that, in principle, older reviews are just as informative as more recent ones. Ultimately, as we show explicitly, the results holds independently of $\beta \in [0,1]$, and therefore it is worth presenting this model extension in its most general form.

that $\beta = 0$ made throughout the paper is without loss of generality for the purpose of studying long-run biases in reviews.

We start with a straightforward Lemma:

Lemma 3. *For every $\beta \in [0,1]$ and every t, $\mathbb{E}_t(\mathcal{R}^C)$ satisfies the following recursive equation:*

$$\mathbb{E}_t(\mathcal{R}^C) = \frac{\beta \mathbb{E}_{t-1}(\mathcal{R}^C) + \mathcal{N}_t(\mathcal{R})\mathbb{E}_t(\mathcal{R})}{\beta + \mathcal{N}_t(\mathcal{R})}. \tag{32}$$

Similarly, $\Delta_t^C(\mathcal{R})$ satisfies

$$\Delta_t^C(\mathcal{R}) = \frac{\beta \mathbb{E}_{t-1}(\mathcal{R}_1^C) + \mathcal{N}_t(\mathcal{R}_1)\mathbb{E}_t(\mathcal{R}_1)}{\beta + \mathcal{N}_t(\mathcal{R}_1)} - \frac{\beta \mathbb{E}_{t-1}(\mathcal{R}_2^C) + \mathcal{N}_t(\mathcal{R}_2)\mathbb{E}_t(\mathcal{R}_2)}{\beta + \mathcal{N}_t(\mathcal{R}_2)}. \tag{33}$$

Lemma 3 follows immediately from the definition of $\mathbb{E}_t(\mathcal{R}^C)$ in Eq. (31), and, in our context, we show that it implies that $\Delta_\infty(\mathcal{R}^C)(\beta)$ converges whenever $\Delta_\infty(\mathcal{R})$ does. Moreover, the two convergence points are the same, as shown in the following Proposition.

Proposition 8. *$\Delta_t(\mathcal{R}^C)(\beta)$ converges whenever $\Delta_t(\mathcal{R})(\beta)$ does. Moreover, they converge to the same long-run outcome:*

$$\Delta_\infty(\mathcal{R}^C)(\beta) = \Delta_\infty(\mathcal{R}) \quad \forall \beta \in [0,1].$$

This result has important implications, since it shows that in the general context of our model, one can study learning from cumulative or period by period reviews interchangeably – at least when the focus in on the long-run properties of reviews. Clearly, if the goal were to study short-term changes in reviews (for example after a popularity shock like a major award, or after a seller obtains fake reviews), then the learning technology would quantitatively matter. We stress, however, that even in this case our results would qualitatively hold true. For example, in the case of fake reviews, the large pool of buyers immediately following would decrease period by period reviews and, albeit by a lesser amount, cumulative ones.

B.3. Learning About Taste

We now discuss the possibility of consumers employing reviews to simultaneously learn about both quality and fit. For instance, while some determinants of consumer-product fit can be easily observable by consumers even absent any reviews (e.g., the genre of a movie, or the cuisine of a restaurant), others might be more subtle, and thus be learned over time through reviews. For example, consumers might pick up more of a movie's characteristics over time, or get more precise information about the atmosphere of a restaurant.

As we highlight in our conclusions (Section 7), modeling social learning about the latter is not straightforward, because taste is *iid* across consumers. Nevertheless, one could think of a simple model in which, in each period, consumers' perceived taste for each product, which we denote by $\tilde{\theta}_{ij}$, is a weighted average of their actual taste θ_{ij}, plus an uninformative signal $\xi_{ij} \sim H(\cdot)$, also with mean 0, which is uncorrelated with θ_{ij}:

$$\tilde{\theta}_{ij} = \rho(t)\theta_{ij} + (1 - \rho(t))\xi_{ij}. \tag{34}$$

$\rho(t)$, the weight assigned to the informative signal, can be assumed to be increasing over time ($\frac{\partial \rho(t)}{\partial t} > 0$) and bounded above by 1 ($\lim_{t\to\infty} \rho(t) \leq 1$).

This is essentially equivalent to assuming that, in each period, each consumer observes a signal of its match for each product, $s_{ij} \sim \mathcal{N}(\theta_{ij}, 1/\rho(t))$, where $\rho(t)$ is an increasing function of t. It is also essentially equivalent to assuming that a fraction $\rho(t)$ of consumers in each period are aware of their taste for each product, while a fraction $1 - \rho(t)$ is not.

It follows from Eq. (34) that all of the conditional taste distributions governing the dynamics of our model would change to

$$\begin{aligned} &\mathbb{E}_t(\theta_{ij}|\tilde{\theta}_{ij} > \tilde{\theta}_{-ij} - \Delta_t(\mathcal{R})) \\ =&\mathbb{E}_{t+1}(\theta_{ij} \mid \rho(t)\theta_{ij} + (1 - \rho(t))\xi_{ij} > \rho(t)\theta_{-ij} + (1 - \rho(t))\xi_{-ij} - \Delta_t(\mathcal{R})). \end{aligned} \tag{35}$$

When $t = 0$, it is easy to see that Eq. (35) implies

$$\begin{aligned} &\mathbb{E}_0(\theta_{ij} \mid \tilde{\theta}_{ij} > \tilde{\theta}_{-ij} - \Delta_t(\mathcal{R})) \\ =&\mathbb{E}_0(\theta_{ij} \mid \xi_{ij} > \xi_{-ij} - \Delta_t(\mathcal{R})) = 0, \end{aligned}$$

due to the fact that ξ_{ij} is uncorrelated with actual taste, θ_{ij}. As a consequence of this fact, as Lemma 1 shows, period 0 reviews are unbiased, since they do not reflect any form of taste-based self-selection. However, as t increases, so does $\rho(t)$. If $\rho(t)$ eventually reaches 1 (say, for $t = \bar{t}$), we recover our original model for every $t \geq \bar{t}$.

None of our findings would be qualitatively affected by this (admittedly simple) modification. What the above formula for the conditional expectation of θ_{ij} implies is that the biases would get stronger over time, as consumers become more aware of their match with each product. This can be seen as a continuous equivalent (or extension) of Lemma 2, which states that absent taste-based self-selection, all biases disappear. In this slightly modified model, the stronger taste-based self-selection, the larger the biases.

B.4. Alternative Reviewing Behavior

We now briefly discuss our model's robustness to changes in its core assumptions. Particularly, we have made three key assumptions for our analysis: *i*) reviews are subjectively honest, that is, each consumer reports their subjective utility upon purchasing a product, *ii*) no self-selection at the writing stage, conditional on purchase:

everyone purchasing a product reviews it and *iii*) consumers are choosing between two options.

Inspired by both empirical realism and the sizable existing literature already presented in Section 2 (and further discussed here), we consider the following extensions.

Self-Selection Into Leaving Reviews. In our model, every consumer leaves a review upon purchasing a product. In reality, very few consumers leave reviews: a variety of surveys estimate this percentage to lie between 1% and 5%, depending on the market.

It is important to stress that, especially in a model (like ours) in which reviews are not subject to noise (see discussion in Section 4.4), this fact *per se* would be inconsequential for our findings whenever self-selection into review conditional on choice is orthogonal to the nature of the review.

However, this need not be the case. Perhaps the most common form of self-selection on writing conditional on choice documented in this context is *extremity bias* (see *e.g.* Brandes et al. (2018) and citations therein). Put simply, consumers with strong feelings towards the product – whether positive or negative – are more likely to express them compared to their peers that feel neutral towards it.

It is interesting to spell out how extremity bias would affect our results. To this end, assume that consumers in both tails (say, consumers that are either below the $10th$ percentile or above the $90th$ in their idiosyncratic taste for the product) are the only ones to leave reviews.[41] Denote by $\mathcal{J}_i^{10-}$ and $\mathcal{J}_i^{90+}$ these two camps of buyers for product i. Then, the average conditional taste for the product as reflected by reviews will be given by

$$\frac{1}{2}\mathbb{E}(\theta_{ij} \mid \theta_{ij} \in \mathcal{J}_i^{10-}) + \frac{1}{2}\mathbb{E}(\theta_{ij} \mid \theta_{ij} \in \mathcal{J}_i^{90+}).$$

How does this compare to the case without extremity bias, $\mathbb{E}(\theta_{ij} \mid \theta_{ij} \in \mathcal{J}_i)$? It is immediate to see that the two are equal for symmetric distributions. So, for instance, all of the numerical results in our Section 3 would be unaffected by this change.

