Influence or Advertise: The Role of Social Learning in Influencer Marketing*

Ron Berman† Aniko Öry‡ Xudong Zheng§

January 15, 2023

*We thank the NET Institute (www.netinst.org) for financial support.
†The Wharton School, University of Pennsylvania, ronber@wharton.upenn.edu
‡Yale School of Management, aniko.oery@yale.edu
§Department of Economics, Johns Hopkins University, xzheng27@jhu.edu

Influence or Advertise: The Role of Social Learning in Influencer Marketing

Abstract

We compare influencer marketing to targeted advertising from information aggregation and product awareness perspectives. Influencer marketing leverages network effects by allowing consumers to socially learn from each other about their experienced content utility, but consumers may not know whether to attribute promotional post popularity to high content or high product quality. If the quality of a product is uncertain (e.g., it belongs to an unknown brand), then a mega influencer with consistent content quality fosters more information aggregation than a targeted ad and thereby yields higher profits. When we compare influencer marketing to untargeted ad campaigns or if the product has low quality uncertainty (e.g., belongs to an established brand), then many micro influencers with inconsistent content quality create more consumer awareness and yield higher profits. For products with low quality uncertainty, the firm wants to avoid information aggregation as it disperses posterior beliefs of consumers and leads to fewer purchases at the optimal price. Our model can also explain why influencer campaigns either "go viral" or "go bust," and how for niche products, micro-influencers with consistent content quality can be a valuable marketing tool.

Keywords: Influencer marketing, social learning, online advertising, targeting, word of mouth, social media marketing.

1 Introduction

Influencer marketing—where a marketer promotes a product by sponsoring posts of social media content creators with many followers—is capturing an increasing share of firm marketing budgets. In 2022 the market was expected to surpass 16 billion US dollars[1] where on Instagram alone brands spent over 8 billion dollars in 2020—an almost 100% increase from 2017.[2] Given the growth of this marketing approach, a natural question for marketers is whether they should shift their online marketing budgets from more established technologies (e.g., targeted display advertising) to this new medium. This question is particularly subtle given that a marketer can target advertisements towards followers of a specific social media influencer by paying the social media platform directly without contracting with the influencer to create original content. Given the recent drive for

[1] `https://influencermarketinghub.com/influencer-marketing-benchmark-report/`, Accessed September 26, 2022.

[2] `https://www.statista.com/statistics/950920/global-instagram-influencer-marketing-spending/`, accessed November 6, 2021.

increased consumer privacy and limitations on ad targeting, influencer marketing might serve as a viable alternative.

In this paper we study the differences between influencer and ad campaigns from a social learning perspective using a game-theoretic model. In particular, we focus on the unique feature of influencer campaigns which allows the followers of an influencer to interact with the influencer's posts—through comments, presses of the "like" button, and content sharing—and to see the response of other followers to the content posted by the influencer. This allows influencers to play the role of social signal aggregators for their followers and facilitate social learning when a product's quality or match is uncertain. In contrast, traditional online advertising, where consumers only observe the ads delivered to them, cannot facilitate the same degree of information exchange. In this context, we ask when it is optimal to substitute influencer campaigns for traditional (targeted) online advertising, and within influencer campaigns, which influencer attributes are more profitable for different product types.

A marketing campaign—both through an influencer and through advertising—achieves two objectives. First, it makes consumers aware of the product, and second, it provides valuable information to consumers. By facilitating social learning, influencers allow marketers to utilize a network effect, similar in spirit to word of mouth (Galeotti et al. 2013, Campbell 2013). Unlike word of mouth, however, where the marketers can entice consumers to speak to each other, e.g., through exclusivity or referral rewards (Campbell et al. 2017, Kamada and Öry 2020, Carroni et al. 2020), the influencer is encouraging interactions among the follower base by curating content and by leveraging the social media platform's content feed algorithm to generate exposure (Berman and Katona 2020). When evaluating whether to choose an influencer campaign over an advertising campaign, a key factor is to compare the information aggregation friction intrinsic to the campaign, and to evaluate whether the campaign should focus on creating awareness or on learning of consumers, given the ex-ante uncertainty about product quality.