Our conclusions become less sharp whenever the skew of the distribution changes. In this case, one can imagine two products with the same quality, same variance in θ_{ij}, same prices, and yet different reviews resulting from asymmetries in $\mathbb{E}(\theta_{ij} \mid \theta_{ij} \in \mathcal{J}_i^{10-})$ and $\mathbb{E}(\theta_{ij} \mid \theta_{ij} \in \mathcal{J}_i^{90+})$.

In this case, for instance, a product that is loved by few and mildly (dis)liked by many might do better than one that is appreciated – but not loved – by most, in line with Proposition 1.

It is *a priori* unclear how this dimension of heterogeneity would interact with the other bias we discuss in this paper, and particularly in Propositions 2. A more in depth analysis of the nature (and dynamics) of reviews in light of this bias is beyond the scope of this paper, and seems like a noteworthy research question.

41. One could also assume that these consumers are simply more likely to post reviews, and not the only ones to do so. This would not affect any of our reasoning below.

Another interesting case is the one in which it is the absolute – not relative – levels of love or hate for the products that shapes self-selection into reviewing. That is, consumer j leaves a review for product i when either $U_{ij} > \bar{U}$ or $U_{ij} < \underline{U}$, for two consumer- and product-independent thresholds $\underline{U} < \bar{U}$.

Under these assumptions, the average reviews of low quality products would be downward biased, while the opposite is true for products of high quality, contrary to Proposition 2 and somewhat similarly (though with slightly different drivers) to Park et al. (2021).

When niche products are also of lower quality – which has been shown to be the case in a variety of contexts, see Johnson and Myatt (2006), Bar-Isaac et al. (2012) and Sun (2012) – the conclusions are ambiguous. Again, spelling these out in greater detail seems like a promising avenue for future research.

Reviewing to Persuade. In our model, consumers are not strategic in their review behavior. They simply report their subjective opinion regarding the chosen option, irrespective of the impact of their reviews on their successors. This assumption is psychologically realistic, and additionally justified by the consumers' desire to receive future personalized recommendations, which is an important driver of review behavior on *Netflix, Yelp* and *Goodreads*, among other platforms.

Nevertheless, it is interesting to briefly discuss the case of consumers leaving reviews with the explicit desire to be persuasive (as is the case, for instance, in Chakraborty et al. (2022)). Generally, consumers motivated by persuading their peers will not rate truthfully. To see this, consider a consumer who believes that a product is of good quality (say, 4 out of 5), and before posting, notices that the product currently has an average review of 3.5. Then, her best response is to inflate her review to 5, to get the *ex-post* average review closer to her subjective quality assessment, 4.

That is, for a product of quality Q_i for which she has taste θ_{ij}, a period $t + 1$ consumer reacting to period t reviews would seek to minimize the strategic (S) loss function

$$L^S(\mathcal{R}_{ij} \mid Q_i, \theta_{ij}, \mathcal{R}_i) := -\big(Q_i + \theta_{ij} - \mathbb{E}_{t+1}(\mathcal{R}_i|\mathcal{R}_{ij})\big)^2$$

instead of the purely individual (I) one

$$L^I(\mathcal{R}_{ij} \mid Q_i, \theta_{ij}, \mathcal{R}_i) := -\big(Q_i + \theta_{ij} - \mathcal{R}_{ij}\big)^2.$$

An in-depth study of social learning with strategic review behavior is beyond the scope of this paper, and seems a promising area for future research (as also suggested by Acemoglu et al. (2022)). Here, we will only add two observations that mitigate concerns regarding the possibility (and impact) of strategic review in this context.

The first one is that $\mathbb{E}_{t+1}(\mathcal{R}_i|\mathcal{R}_{ij}) \approx \mathbb{E}_{t+1}(\mathcal{R}_i)$ whenever the number of reviews that the product had already received is large. In other words, the ability to move the average is limited when such average is built on a high number of reviews. Thus, $L^S(Q_i - \theta_{ij} - P_i \mid Q_i, \theta_{ij}, P_i, \mathcal{R}_i) \approx \max_{\mathcal{R}_{ij}} L^S(\mathcal{R}_{ij} \mid Q_i, \theta_{ij}, \mathcal{R}_i)$. This is usually the case on many online platforms such as *Netflix, Goodreads* and *IMDb*, in which every product has several thousands (and often millions) of reviews.

Second, notice that for each $j^* \in \mathcal{J}_i$, it is straightforward to sign the difference between individual and strategic reviews, $\mathcal{R}_{ij}^I - \mathcal{R}_{ij}^S$:

$$\mathcal{R}_{ij}^I < (>) \mathcal{R}_{ij}^S \;\Leftrightarrow\; \mathbb{E}(\mathcal{R}_i) < (>) Q_i + \theta_{ij^*} \;\Leftrightarrow\; \mathbb{E}(\theta_{ij} | j \in \mathcal{J}_i^t) < (>) \theta_{ij^*}.$$

In other words, consumer j^* strategic review is lower than the truthful one if and only if consumer j^* has a lower taste for the product than the average period t rater.

Much like we have seen in Proposition 2 and Corollary 1, this also gives rise to self-defeating review dynamics: products with very high reviews will motivate future strategic consumers to skew their reviews down in order to have an impact, and the opposite is true for products with low reviews. Therefore, assuming strategic motives strengthen our conclusions that reviews are pushed to the middle, understating quality differences and thus penalizing higher quality products.

Social Influence. The deviation from truthful review behavior that we have just highlighted is not the only possible one. Contrary to the contrarian-like behavior of a reviewer who has a desire to sway future consumers towards her preferred options, one can imagine at least some reviewers' opinions are at least partly reflective of (that is, anchored to) those of their predecessors.

This phenomenon is an example of social influence (see Muchnik et al. (2013) and citations therein) and can be conceptualized as "biasing the judgement of an experience – and, thus, adapting one's review – in the direction of what previous consumers have reported".

For instance, if every consumer in the previous generation has left a product glowing reviews, future consumers will rate the product higher if they were to consume it in isolation. That is,

$$\frac{\partial \mathbb{E}_{t+1}(\mathcal{R})}{\partial \mathbb{E}_t(\mathcal{R})} > 0.$$ [42]

Social influence is an important force in the digital world. For example, Muchnik et al. (2013) demonstrate, using a large scale field experiment, that randomly manipulating the first upvote or downvote received by a user post on a popular online forum influences the post's long-term upvotes to downvotes ratio. Similarly, Jacobsen (2015) shows that when famous beer bloggers review a beer more positively or negatively than the average of consumers, future consumer reviews shift in the direction of the bloggers' opinion.

This type of review behavior is often opposite to the one described in the subsection above. There, consumers effectively look as contrarians (despite their lack of

42. While beyong the scope of our paper, it is interesting to notice that this could be either because the *perceived* consumption utility went up, $\frac{\partial U_{ij}^{t+1}\left(Q_i, \theta_{ij}, \mathbb{E}_t(\mathcal{R}_i)\right)}{\partial \mathbb{E}_t(\mathcal{R}_i)} > 0$, or because reviews went up for a given U_{ij}, reflecting the rater's desire to conform to the raters in the previous period.

social image concerns), since that is what is required to affect the average review. Here, consumers have a desire to conform (or they perceive products differently depending on the previous reviews), and thus they conform to the crowd preceding them. From a learning standpoint, conformity is dangerous in this setting, because – much like in the classic work of Banerjee (1992) and Bikhchandani et al. (1992) – it leads to a halt in the aggregation of information.