In our model, a marketer can promote a product with uncertain quality or popularity through an influencer campaign or an ad campaign. If they choose an influencer campaign, the followers of an influencer see the promoted post and update their beliefs about the product's quality to make a purchase decision. The posterior belief of a consumer about the product's quality depends on a private signal generated from the post reflecting their liking of the post, and on the observed responses of other followers of the influencer to the post. If the marketer used a targeted ad campaign, the consumers update their beliefs about product quality based only on a private signal received from the ad they see. We model each signal as the sum of the post's content quality, the product quality, and an idiosyncratic taste component.

Learning about product quality through a campaign affects the distribution of willingness to pay of consumers which can decrease total purchases at the profit-maximizing price, but may also allow the marketer to charge a higher price if the option value from learning is high. How much consumers can learn from a campaign depends on how strong the learning frictions are, which may come from multiple sources. First, since influencers have flexibility in the content they create to promote a product, the content quality is affected by their creativity. Higher creativity, however, can be a double-edged sword as it can make the content quality of influencers less predictable and variable, which affects the inference of consumers about the promoted product's quality. For example, when a consumer sees a viral post by an influencer for a product, they might infer that the post is popular because of its content quality, and not because the product being presented is enticing. Second, influencers often build their followership by focusing on specific topics, resulting in more homogeneous content tastes of their follower base. A mismatch between the content preferences of the user and the influencer's post might be mistakenly attributed to lower product quality and would greatly affect the effectiveness of social learning. Third, the number of followers of an influencer determines the number of signals a follower observes from other followers. More followers decrease information frictions. By contrast, learning frictions operate differently for advertising campaigns. In these campaigns the degree of targeting determines how homogeneous the content taste of potential consumers is and affects their learning frictions (Shin and Yu 2021). We view this as the key determinant for information frictions coming from ads.

Furthermore, the network effect in influencer marketing has two unique features. First, the information learned by consumers is not linear in the number of followers, making micro influencers particularly valuable for making consumers aware of products, while making mega influencers particularly valuable for a marketer who wants to benefit from the network effects of social learning. We show that a campaign should focus on creating awareness when there is little uncertainty about product quality, e.g., for products from established brands. By contrast, when the product quality is very uncertain, e.g., for products from unknown brands, then social learning through a mega influencer with consistent content quality is more valuable than targeted advertising.

Second, as followers of an influencer observe each others' information, the homogeneity of consumer information sets increases and leads to ex-post more concentrated demand. As a result, influencer campaigns tend to "go viral" or "go bust", and this effect depends on how products are priced. In our main model we assume that the marketer chooses a price simultaneously with the campaign launch. Thus, the price cannot condition on whether the campaign is successful or not. If the initial price was set too high, social learning can cause most consumers to not buy. In that sense, an influencer campaign can be detrimental and "go-bust." However, a marketer can extract all rents from consumers by charging a single price *after* seeing whether the campaign leads to collectively high or low willingness to pay of consumers. Such ex-post pricing can be implemented, for example, with a discount code that is sent to the influencer if the product is less popular than expected. Thus, social learning offers an alternative reason for the profitability of influencer promo codes studied in Jiang et al. (2021).

Finally, we consider the benefit to consumers from influencer campaigns in terms of consumer surplus, and extend our analysis to a model where consumers have heterogeneous product tastes. On consumer surplus, we find that learning frictions might generate a non-monotone effect, and that sometimes (but not always), some level of learning friction might be better for consumers. With heterogeneity in consumption utility of the product, we show that most of our results generalize except for one: While creative mega influencers are preferred for promoting mass market products—when consumers have homogeneous tastes—for niche products, consistent micro influencers are preferred. The reason is that in niche markets there is a lot of option value from learning about the product's quality, but social learning is less valuable as other consumers evaluate the product itself differently because of taste heterogeneity.

Our paper makes two important contributions to the research on influencer marketing and social learning. First, we focus on the tradeoffs when choosing between targeted ad campaigns and influencer campaigns, a decision many marketers face today. Second, our findings capitalize on the social learning aspect of influencer marketing and are complementary to the view that influencers are more persuasive than ads because of the influencer's personality or authenticity. Other research on influencer marketing analyzed aspects such as the interaction between an influencer's idiosyncratic preferences and the product variety of a firm (Kuksov and Liao 2019); the optimal level of affiliation of a marketer with social media influencers and its implications on consumer welfare (Pei and Mayzlin 2022); and the optimal regulation regarding disclosure of an influencer's affiliation with a marketer (Fainmesser and Galeotti 2021, Mitchell 2021). By investigating the social learning aspect of campaigns we introduce novel criteria for evaluating the effectiveness of advertising versus influencer campaigns.

The social learning aspect we analyze also relates to the extensive literature on word of mouth and observational learning. Examples of research in this area include Banerjee (1992), Bikhchandani et al. (1992) and Zhang (2010). Similarly to us, Crapis et al. (2017), Ifrach et al. (2019) and Fainmesser et al. (2021) study how consumers learn from the signals of others. Their setting is product reviews which are posted conditional on purchase, and hence the price of the product affects the amount and quality of learning. A similar mechanism is also analyzed in Nistor and Selove (2022), but this time in the context of influencer marketing. In contrast to these papers, learning occurs simultaneously in our setting, which can explain the potential "virality" of influencer posts and their effect on product demand.

Our findings also have potential empirical implications, because they provide insight and predictions about when one should expect influencer campaigns to be beneficial, and how these benefits will appear in consumer behavior. The effectiveness of advertising campaigns is notoriously hard to measure, and this problem becomes even harder for influencer campaigns. Our insights can help marketers to determine not only which type of campaign to launch, but can also help researchers devise new methods to estimate the value of social learning and their effects in these campaigns.

The paper proceeds to present the model and its timing in Section 2, followed by the equilibrium analysis in Section 3. Section 4 compares the profitability of advertising to influencer campaigns. Section 5 summarizes the results and considers consumer surplus. Section 6 extends the model to allow consumers to have heterogeneous tastes for products. The analysis focuses on determining which products should be promoted, and through which influencers. Section 7 discusses our findings in a broader context and concludes. The proofs of formal results appear in the Appendix.

2 Model

Overview. A firm (marketer) would like to sell a product of unknown quality q_p — distributed according to $q_p \sim \mathcal{N}(\mu_p, \sigma_p^2)$. The coefficient of variation of this distribution $\frac{\sigma_p}{\mu_p}$ captures how unknown the brand is. A product with high σ_p comes with a lot of uncertainty about quality but a high upside potential, for example, because it is produced by a new innovative firm without too much track record. A low σ_p implies that the brand is already established and consumers have a good idea about the quality of the product. μ_p normalizes this standard deviation. We normalize the marginal cost of production to zero. The firm chooses an optimal price m and a marketing campaign type. The marketing campaign has two purposes: First, it makes potential consumers aware of the product, and second, it provides consumers with information about the potential utility they will receive from buying the product.

There are two types of marketing campaigns denoted by c — a direct (targeted) advertising campaign ($c = \text{ad}$) and an influencer campaign ($c = \text{inf}$). If the firm chooses a direct advertising campaign with reach N_{ad}, each of the N_{ad} consumers sees an ad by the firm which is informative about the quality of the product q_p. If the firm partners with an influencer, the influencer promotes the firm's product to its N_{inf} followers. The key difference between the two marketing strategies is that consumers can respond to the influencer's content—e.g., by pressing the "like" button or commenting on the post—and can also see the responses of other followers of the influencer.[3] As a result, each consumer can learn about the quality of the product from other consumers' reactions with the details provided below. We will refer to advertising campaigns as ad campaigns for brevity.

Consumption and content utility. Each consumer i experiences two types of utility—utility from consuming the product (consumption utility u_i) and utility from consuming content (content utility v_i). In the main model all consumers receive the same consumption utility $u_i = q_p$ if they buy the product at price m. This allows us to focus on the difference between ad and influencer campaigns from an information aggregation perspective. We extend our model to heterogeneous product utilities in Section 6.

The content utility from a promotional influencer post or an ad is experienced separately from purchase. It depends on both the quality of the promoted product and the quality of the content itself. Specifically, consumer i's content utility is given by

$$v_i = q_c + q_p + \epsilon_i, \tag{1}$$

[3] A few social media platforms do allow interactive responses to ads. However, the response rate is very low and viewers of the ads rarely know anything about other viewers tastes and cannot learn from them.

which is the sum of content quality q_c, product quality q_p, and an idiosyncratic taste shock, ϵ_i. We assume that $q_c \sim \mathcal{N}(\mu_c, \sigma_c^2)$, and $\epsilon_i \sim \mathcal{N}(0, \tau_c^2)$, where the value of μ_c, σ_c^2 and τ_c^2 depend on the source of the content (influencer or ad campaign). To simplify exposition, we focus on some specific values of μ_c, σ_c^2 and τ_c^2 for each campaign type, but the analysis is generalized in the Appendix. We assume that $\mu_c = 0$ for both campaign types. This assumption is without loss of generality, as the Appendix shows that the marketer cannot increase their profit using campaigns with higher average content quality.