Our model is *not* robust to social influence, and in fact generates prediction that are to it, as discussed at length in both Section 1 and Section 5.2. Clearly, the presence of social influence leads to *winners-take-all* dynamics: better reviews today translate into (more and) better ones tomorrow. Particularly, the opinions of particularly influential members should sway not only readers' choices, but also their very perceptions conditional on that.

We believe that a variety of empirical findings – including those of Kovács and Sharkey (2014), Rossi (2021), and He et al. (2022) – offer substantial evidence that social influence is not prevalent in this context and, if anything, high reviews (and thus sales) end up *hurting* a product's future review, as described in our Proposition 2 and Corollaries 1 and 2.

Understanding when social influence is the dominant force, and when, on the contrary, taste-based self-selection leads to robust (but potentially biased) reviews seems like a promising research question moving forward.

Online Appendix for *Alone, Together*: A Model of Social (Mis)Learning from Consumer Reviews

A. Optimal Pricing

Throughout our analysis up to this point, we have fully focused on the demand side, that is, we have studied consumer (mis)learning in a context in which firms were "passive". It is interesting to now consider a more strategic model in which firms react to – and, as we will see, influence – the information environment by setting their optimal prices. For analytical simplicity, we go back to the case of duopoly without an outside option – which also nests the case of monopoly with an outside options, if we assume that one of the two products, i, is such that $\theta_{ij} \equiv 0 \ \forall j$.

Our goal is two-fold: first, we aim to gather insights on the robustness (or lack thereof) of our findings to endogenous prices, that is, study how optimal prices influence properties of long-run reviews. Second, conversely, we want to study how the presence of reviews influence firms' pricing.

To this end, assume, as in the previous analysis, that firm i has quality Q_i, design s_i and set price P_i, $i = 1, 2$. Denote by $\Delta(Q) := Q_1 - Q_2$ the quality advantage of firm 1 and assume, without loss of generality, that it is non-negative. We define the *relative* taste for firm 1 by $\theta_j \equiv \theta_{1j} - \theta_{2j}$, with $\theta_{ij} \sim F_i(\cdot)$, and by $G(\cdot)$ and $g(\cdot)$ its cumulative and density respectively. We assume that $G(\cdot)$ is symmetric and that it satisfies the monotone hazard rate, that is, $h(\cdot) := \frac{g(\cdot)}{1-G(\cdot)}$ is non-decreasing. For brevity, we will drop the consumer subscript j from this point on.

To emphasize the role of subjective reviews on prices (and, conversely, that of optimal prices on reviews), we compare this case to two opposite benchmarks: no information and full information. Thus, we characterize pricing equilibria in three informational environments: 1) **no information** (N), in which consumers are unaware of quality differences, 2) **full information** (F) , in which the consumers are aware of the firms' qualities Q_1 and Q_2, and thus their difference $\Delta(Q)$, and 3) **subjective reviews** $(\mathcal{R})$, in which buyers of each products report their utility from it, as previously described in this paper. In this third case, consistent with the rest of our paper, we assume consumers take reviews at face value.[1]

Throughout the analysis, we denote consumers' beliefs about firm 1's quality advantage by $\tilde{\Delta}(Q)$. In particular, we have $\tilde{\Delta}(Q) = 0$, $\tilde{\Delta}(Q) = \Delta(Q)$ and $\tilde{\Delta}(Q) = \Delta_\infty(\mathcal{R})$ in the no information, full information and subjective reviews cases, respectively.[2]

1. The second case can be seen as one in which all consumers are Bayesian, and thus fully internalize, and correct for, potential biases in subjective reviews.
2. Throughout this Section, to ensure consistency between the three cases, reviews should be thought of as reflecting gross utilities: $\mathcal{R}_{ij} = Q_i + \theta_{ij}$. Equivalently, we assume that, since prices are observable, if reviews were reflecting net utilities instead, consumers could simply account for prices. We refer to Section 4.2 for a more detailed discussion of this point.

First, we fully characterize the properties of optimal prices in each of these three environments. Then, we compare prices across environments.

A.1. No Information

Given the lack of information regarding firms' (relative) qualities, each consumer simply trades off price and fit for each product. That is, a consumer buys from firm 1 if and only if

$$\theta_j \geq P_1 - P_2.$$

Thus, we have

$$\pi_1(P_1, P_2) = P_1 \cdot \big(1 - G(P_1 - P_2)\big)$$

and

$$\pi_1(P_1, P_2) = P_2 \cdot G(P_1 - P_2).$$

We have the following

Proposition 9. *In the unique equilibrium,*

$$P_1^N = P_2^N = \frac{G(0)}{g(0)} = \frac{1}{2g(0)}. \tag{36}$$

A.2. Full Information

We now turn to the case in which consumers are fully aware of the quality difference, $\Delta(Q)$. We have the following:

Proposition 10. *In the unique equilibrium, we have*

$$P_1^F = \frac{1 - G(\Delta(P) - \Delta(Q))}{g(\Delta(P) - \Delta(Q))} \tag{37}$$

and

$$P_2^F = \frac{G(\Delta(P) - \Delta(Q))}{g(\Delta(P) - \Delta(Q))}. \tag{38}$$

Comparing to the no-information case, we have $P_2^F < P_2^N = P_1^N < P_1^F$. Last, firm 1 captures more than half of the market.

A.3. Subjective Reviews

Now consider the case in which reviews are present, and consumers take them at face value, that is, equating $\tilde{\Delta}(Q) = \Delta_\infty(\mathcal{R})$.

We have

$$\mathbb{E}_\infty(\mathcal{R}_1) = Q_1 + \mathbb{E}(\theta_1 \mid \mathbb{E}_\infty(\mathcal{R}_1) + \theta_1 - P_1 > \mathbb{E}_\infty(\mathcal{R}_2) + \theta_2 - P_2))$$

and

$$\mathbb{E}_\infty(\mathcal{R}_2) = Q_2 + \mathbb{E}(\theta_2 \mid \mathbb{E}_\infty(\mathcal{R}_2) + \theta_2 - P_2 > \mathbb{E}_\infty(\mathcal{R}_1) + \theta_1 - P_1)).$$

This is in line with the rest of our paper. In this context, the key observation is that reviews do not *directly* reflect prices. While not a crucial assumption, this is justified by the fact that consumers observe prices directly, so that, if reviews reflected them, consumers could easily invert reviews to then employ them as (potentially biased) quality signals.[3]

It is useful to combine the previous two equations into

$$\begin{aligned}\Delta_\infty(\mathcal{R}) = \Delta(Q) + \mathbb{E}\big(\mathbb{E}\big(\theta_1 \mid \Delta\theta > \Delta(P) - \Delta_\infty(\mathcal{R})\big) \\ - \mathbb{E}\big(\theta_2 \mid \Delta\theta < \Delta(P) - \Delta_\infty(\mathcal{R})\big)\end{aligned} \tag{39}$$

Crucially, notice how firm 1 can increase the RHS of Eq. 39 by increasing P_1, and the symmetric is true for firm 2 increasing P_2. By setting a high price a firm improves the self-selection of its consumers, while worsening the self-selection of its competitor's consumers.

We start with three Lemma's, each instrumental to proving our main Proposition:

Lemma 4. $0 < \frac{\partial \Delta_\infty(\mathcal{R})}{\partial \Delta(P)} < 1$.

Lemma 5. *If* $\Delta(P^{\mathcal{R}}) < \Delta(Q)$, *then* $\Delta(P^{\mathcal{R}}) < \Delta_\infty(\mathcal{R}) < \Delta(Q)$; *If* $\Delta(P^{\mathcal{R}}) > \Delta(Q)$, *then* $\Delta(Q) < \Delta_\infty(\mathcal{R}) < \Delta(P^{\mathcal{R}})$.

Lemma 6. *In equilibrium, we have* $0 < \Delta(P^{\mathcal{R}}) < \Delta(Q)$.