The main differences between influencer and ad campaigns will be the source of variability of the content utility σ_c and the source of variability of content preferences τ_c. For influencer campaigns, the marketer has control over which influencer to work with, but uncertainty about the quality of content may remain. The variance σ_{inf}^2 of content quality $q_c = q_{\text{inf}} \sim \mathcal{N}(0, \sigma_{\text{inf}}^2)$ can be interpreted as the *creativity of the influencer* since the willingness to try new and varied approaches to social media posts by the influencer may result in high variance of quality of posts. For ad campaigns, we assume that there is no variability in content quality and set $\sigma_{\text{ad}} = 0$.

With respect to idiosyncratic content preferences, a heterogeneous content taste (τ_c large) means that a mismatch between the content preferences of the user and the content might be attributed to lower product quality. The variance of the idiosyncratic content taste τ_{inf} captures *follower heterogeneity* with respect to the content of the influencer. We normalize this value to 1.

For ad campaigns the degree of targeting may affect the heterogeneity of content preferences. The consumer's idiosyncratic utility from ad content is drawn from $\epsilon_i \sim \mathcal{N}(0, \tau_{\text{ad}}^2)$, where the precision $\frac{1}{\tau_{\text{ad}}^2}$ measures the similarity in content preference between the consumers. Hence, it measures how *targeted* the advertising is. If $\tau_{ad}^2 > 1$, then the ad campaign is less targeted than the influencer campaign, and if $\tau_{ad}^2 < 1$, it is more targeted.

Note that an individual consumer's content utility can be high because the content quality is high, because the product quality is high or because it matches their taste well, or a combination of those. The key premise is that it is impossible for the consumers to disentangle what the source of liking of content is.

Consumer learning and purchase decisions. A consumer's willingness to pay for the product depends on her posterior belief about $u_i = q_p$ after observing and interacting with the influencer campaign, or after seeing an ad. If the firm chooses an influencer campaign, each follower i is able to see the content utility of all other followers generating the information set $\mathbf{v}_i^{\text{inf}} = \{v_1, \ldots, v_{N_{\text{inf}}}\}$. If the firm chooses an ad campaign, each consumer observes only the single ad displayed to them, and the information they can use to update their beliefs is the singleton set $\mathbf{v}_i^{\text{ad}} = \{v_i\}$. For an arbitrary information set $\mathbf{v}$, we denote the size of the set (i.e., the number of observed content utility values) by $n = |\mathbf{v}|$, i.e., $n = N_{\text{inf}}$ for influencer campaigns and $n = 1$ for ad campaigns. Further, we denote the generic set of customers who are affected by the campaign by N, so that $N = N_{\text{inf}}$ for an influencer campaign and $N = N_{\text{ad}}$ for an ad campaign. All in all, we have:

1. For influencer campaigns $(\sigma_c^2, \tau_c^2, n, N) = (\sigma_{\text{inf}}^2, 1, N_{\text{inf}}, N_{\text{inf}})$.

2. For ad campaigns $(\sigma_c^2, \tau_c^2, n, N) = (0, \tau_{\text{ad}}^2, 1, N_{\text{ad}})$.

After observing the information in $\mathbf{v_i^c}$, consumers update their belief about the product quality q_p and make a purchase if they expect the posterior consumption utility to be weakly higher than the price, i.e., if

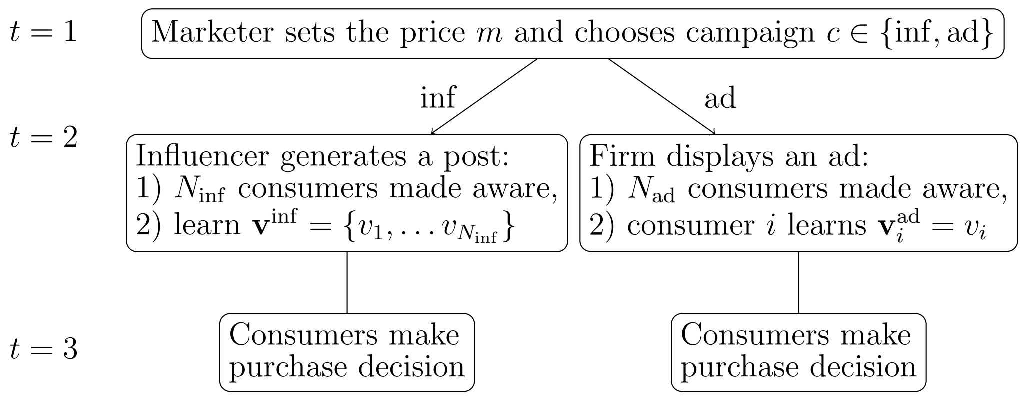
$$\mathbb{E}_{q_p}[u_i|\mathbf{v_i^c}] - m \geq 0. \tag{2}$$