We are now ready to state the following:

Proposition 11. *Equilibrium prices are given by*

$$P_1^{\mathcal{R}} = \frac{1 - G(\Delta(P) - \Delta_\infty(\mathcal{R}))}{g(\Delta(P) - \Delta_\infty(\mathcal{R}))\Big(1 - \frac{\partial \Delta_\infty(\mathcal{R})}{\partial \Delta(P)}\Big)} \tag{40}$$

and

$$P_2^{\mathcal{R}} = \frac{G(\Delta(P) - \Delta_\infty(\mathcal{R}))}{g(\Delta(P) - \Delta_\infty(\mathcal{R}))\Big(1 - \frac{\partial \Delta_\infty(\mathcal{R})}{\partial \Delta(P)}\Big)}. \tag{41}$$

In equilibrium, $0 < \Delta(P^{\mathcal{R}}) < \Delta_\infty(\mathcal{R}) < \Delta(Q)$.

3. For an alternative modeling choice, in which consumers are unaware of whether reviews include prices, and thus fail to correct for them, see Carnehl et al. (2023).

First and foremost, Proposition 11 demonstrates the robustness of our main results to the introduction of optimal pricing by firms. This is an important insight, as one could have suspected prices to work as a corrective mechanism, undoing the reviews' biases. Proposition 11 also has a host of additional, interesting implications. We start by noticing that, because $\Delta(P^{\mathcal{R}}) - \Delta_\infty(\mathcal{R}) < 0$, we have

$$P_1^{\mathcal{R}} > \frac{1 - G(0)}{g(0)\Big(1 - \frac{\partial \Delta \mathcal{R}}{\partial \Delta(P)}\Big)} > \frac{1 - G(0)}{g(0)} = P_1^N. \tag{42}$$

The two inequalities emphasize the presence of two totally distinct forces, each pushing Firm 1's prices up. First, because $\Delta_\infty(\mathcal{R}) > 0$, the quality advantage of Firm 1 is now known to consumers. Second, both firms have additional incentives to increase their prices to improve the matches with their consumers ("price as a matching device"), while decreasing their competitor's matches, thus increasing their relative reviews: this is quantified by the $1/\big(1 - \frac{\partial \Delta_\infty(\mathcal{R})}{\partial \Delta(P)}\big)$ term appearing in both $P_1^{\mathcal{R}}$ and $P_2^{\mathcal{R}}$.

The analysis for firm 2 is more ambiguous, since the two aforementioned forces point in opposite directions: while (part of) the quality disadvantage of firm 2 is revealed, firm 2 also has the same incentive to increase prices to improve matches, and thus reviews, as firm 1.

A.4. Combining the Three Cases

Despite the aforementioned ambiguity, we find that, strikingly, when $\Delta(Q)$ is not too large, both firms' prices are highest in the case of subjective reviews.

Proposition 12. *Suppose that* $\frac{d\mathbb{E}(\theta_1|\Delta\theta>k)}{dk} + \frac{d\mathbb{E}(\theta_2|\Delta\theta<k)}{dk} > \epsilon > 0$ *in a neighborhood of* 0 *(for k).*[4] *Then, there exists a* $\Delta(Q^*)$ *such that whenever* $\Delta(Q) < \Delta(Q^*)$*, we have*

$$P_1^{\mathcal{R}} > P_1^F > P_1^N, \quad P_2^{\mathcal{R}} > P_2^N > P_2^F. \tag{43}$$

Taken together, our findings in Section A highlight several interesting facts. First and most obviously, the combination of subjective reviews and consumer naïvete can lead both firms to price higher compared to both the no information and the full information case. This happens despite the fact that, from an informational point of view, we have shown the subjective review case to always lie between the other two: $0 < \Delta_\infty(\mathcal{R}) < \Delta(Q)$. The specific price rankings highlighted in Proposition 12 crucially depend on $\Delta(Q)$ to be small – if this condition is violated, then firm 2 would be revealed to be at a large quality disadvantage, which would decrease its prices compared to the no information case, leading to $P_2^N > P_2^{\mathcal{R}}$, while firm 1's perceived quality advantage would be much smaller than in the full information case,

4. The condition that $\frac{d\mathbb{E}(\theta_1|\Delta\theta>k)}{dk} + \frac{d\mathbb{E}(\theta_2|\Delta\theta<k)}{dk} > \epsilon > 0$ is easily satisfied by many distributions. For example, it is satisfied if θ_1 and θ_2 are both uniformly distributed or both normally distributed.

leading to $P_1^F > P_1^{\mathcal{R}}$. But the more general point does not: with subjective reviews that depend on average conditional taste distributions, both firms have incentives to price higher than they would otherwise, since prices act as *matching devices*, or, put differently, as a reviews management tool: a higher relative taste is required to pick a product when its relative price is higher.

Perhaps more importantly for our analysis, Proposition 11 highlights that Proposition 1 and 2 are robust to the introduction of optimal pricing by both firms. It is interesting to justaxpose our findings with those of Sayedi (2018), who shows that the pathological outcomes in the classic observational learning models of Banerjee (1992) and Bikhchandani et al. (1992) disappear whenever the two firms are allowed to price optimally. This is because, with observational learning, the higher quality firm is incentivized to dramatically lower its prices whenever a cascade on the lower quality product starts. Here, on the other hand, when quality differences are not overwhelming, both firms are incentivized to increase their equilibrium prices to obtain better matches, and thus higher reviews, than they would if reviews were absent altogether.

Because equilibrium reviews reveal some information ($\Delta_\infty(\mathcal{R}) > 0$), firm 1 still partly cashes in on its revealed quality advantage, and thus sets higher prices than firm 2. Therefore, the number of reviews for the two products are closer to $0.50 - -0.50$ than they would be with symmetric prices. But the point remains that this correction is only partial, and all the biases highlighted in our analysis ($\Delta_\infty(\mathcal{R}) < \Delta(Q)$) are robust to optimal pricing by both firms.

B. Additional Proofs

B.1. Proofs for Main Text Results

Proof of Corollary 1

We have that the period $t+1$ set of consumers for product i is given by:

$$\mathcal{J}_i^{t+1} = \{j \mid \theta_{ij} + \mathbb{E}_t(\mathcal{R}_i) > \theta_{-ij} + \mathbb{E}_t(\mathcal{R}_{-i})\}$$

Now let $\mathbb{E}_t(\mathcal{R}_i)$ increase to $\mathbb{E}_t(\mathcal{R}_i) + \epsilon$. Clearly, this implies that $\mathcal{J}_i^{t+1,\epsilon} \supset \mathcal{J}_i^{t+1}$. It also implies that the crowd of product i buyers becomes less self-selected:

$$\mathbb{E}(\theta_{ij} \mid \theta_{ij} + \mathbb{E}_t(\mathcal{R}_i) + \epsilon > \theta_{-ij} + \mathbb{E}_t(\mathcal{R}_{-i}))) < \mathbb{E}(\theta_{ij} \mid \theta_{ij} + \mathbb{E}_t(\mathcal{R}_i) > \theta_{-ij} + \mathbb{E}_t(\mathcal{R}_{-i}))),$$

which in turns causes $\mathbb{E}_{t+1}(\mathcal{R}_i)$ to decline. The case of $\mathbb{E}_{t+1}(\mathcal{R}_{-i})$ can be handled symmetrically.

Clearly, because $\mathcal{B}_{t+1} := (\mathbb{E}_{t+1}(\mathcal{R}_1) - \mathbb{E}_{t+1}(\mathcal{R}_2)) - (Q_1 - Q_2)$, a decrease in $\mathbb{E}_{t+1}(\mathcal{R}_1)$ and an increase in $\mathbb{E}_{t+1}(\mathcal{R}_2)$ will compound to cause a larger decrease in $\mathcal{B}_{t+1}$. ■

Proof of Corollary 2

The proof follows directly from our previous results. ■

Proof of Corollary 3

Assume that product quality is Q, and, without loss of generality, that initial belief is 0, compared to an outside option $c \in \mathbb{R}$. We have

$$\mathbb{E}_0(\mathcal{R}) = Q + \mathbb{E}(\theta|\theta > c) > Q.$$

But then,

$$\mathbb{E}_1(\mathcal{R}) = Q + \mathbb{E}(\theta|\theta + \mathbb{E}_0(\mathcal{R}) > c).$$

Clearly, $\mathbb{E}_1(\mathcal{R}) \in (0, \mathbb{E}_0(\mathcal{R}))$. Thus, a similar argument to the one applied in the proof of Proposition 1 implies that reviews converge.

The convergence point is given by the fixed point of the above equation:

$$\mathbb{E}_\infty(\mathcal{R}) = Q + \mathbb{E}(\theta \mid \theta + \mathbb{E}_\infty(\mathcal{R}) > c).$$

Assume by contradiction that $\mathbb{E}_\infty(\mathcal{R}_H) \geq \mathbb{E}_\infty(\mathcal{R}_L)$. This implies

$$\mathbb{E}\big(\theta_H \mid \mathbb{E}_\infty(\mathcal{R}_H) + \theta_H > c\big) \geq \mathbb{E}\big(\theta_L \mid \mathbb{E}_\infty(\mathcal{R}_L) + \theta_L > c\big).$$

But then, *a fortiori*, we have

$$\mathbb{E}\big(\theta_H \mid \theta_H > c\big) \geq \mathbb{E}\big(\theta_L \mid \theta_L > c\big).$$

Therefore, to reach a contradiction we have to show that, when c is high enough,

$$\mathbb{E}(\theta_L \mid \theta_L > c) > \mathbb{E}(\theta_H \mid \theta_H > c)$$

We have that

$$\mathbb{E}(\theta_L \mid \theta_L > c) = \frac{\int_c^{\bar{\theta}} \big(1 - F_{s_L}(\theta)\big) d\theta}{1 - F_{s_L}(c)} \tag{44}$$

and similarly

$$\mathbb{E}(\theta_1 \mid \theta_1 > c) = \frac{\int_c^{\bar{\theta}} \big(1 - F_{s_H}(\theta)\big) d\theta}{1 - F_{s_H}(c)}. \tag{45}$$

Note that, by the definition of demand rotations, there exists a unique $\theta^\dagger$ such that $F_{s_L}(\theta^\dagger) = F_{s_H}(\theta^\dagger)$ and $F_{s_L}(\theta) < F_{s_H}(\theta)$ for every $\theta > \theta^\dagger$ (in particular, $\theta^\dagger = 0$ whenever s is symmetric).

As a result,

$$\int_{\theta^\dagger}^{\bar{\theta}} \big(1 - F_{s_H}(\theta)\big) d\theta < \int_{\theta^\dagger}^{\bar{\theta}} \big(1 - F_{s_L}(\theta)\big) d\theta$$

and thus, substituting $c = \theta^\dagger$ in Equations (44) and (45), we have

$$\mathbb{E}(\theta_L \mid \theta_L > \theta^\dagger) > \mathbb{E}(\theta_H \mid \theta_H > \theta^\dagger).$$

The result is true *a fortiori* for every $c > \theta^\dagger$. By continuity of both conditional expected values – which follows from the smoothness of both $F_L(\cdot)$ and $F_H(\cdot)$ – we have that there exists a $c^* \in [\underline{\theta}, \theta^\dagger)$ such that the result is true for every $c > c^*$. Thus, $\mathbb{E}_\infty(\mathcal{R}(s_L)) > \mathbb{E}_\infty(\mathcal{R}(s_H))$ as desired. This concludes the proof. ■

Proof of Corollary 4

Assume product quality is Q. Initial belief 0 (wlog) compared to outside option $c \in \mathbb{R}$. We have

$$\mathbb{E}_0(\mathcal{R}) = Q + \mathbb{E}(\theta|\theta > c) > Q.$$

But then,

$$\mathbb{E}_1(\mathcal{R}) = Q + \mathbb{E}(\theta|\theta + \mathbb{E}_0(\mathcal{R}) > c).$$

Clearly, $\mathbb{E}_1(\mathcal{R}) \in (0, \mathbb{E}_1(\mathcal{R}))$. A similar logic implies $\mathbb{E}_2(\mathcal{R}) = Q + \mathbb{E}(\theta|\theta + \mathbb{E}_1(\mathcal{R}) > c) \in (\mathbb{E}_1(\mathcal{R})), \mathbb{E}_0(\mathcal{R})))$. Thus, a similar argument to the one applied in the proof of Proposition 2 implies that reviews converge.

The convergence point is given by the fixed point of the above equation:

$$\mathbb{E}_\infty(\mathcal{R}) = Q + \mathbb{E}(\theta|\theta + \mathbb{E}_\infty(\mathcal{R}) > c). \tag{46}$$

We want to show that $\mathcal{B}_\infty = \mathbb{E}_\infty(\mathcal{R}) - Q$ is decreasing in Q.

To show this in a straightforward fashion, notice that as Q increases by ϵ, so does the RHS of Eq. (46).

Now suppose $\mathbb{E}_\infty(\mathcal{R})$ also increases by ϵ: then, first, the LHS goes up by ϵ; and second, the RHS decreases, since $\mathbb{E}(\theta \mid \theta + \mathbb{E}_\infty(\mathcal{R}) + \epsilon > c)$ is decreasing in ϵ. Thus, if Q and $\mathbb{E}_\infty(\mathcal{R})$ were to increase equally, the LHS would exceed the RHS.

At the same time, if Q increased by ϵ and $\mathbb{E}_\infty(\mathcal{R})$ were unchaged, the RHS would exceed the LHS.

Thus, we have

$$\frac{\partial \mathbb{E}_\infty(\mathcal{R})}{\partial Q} \in (0,1) \quad \Rightarrow \quad \frac{\partial \mathcal{B}_\infty(Q)}{\partial Q} = \frac{\partial \mathbb{E}_\infty(\mathcal{R})}{\partial Q} - 1 < 0.$$

The number of reviews is given by

$$\mathcal{N}_\infty(\mathcal{R})(Q) = Prob(\theta + \mathbb{E}_\infty(\mathcal{R}) > c),$$

which is increasing in Q because $\mathbb{E}_\infty(\mathcal{R})$ is. This concludes the proof. ■

Proof of Corollary 5

We show a direct example of welfare reducing social learning from reviews, and then extend it.

We have seen in Proposition 2, Claim 5 that whenever $s_1 = s_2$ learning from reviews is welfare enhancing. Therefore, assume now that the products differ in their designs $s_1 = L$, $s_2 = H$. Assume furthermore that $Q_1 = Q_2$. Then, we have that

$$\mathcal{J}_1^{NR} = \emptyset \tag{47}$$

whereas in the presence of biased reviews we have

$$\mathcal{J}_1^{BR} = \{j \mid \theta_{1j} < \theta_{2j} - \Delta_\infty(\mathcal{R}) \;\;\&\;\; \theta_{2j} < \theta_{1j}\}. \tag{48}$$

Clearly, $\mathcal{J}_1^{BR}$ is non-empty because we know from Proposition 1 that in this case, long-run reviews are biased in favor of the more polarizing product: $\mathcal{B}_\infty = \Delta_\infty(R) < 0$.

Welfare in this case is given by the probability of a consumer making the correct choice. This is given by $1 - Prob(j \in \mathcal{J}_1^{BR})$ in the case of biased reviews and 1 otherwise. Thus, in this case reviews reduce welfare.

Clearly, the above example is not particularly surprising: when quality differences are 0 to begin with, and reviews help consumers make inference about quality differences, no improvement over the prior $Q_1 = Q_2$ is possible.

Thus, we extend this result to the case of quality asymmetries: $\Delta(Q) = Q_1 - Q_2 > 0$. Now define by

$$\mathcal{J}_1^{NR}(\Delta(Q)) = \{j \mid \theta_{1j} > \theta_{2j} \;\;\&\;\; \theta_{1j} > \theta_{2j} - \Delta(Q)\},$$

and

$$\mathcal{J}_1^{BR}(\Delta(Q)) = \{j \mid \theta_{1j} > \theta_{2j} - \Delta_\infty(\mathcal{R})(\Delta(Q)) \;\; \& \;\; \theta_{2j} > \theta_{1j}\}.$$

We have $\mathcal{J}_1^{BR}(0) \subset \mathcal{J}_1^{NR}(0)$. By continuity, there exists a $\Delta^*(Q)$ such that the result holds for any $\Delta^*(Q) < \Delta(Q)$. (We have suppressed the dependence of $\Delta^*(Q)$ on $\mathcal{B}_\infty$.)