Timing. The game proceeds in 3 time periods, as depicted in Figure 1. At $t = 1$, the marketer sets a price m and decides whether to promote the product through targeted advertising or through influencer marketing. At $t = 2$, a campaign $c \in \{\text{ad}, \text{inf}\}$ generates a post with quality q_c. Each consumer i observe the realized content utilities in $\mathbf{v}_i^c$, and update their beliefs about q_p. Then, in period $t = 3$, consumers decide whether to purchase the product or not, and their consumption utilities and marketer profits are realized. Our analysis focuses on finding subgame-perfect equilibria.

3 Equilibrium analysis

In this section we characterize the equilibrium behavior of consumers and the marketer by backward induction. First, we determine the demand the marketer faces. Then we characterize the profit-maximizing price and the profit of the marketer for each campaign.

Figure 1: Game timeline



3.1 Consumption decision ($t = 3$)

In order to consume the product, consumers need to be aware of it and want to buy it. Conditional on being aware, the purchasing decision of a consumer depends on their posterior expected quality of the product after observing the information set $\mathbf{v}_i^c$. We denote this posterior expectation by $\mu_p(\mathbf{v}_i^c) = \mathbb{E}_{q_p}[u|\mathbf{v}_i^c]$. Since by (2) a consumer i purchases if $\mu_p(\mathbf{v}_i^c) \geq m$ for $c \in \{\text{inf}, \text{ad}\}$, the expected demand for the product given a campaign c is:

$$D_c(m) = \sum_{i=1}^{N_c} Pr(\mu_p(\mathbf{v}_i^c) \geq m). \tag{3}$$

It is worth noting that for an ad campaign, each consumer independently observes a different realization of v_i, resulting in different realized values of posterior means $\mu_p(\mathbf{v}_i^c)$. Hence, each consumer has a different realized willingness to pay for the product. In contrast, for influencer campaigns, all consumers observe the same information set $\mathbf{v}_i^{\text{inf}} = \mathbf{v}_1^{\text{inf}}$. Hence, each consumer's willingness to pay is exactly the same ex-post. Consequently, at any given price influencer campaigns cause the product to either "go viral" with everyone buying it, or to "go bust" with no one buying it.[4] Since we assume that the marketer has to commit to a price m ex-ante, they cannot price based on the realized demand. In Section 7 we discuss in which situations this feature of influencer marketing can increase marketer profits significantly if the marketer can provide the influencer with a price

[4] In reality not all consumers will buy the product because of heterogeneous preferences and information sets. Nevertheless, demand will be more concentrated because of social learning. We analyze this more realistic case in Section 6.

discount (e.g., through a coupon code) based on the success of their post.

3.2 (Social) learning stage ($t = 2$)

In order to characterize the expected posterior belief about product quality $\mu_p(\mathbf{v}_i^c)$, we denote the average content utility of all items in an information set $\mathbf{v}$ by $\bar{\mathbf{v}} = \frac{\sum_{v\in\mathbf{v}} v}{n}$. We also define an auxiliary quantity, the *relative learning friction* of a campaign c as κ_c^2. It captures the main differences between campaign types that drive demand and profit. For influencer campaigns

$$\kappa_{\text{inf}}^2 = \frac{\sigma_{\text{inf}}^2 + \frac{1}{N_{\text{inf}}}}{\sigma_p^2} + 1$$

while for ad campaigns

$$\kappa_{\text{ad}}^2 = \frac{\tau_{\text{ad}}^2}{\sigma_p^2} + 1.$$

In the Appendix we derive demand using the general expression $\kappa_c^2 = \frac{\sigma_c^2 + \frac{\tau_c^2}{n}}{\sigma_p^2} + 1$ for a campaign of type c. We elaborate on how κ_c affects consumer learning and demand after the following demand characterization:

Proposition 1 (Demand)**.**

In a campaign $c \in \{\text{inf}, \text{ad}\}$, the posterior belief about the product's quality $q_p\big|\mathbf{v}_i^c$ of consumer i is normally distributed with mean

$$\mu_p(\mathbf{v}_i^c) \quad = \mu_p + \frac{\bar{\mathbf{v}}_i^c - \mu_p}{\kappa_c^2}. \tag{4}$$

The expected demand when the marketer sets price m is

$$D_c(m) = N\left(1 - \Phi\left(\frac{m - \mu_p}{\sigma_p/\kappa_c}\right)\right), \tag{5}$$

where $\Phi(\cdot)$ is the CDF of the standard Normal distribution.

The proposition summarizes how the posterior belief of the consumers and the expected demand of the marketer vary with the product and campaign parameters—through the learning friction κ_c. The expected posterior product quality is given by the sum of the prior mean of product quality μ_p and the strength of the campaign information signal $\frac{\bar{\mathbf{v}}_i^c - \mu_p}{\kappa_c^2}$.

The learning friction κ_c prevents the consumer from learning about the quality of the product perfectly. It measures the informativeness of the campaign relative to the product quality uncertainty. There are two factors that affect the learning friction κ_c: First, learning is imperfect due to the finite number of signals n in $\mathbf{v}_i^c$. We call this friction the *finite information friction.* Importantly, for an advertising campaign $n = 1$, so that the finite information friction is always significant. Second, even if the consumer observed infinitely many signal realizations in $\mathbf{v}_i^c$, e.g., because the influencer has a very large number of followers, learning can be imperfect. In this case the learning friction κ_c^2 equals $\frac{\sigma_c^2}{\sigma_p^2} + 1$ and remains greater than one if $\sigma_c^2 > 0$. As a result, $\lim_{n\to\infty} \mu_p(\mathbf{v}_i^c) = \mu_p + \frac{q_p+q_c-\mu_p}{\kappa_c^2} \neq q_p$ and the consumer never exactly learns the true quality of the product.[5] The reason is that the consumer cannot be sure whether to attribute the high content utility to the content quality q_c or to the product quality q_p. We call this the *attribution friction* of social learning. In the presence of the finite information friction, there is an additional idiosynchratic attribution friction due to τ_c. We compare the learning friction for advertising and influencer campaigns in detail in Lemma 1 below.

The overall dispersion of demand ($\frac{\sigma_p}{\kappa_c}$) captures how differentiated consumer preferences are after learning, and hence how much surplus the marketer can extract with a single price. It is determined by the ex-ante uncertainty about product quality σ_p and how much consumer preferences disperse due to learning. If the standard deviation $\frac{\sigma_p}{\kappa_c}$ is large, the marketer can extract little surplus with a single price, while with small $\frac{\sigma_p}{\kappa_c}$, the marketer can extract more consumer surplus.

Furthermore, note that demand decreases in the price m as we would intuitively expect. Interestingly, we show in the Appendix that the expected demand is independent of the average content quality μ_c for both influencer and ad campaigns. The reason is that the marketer cannot extract surplus from content utility through product sales, because μ_c does not affect the expected posterior belief about the product quality $\mu_p(\mathbf{v}_i^c)$. This means that marketers will not find an advantage for influencer vs. ad campaigns based on different levels of average content quality. It also provides a rationale for assuming $\mu_c = 0$ without loss of generality.

Whether demand is increasing or decreasing in the information friction κ_c depends on whether the price m is larger or smaller than the prior belief about product quality μ_p. Expression (5) shows that when $m > \mu_p$, the expected demand is decreasing in κ_c. If $m < \mu_p$, demand is increasing in

[5] By the Law of Large Numbers $\lim_{n\to\infty} \bar{\mathbf{v}}_i^c = q_c + q_p$.

κ_c. In that case, learning on average convinces consumers not to buy.