But then,

$$\mathcal{J}_{NR}(\Delta^*(Q)) \subset \mathcal{J}_{QR}(\Delta^*(Q)) \quad \forall \Delta^*(Q) < \Delta(Q). \tag{49}$$

Thus,

$$1 - Prob(\mathcal{J}_{BR}(\Delta^*(Q))) < 1 - Prob(\mathcal{J}_{NR}(\Delta^*(Q))) \quad \forall \Delta^*(Q) < \Delta(Q). \tag{50}$$

Thus, reviews are welfare reducing whenever differences in designs are large, and quality differences are small. ■

B.2. Proofs for Appendix B Results

Proof of Proposition 5

Proof:

The proof of Proposition 5 essentially replicates the steps of that of Proposition 1.

- The fact that both reviews are higher than they would have been without an outside option follows from the fact that

$$\mathbb{E}(\theta_1|\theta_1 > (\max \theta_2 - \Delta_\infty(\mathcal{R}), c)) > \mathbb{E}(\theta_1|\theta_1 > \theta_2 - \Delta_\infty(\mathcal{R})),$$

  and similarly for product 2, as the presence of a non-trivial outside option c increases the lower bound of integration for at least some consumers – the more so the higher c.
- Following the reasoning in the proof of Proposition 1, this is equivalent to showing that

$$\mathbb{E}(\theta_2|\theta_2 > \max(\theta_1, c)) > \mathbb{E}(\theta_1|\theta_1 > \max(\theta_2, c)).$$

  The LHS can be rewritten as

$$\mathbb{E}(\theta_2|\theta_2 > \max(\theta_1, c)) = P(\theta_1 > c)\mathbb{E}(\theta_2|\theta_2 > \theta_1) + P(\theta_1 < c)\mathbb{E}(\theta_2|\theta_2 > c).$$

  We know from our Proofs of Propositions 1 and Corollary 3 that $\mathbb{E}(\theta_2|\theta_2 > \theta_1) > \mathbb{E}(\theta_1|\theta_1 > \theta_2)$ and $\mathbb{E}(\theta_2|\theta_2 > c) > \mathbb{E}(\theta_1|\theta_1 > c)$. The result follows.

- The proof for this result is straightforward, and can be found in the proof of Proposition 1.
- We have that

$$\mathbb{E}_0(\mathcal{R}_1) = Q_1 + \mathbb{E}(\theta_{1j}|\theta_{1j} > max(\theta_{2j}, 0)) \tag{51}$$

and similarly for product 2. Thus, $\mathcal{B}_0 = \mathbb{E}_1(\mathcal{R}_1) - \mathbb{E}_1(\mathcal{R}_2)$ implies

$$\mathcal{B}_0 = \mathbb{E}(\theta_{1j} \mid \theta_{1j} > max(\theta_{2j}, c))) - \mathbb{E}(\theta_{2j} \mid \theta_{2j} > max(\theta_{1j}, c))), \tag{52}$$

where we have simplified the RHS using the fact that $Q_1 = Q_2$ by assumption. To show that $\mathcal{B}_0 < \mathcal{B}_\infty$, assume by contradiction $\mathcal{B}_0 = \mathcal{B}_\infty$. But then, we obtain

$$\begin{aligned} \mathcal{B}_0 =& \mathbb{E}(\theta_{1j} \mid \theta_{1j} > max(\theta_{2j} - \mathcal{B}_0, c - \mathbb{E}_\infty(\mathcal{R}_1)) - \\ & \mathbb{E}(\theta_{2j} \mid \theta_{2j} > max(\theta_{1j} + \mathcal{B}_0, c - \mathbb{E}_\infty(\mathcal{R}_2)). \end{aligned} \tag{53}$$

Therefore,

$$\begin{aligned} \mathcal{B}_0 &= \mathbb{E}(\theta_{1j} \mid \theta_{1j} > \theta_{2j}) - \mathbb{E}(\theta_{2j} \mid \theta_{2j} > \theta_{1j}) \\ &< \mathbb{E}(\theta_{1j} \mid \theta_{1j} > max(\theta_{2j} - \mathcal{B}_0, c - \mathbb{E}_\infty(\mathcal{R}_1)) \\ &- \mathbb{E}(\theta_{2j} \mid \theta_{2j} > max(\theta_{1j} + \mathcal{B}_0, c - \mathbb{E}_\infty(\mathcal{R}_2)) \\ &= \mathcal{B}_\infty, \end{aligned}$$

where the inequality follows from the fact that $\mathcal{B}_\infty < 0$ (equivalently, $\mathbb{E}_\infty(\mathcal{R}_1) < \mathbb{E}_\infty(\mathcal{R}_2)$, as established in the Proof of **Claim 1**. The conclusions hold *a fortiori* if $\mathcal{B}_0 > \mathcal{B}_\infty$. This proves that $\mathcal{B}_0 < \mathcal{B}_\infty$. The fact that $\mathcal{N}_1(\mathcal{R}_2) > \mathcal{N}_\infty(\mathcal{R}_2)$ follows straightforwardly from $\mathcal{B}_0 < \mathcal{B}_\infty$ using the same argument as in the Proof of **Claim 2**. ■

■

Proof of Proposition 6

Proof:

The proof of Proposition 6 essentially replicates the steps of that of Proposition 2, with some minor modifications.

- The fact that both reviews are higher than they would have been without an outside option follows from the fact that

$$\mathbb{E}(\theta_1|\theta_1 > \max(\theta_2 - \Delta_\infty(\mathcal{R}), c)) > \mathbb{E}(\theta_1|\theta_1 > \theta_2 - \Delta_\infty(\mathcal{R})),$$

as the presence of a non-trivial outside option c increases the lower bound of integration for at least some consumers – the more so the higher c.

- To see that product 1 has higher long-run reviews, assume by contradiction that $\mathbb{E}(\mathcal{R}_1) = \mathbb{E}(\mathcal{R}_2)$. Then, we have that

$$0 > -(Q_1 - Q_2) = \mathcal{B}_\infty = \mathbb{E}(\theta_1|\theta_1 > \max(\theta_2, c)) - \mathbb{E}(\theta_2|\theta_2 > \max(\theta_1, c)) = 0$$

  where the last equality comes from the fact that $s_1 = s_2$. We have reached a contradiction. Just like in the case of Proposition 2, one can show that if $\mathbb{E}(\mathcal{R}_1) < \mathbb{E}(\mathcal{R}_2)$, the LHS decreases, while the RHS increases.

- To show that $\mathcal{B}_\infty < 0$, assume by contradiction that $\mathcal{B}_\infty \geq 0$. This is equivalent to

$$\begin{aligned}&\mathbb{E}(\theta_1|\theta_1 > \max(\theta_2 - \Delta_\infty(\mathcal{R}), c - \mathbb{E}_\infty(\mathcal{R}_1)) - \\ &\mathbb{E}(\theta_2|\theta_2 > \max(\theta_1 + \Delta_\infty(\mathcal{R}), c - \mathbb{E}_\infty(\mathcal{R}_2)) > 0.\end{aligned}$$

  But this is a contradiction, because $s_1 = s_2$, $\Delta_\infty(\mathcal{R}) > \Delta(Q) > 0$ and thus $c - \mathbb{E}_\infty(\mathcal{R}_1) < c - \mathbb{E}_\infty(\mathcal{R}_2)$, implying that the second conditional expected value exceeds the first.

- The proof for this result is straightforward, and can be found in the proof of Proposition 2.

  We have that $\mathcal{B}_0 = 0$ since $\mathbb{E}(\theta_1|\theta_1 > max(\theta_2, c)) = \mathbb{E}(\theta_2|\theta_2 > max(\theta_1, c))$. That is, $\mathbb{E}_0(\mathcal{R}_1) - \mathbb{E}_0(\mathcal{R}_2) = Q_1 - Q_2 > 0$. Therefore,

$$\begin{aligned}\mathbb{E}_1(\mathcal{R}_1) &= Q_1 + \mathbb{E}(\theta_1 \mid \theta_1 + \mathbb{E}_0(\mathcal{R}_1) > max(\theta_2 + \mathbb{E}_0(\mathcal{R}_2), c) \\ &= Q_1 + \mathbb{E}(\theta_1 \mid \theta_1 > max(\theta_2 - \Delta(Q), c))\end{aligned}$$

  and similarly

$$\begin{aligned}\mathbb{E}_1(\mathcal{R}_2) &= Q_1 + \mathbb{E}(\theta_2 \mid \theta_2 + \mathbb{E}_0(\mathcal{R}_2) > max(\theta_1 + \mathbb{E}_0(\mathcal{R}_1), c) \\ &= Q_2 + \mathbb{E}(\theta_2 \mid \theta_2 > max(\theta_1 - \Delta(Q), c))\end{aligned}$$

  Jointly, these two imply that

$$\mathcal{B}_1 = \mathbb{E}(\theta_1 \mid \theta_1 > max(\theta_2 - \Delta(Q), c)) - \mathbb{E}(\theta_2 \mid \theta_2 > max(\theta_1 - \Delta(Q), c)). \tag{54}$$

  But then, $\mathcal{B}_1 < \mathcal{B}_\infty$ if and only if $\Delta(Q) > \mathbb{E}_\infty(\mathcal{R}_1) - \mathbb{E}_\infty(\mathcal{R}_2)$. But this is equivalent to $\mathcal{B}_\infty < 0$, which we have shown to be true in the Proof of **Claim 1**.

  As in the previous cases, the fact that $\mathcal{N}_2(\mathcal{R}_1) < \mathcal{N}_\infty(\mathcal{R}_1)$ follows straightforwardly from $\mathcal{B}_1 < \mathcal{B}_\infty$.

- The proof for this claim is exactly the same as the one for the corresponding claim in Proposition 2.

■

Proof of Proposition 7

Proof:

We want to show that

$$Var\big(\theta_L \mid \theta_L > max(\theta_H + \Delta(Q), c)\big) < Var\big(\theta_H \mid \theta_H > max(\theta_H - \Delta(Q), c)\big), \quad \forall \Delta(Q) > \Delta^*(Q).$$

First notice that, when $\Delta(Q)$ approaches $\bar{\theta} - \underline{\theta}$, we have

$$Var\big(\theta_H \mid \theta_H > max(\theta_H - \Delta(Q), c)\big) \rightarrow Var(\theta_H | \theta_H > c).$$

On the other hand, the fact that $\theta_L > \theta_H + \Delta(Q)$ implies $\theta_L \in (\underline{\theta} + \Delta(Q),\ \bar{\theta})$. Therefore, Popoviciu's Inequality Popoviciu (1935) implies that

$$Var\big(\theta_L \mid \theta_L > \theta_H + \Delta(Q)\big) \leq \frac{1}{4}\big(\bar{\theta} - \underline{\theta} - \Delta(Q)\big)^2.$$

Notice that the right hand side gets arbitrarily small as $\Delta(Q) \rightarrow \bar{\theta} - \underline{\theta}$, implying the existence of a $\Delta^*(Q)$ such that $Var\big(\theta_L | \theta_L > \theta_H + \Delta(Q)\big) < Var(\theta_H)$ for every $\Delta(Q) > \Delta^*(Q)$. A fortiori, this is true for $Var\big(\theta_L \mid \theta_L > max(\theta_H + \Delta(Q), c)\big)$. ■

Proof of Proposition 8

Proof: Assume, for now, that $\Delta(Q) > 0$ and $s_1 = s_2$. We start by showing uniqueness. The proofs follows straightforwardly from the uniqueness of a solution for the equation

$$\Delta(\mathcal{R}) = \Delta(Q) + \mathbb{E}(\theta_1 | \theta_1 > \theta_2 - \Delta(\mathcal{R})) - \mathbb{E}(\theta_2 | \theta_2 > \theta_1 + \Delta(\mathcal{R})). \tag{55}$$

As discussed in our Proofs of Propositions 2, 1 and 11, this uniqueness is guaranteed by the fact *i)* the LHS is increasing in $\Delta(\mathcal{R})$, while the RHS is decreasing, *ii)* the RHS exceeds the LHS when $\Delta(\mathcal{R}) = 0$ and *iii)* the LHS exceeds the RHS when $\Delta(\mathcal{R}) = \Delta(Q)$, and similarly if the products differ, instead, in their design.

Next, we want to show that $\Delta_t(\mathcal{R}^C)(\beta)$ converges whenever $\Delta_t(\mathcal{R})$ does. To see this, we start by showing that $\Delta_2(\mathcal{R}^C)(\beta) \in (\Delta_1(\mathcal{R}^C)(\beta),\ \Delta_0(\mathcal{R}^C)(\beta))$.

But this is immediate since (as shown in the proof of Proposition 2), when $\Delta(Q) > 0$ we have $\Delta_0(\mathcal{R}) = \Delta(Q)$, $0 < \Delta_1(\mathcal{R}) < \Delta(Q)$ and $\Delta_2(\mathcal{R}) \in (\Delta_1(\mathcal{R}),\ \Delta_0(\mathcal{R}))$. Thus, in the cumulative case, $\Delta_1(\mathcal{R}^C)(\beta) \in (0,\ \Delta_0(\mathcal{R}^C)(\beta))$, as $\Delta_1(\mathcal{R}^C)(\beta)$ is a weighted average of $\Delta_0(\mathcal{R})$ and $\Delta_1(\mathcal{R})$ for every β. This immediately implies that $\Delta_2(\mathcal{R}^C)(\beta) < \Delta_0(\mathcal{R}^C)(\beta) = \Delta(Q)$. Thus, a similar reasoning to the one employed in the proof of Proposition 2 implies that the sequence $\{\Delta_t(\mathcal{R}^C)(\beta)\}_{t=0}^{\infty}$ converges for every $\beta \in [0, 1]$.

The proof for the case for $s_1 \neq s_2$ follows very similar steps, and be easily derived by adapting the previous case and combining it with the proof of Proposition 1.

■

C. Proofs for Online Appendix A: Optimal Pricing

Proof of Proposition 9

First order conditions for firm 1 and 2 are given by, respectively,

$$1 - F(P_1^N - P_2^N) - P_1^N f(P_1^N - P_2^N) = 0 \tag{56}$$

and

$$F(P_1^N - P_2^N) - P_2^N f(P_1^N - P_2^N) = 0 \tag{57}$$

Jointly, these imply $P_1^N = \frac{1-F(P_1^N - P_2^N)}{f(P_2^N - P_1^N)}$ and $P_2^N = \frac{F(P_1^N - P_2^N)}{f(P_1^N - P_2^N)}$. Now, define $\Delta(P) := P_1^N - P_2^N$. Subtracting the two expressions for P_1^N and P_2^N we have just found, we get

$$\Delta(P)^N f(\Delta(P)^N) = 1 - 2F(\Delta(P)^N). \tag{58}$$

Now, notice that the LHS is positive if and only if $\Delta(P)^N$ is. Conversely, the RHS is positive whenever $F(\Delta(P)^N) < 1/2$, and negative afterwards. Because $F(0) = 1/2$, this implies $\Delta(P)^N = 0$. Thus, the two only intersect at 0, which implies that $P_1^N = P_2^N$ in every equilibrium.

Now, plugging this back into Eq. (57), we get that in equilibrium

$$P_1^N = P_2^N = \frac{F(0)}{f(0)} = \frac{1}{2f(0)}, \tag{59}$$

when the equality comes from the symmetry of $F(\cdot)$. ■

Proof of Proposition 10

The first order conditions are given by

$$1 - F(\Delta(P)^F - \Delta(Q)) - P_1 f(\Delta(P)^F - \Delta(Q)) = 0 \tag{60}$$

and

$$F(\Delta(P)^F - \Delta(Q)) - P_2 f(\Delta(P)^F - \Delta(Q)) = 0 \tag{61}$$

which can be combined into

$$\Delta(P)^F f(\Delta(P)^F - \Delta(Q)) = 1 - 2F(\Delta(P)^F - \Delta(Q)). \tag{62}$$

In equilibrium, we have $0 < \Delta(P)^F < \Delta(Q)$ whenever $\Delta(Q) > 0$. This is because the RHS is weakly negative whenever $\Delta(P)^F \geq \Delta(Q)$, while the LHS is always positive when $\Delta(P)^F > 0$. Thus, $\Delta(P)^F < \Delta(Q)$. A similar argument can rule out $\Delta(P)^F < 0$.

Moreover, explicitly solving for P_1 and P_2 we obtain

$$P_1^F = \frac{1 - F(\Delta(P)^F - \Delta(Q))}{f(\Delta(P)^F - \Delta(Q))} \tag{63}$$

and

$$P_2^F = \frac{F(\Delta(P)^F - \Delta(Q))}{f(\Delta(P)^F - \Delta(Q))}. \tag{64}$$

Comparing to the no-information case, we have:

$$P_2^F < P_2^N = P_1^N < P_1^F. \tag{65}$$

To see this, notice that the monotone hazard rate implies

$$P_2^F = \frac{F(\Delta(P)^F - \Delta(Q))}{f(\Delta(P)^F - \Delta(Q))} < \frac{F(0)}{f(0)} = \frac{1}{2f(0)} = P_2^N,$$

and similarly $P_1^F > P_1^N$. ■

Proof of Lemma 4

First, to show that $\frac{\partial \Delta \mathcal{R}}{\partial \Delta(P)} > 0$, assume by contradiction that $\Delta(P^{\mathcal{R}})$ increases and $\Delta \mathcal{R}$ does not. Then, the LHS of Eq. 39 does not change, while the RHS increases, violating their equality.

To show that $\frac{\partial \Delta \mathcal{R}}{\partial \Delta(P^{\mathcal{R}})} < 1$, assume by contradiction it is not. Then, $\Delta(P^{\mathcal{R}}) - \Delta \mathcal{R}$ goes down (or remains unchanged) following an increase in $\Delta(P^{\mathcal{R}})$. Thus, as a whole, the RHS of Eq. 39 decreases. On the contrary, the LHS increases, again reaching a contradiction. ■

Proof of Lemma 5

From Eq. 39, we see that if $\Delta(P^{\mathcal{R}}) = \Delta(Q)$, then $\Delta \mathcal{R} = \Delta(Q)$. The result then follows immediately from Lemma 4. ■

Proof of Lemma 6

Suppose (by contradiction) that $\Delta(P^{\mathcal{R}}) \leq 0$. Then $\Delta(P^{\mathcal{R}}) < \Delta(Q)$, and by Lemma 5, $\Delta(P^{\mathcal{R}}) - \Delta \mathcal{R} < 0$. In this case, the LHS of Eq. (68) is positive but the RHS is weakly negative, a contradiction, so $\Delta(P^{\mathcal{R}}) > 0$. ■

Proof of Proposition 11

The first order conditions are given by

$$\frac{\partial \pi_1}{\partial P_1^{\mathcal{R}}} = 1 - F(\Delta(P^{\mathcal{R}}) - \Delta \mathcal{R}) - f(\Delta(P^{\mathcal{R}}) - \Delta \mathcal{R}) \cdot \left(1 - \frac{\partial \Delta(R)}{\partial \Delta(P^{\mathcal{R}})}\right) \cdot P_1 = 0, \tag{66}$$

$$\frac{\partial \pi_2}{\partial P_2^{\mathcal{R}}} = F(\Delta(P^{\mathcal{R}}) - \Delta \mathcal{R}) - f(\Delta(P^{\mathcal{R}}) - \Delta \mathcal{R}) \cdot \left(1 - \frac{\partial \Delta(R)}{\partial \Delta(P^{\mathcal{R}})}\right) \cdot P_2^{\mathcal{R}} = 0. \tag{67}$$

where we use the fact that $\frac{\partial \Delta \mathcal{R}}{\partial P_1^{\mathcal{R}}} = \frac{\partial \Delta \mathcal{R}}{\partial \Delta(P^{\mathcal{R}})} = -\frac{\partial \Delta \mathcal{R}}{\partial P_2^{\mathcal{R}}}$. Then, in equilibrium,

$$1 - 2F(\Delta(P^{\mathcal{R}}) - \Delta\mathcal{R}) = f(\Delta(P^{\mathcal{R}}) - \Delta\mathcal{R})\Big(1 - \frac{\partial\Delta\mathcal{R}}{\partial\Delta(P^{\mathcal{R}})}\Big) \cdot \Delta(P^{\mathcal{R}}) \tag{68}$$

Suppose (by contradiction) that $\Delta(P^{\mathcal{R}}) \geq \Delta(Q)$. By Lemma 5, $\Delta(P^{\mathcal{R}}) - \Delta\mathcal{R} \geq 0$. In this case, the LHS of Eq. (68) is weakly negative but the RHS is positive, a contradiction, so $\Delta(P^{\mathcal{R}}) < Q$.

Therefore, we have

$$0 < \Delta(P^{\mathcal{R}}) < \Delta\mathcal{R} = \tilde{\Delta}Q < \Delta(Q). \tag{69}$$

Solving the FOCs to derive prices, we get

$$P_1^{\mathcal{R}} = \frac{1 - F(\Delta(P^{\mathcal{R}}) - \Delta\mathcal{R})}{f(\Delta(P^{\mathcal{R}}) - \Delta\mathcal{R})\Big(1 - \frac{\partial\Delta\mathcal{R}}{\partial\Delta(P^{\mathcal{R}})}\Big)} \tag{70}$$

and

$$P_2^{\mathcal{R}} = \frac{F(\Delta(P^{\mathcal{R}}) - \Delta\mathcal{R})}{f(\Delta(P^{\mathcal{R}}) - \Delta\mathcal{R})\Big(1 - \frac{\partial\Delta\mathcal{R}}{\partial\Delta(P^{\mathcal{R}})}\Big)}. \tag{71}$$

■

Proof of Proposition 12

Taking derivatives w.r.t. $\Delta(P)$ on both sides of Eq. 39 and rearranging, we have

$$1 - \frac{\partial\Delta\mathcal{R}}{\partial\Delta(P)} = \Big(1 + \frac{d\mathbb{E}(\theta_1|\Delta\theta > \Delta(P) - \Delta\mathcal{R})}{dk} + \frac{d\mathbb{E}(\theta_2|\Delta\theta < \Delta(P) - \Delta\mathcal{R})}{dk}\Big)^{-1} \tag{72}$$

When $\Delta(Q)$ is close to 0, the previous analysis tells us that in equilibrium $\Delta(P^{\mathcal{R}}) - \Delta\mathcal{R}$ and $\Delta(P)^F - \Delta(Q)$ are close to 0. In contrast, $1 - \frac{\partial\Delta\mathcal{R}}{\partial\Delta(P)} < (1+\epsilon)^{-1}$. Therefore, we have

$$P_1^{\mathcal{R}} > \frac{1 - F(\Delta(P^{\mathcal{R}}) - \Delta\mathcal{R})}{f(\Delta(P^{\mathcal{R}}) - \Delta\mathcal{R})}(1+\epsilon) \approx \frac{1 - F(\Delta(P)^F - \Delta(Q))}{f(\Delta(P)^F - \Delta(Q))}(1+\epsilon) > P_1^F > P_1^N$$

and

$$P_2^{\mathcal{R}} > \frac{F(\Delta(P^{\mathcal{R}}) - \Delta\mathcal{R})}{f(\Delta(P^{\mathcal{R}}) - \Delta\mathcal{R})}(1+\epsilon) \approx \frac{F(0)}{f(0)}(1+\epsilon) > P_2^N > P_2^F.$$

■