Corollary 1.1. *If the price is lower than the average product quality ($m < \mu_p$), then demand $D_c(m)$ is increasing in κ_c. If $m > \mu_p$, $D_c(m)$ is decreasing in κ_c.*

When we compare an ad campaign with an influencer campaign, one important comparison is between the learning frictions. As discussed before, the learning friction of an ad campaign is driven by the accuracy $1/\tau_{\text{ad}}$, while for influencer campaigns, the creativity of the influencer σ_{inf} and the number of followers N_{inf} matter. We summarize the comparison between the learning frictions in influencer vs. ad campaigns in the following lemma:

Lemma 1. *An advertising campaign has a larger relative learning friction than an influencer campaign when $\tau_{\text{ad}}^2 > \sigma_{\text{inf}}^2 + \frac{1}{N_{\text{inf}}}$.*

Lemma 1 highlights that learning frictions between the two marketing campaigns do not depend on product characteristics when the products are homogeneous, but can still be different due to the inherent features of the learning process. It also emphasizes that network effects of influencer campaigns (large N_{inf}) decrease the learning friction, but that the creativity, or inconsistency, of influencers increases the learning friction. By contrast, the reach N_{ad} of an advertising campaign does not affect the learning friction at all. For an ad campaign, higher reach only makes more consumers aware of the product without affecting learning. For an influencer campaign, the impact of reach on profits is more nuanced as we discuss in the next section.

3.3 Marketer profit-maximization ($t = 1$)

Based on the expected market demand, the firm sets a price m to maximize the expected profit:

$$\max_m m \cdot D_c(m) \tag{6}$$

The optimal price m_c^* cannot be expressed in closed form, but can be characterized as a function of m_c^0 which we define implicitly as the unique solution to:

$$m_c^0 - \frac{1}{h(m_c^0)} = -\frac{\mu_p}{\sigma_p}\kappa_c, \tag{7}$$

where $h(\cdot) = \frac{\phi(\cdot)}{1-\Phi(\cdot)}$ is the increasing hazard function of the standard normal distribution. The right-hand side of the equation is determined by two factors: the level of uncertainty of the product's quality $\frac{\sigma_p}{\mu_p}$ (i.e., whether the brand is established) and the relative learning friction κ_c. Using m_c^0, we characterize the optimal price m_c^* as follows:

Proposition 2 (Profit-maximizing price)**.**

The unique profit-maximizing price m_c^* is given by

$$m_c^* = \mu_p + \frac{\sigma_p}{\kappa_c} m_c^0. \tag{8}$$

The price m_c^* exceeds the a priori expected product quality μ_p if and only if

$$\sqrt{\frac{\pi}{2}} \frac{\sigma_p}{\mu_p} > \kappa_c. \tag{9}$$

Proposition 2 shows how the optimal price m_c^* depends on the relationship between the product's quality uncertainty $\frac{\sigma_p}{\mu_p}$ and the relative learning friction κ_c. In particular, the optimal price can be larger or smaller than the prior expected quality μ_p depending on their relationship. For example, when the quality is very uncertain there is a lot of option value from learning about the product's quality, and if the learning friction is small, then this option value can be exploited. As a result, the marketer can charge a price higher than μ_p.

Similarly, the optimal price is increasing or decreasing in information frictions depending on the level of quality uncertainty. For products with little uncertainty, e.g., products from established brands, more information friction allows the marketer to sell to more consumers at a higher price because there is little option value from learning. In contrast, for products with high uncertainty about their quality, e.g., products from unknown brands, reducing information frictions, e.g., by choosing an influencer with a higher reach N, can allow the marketer to charge a significantly higher price. This is summarized in the following corollary.

Corollary 2.1 (Comparative statics of profit-maximizing price with respect to κ_c)**.**

The profit-maximizing price m_c^* increases in κ_c when $\lambda \frac{\sigma_p}{\mu_p} < \kappa_c$, and decreases in κ_c otherwise, where $\lambda \approx 3.7 > \sqrt{\frac{\pi}{2}}$.[6]

[6] The exact value of λ appears in the proof in the Appendix.

By plugging-in the equilibrium price into the expression for the demand, we we can write equilibrium demand as $1-\Phi\left(m_c^0\right)$. Since m_c^0 is decreasing in κ_c by Equation (7), we can state the following result:

Corollary 2.2 (Comparative statics of equilibrium demand with respect to κ_c)**.**

The equilibrium demand increases in κ_c.

Finally, we can analyze how the marketer's optimal profit changes with κ_c by plugging in the equilibrium prices and demands into (6):

$$\Pi_c^* = N_c \cdot \pi_c^*(\kappa_c) = N_c \left(1 - \Phi\left(\frac{m_c^* - \mu_p}{\sigma_p/\kappa_c}\right)\right) m_c^*. \tag{10}$$

The analysis shows the following comparative statics in κ_c:

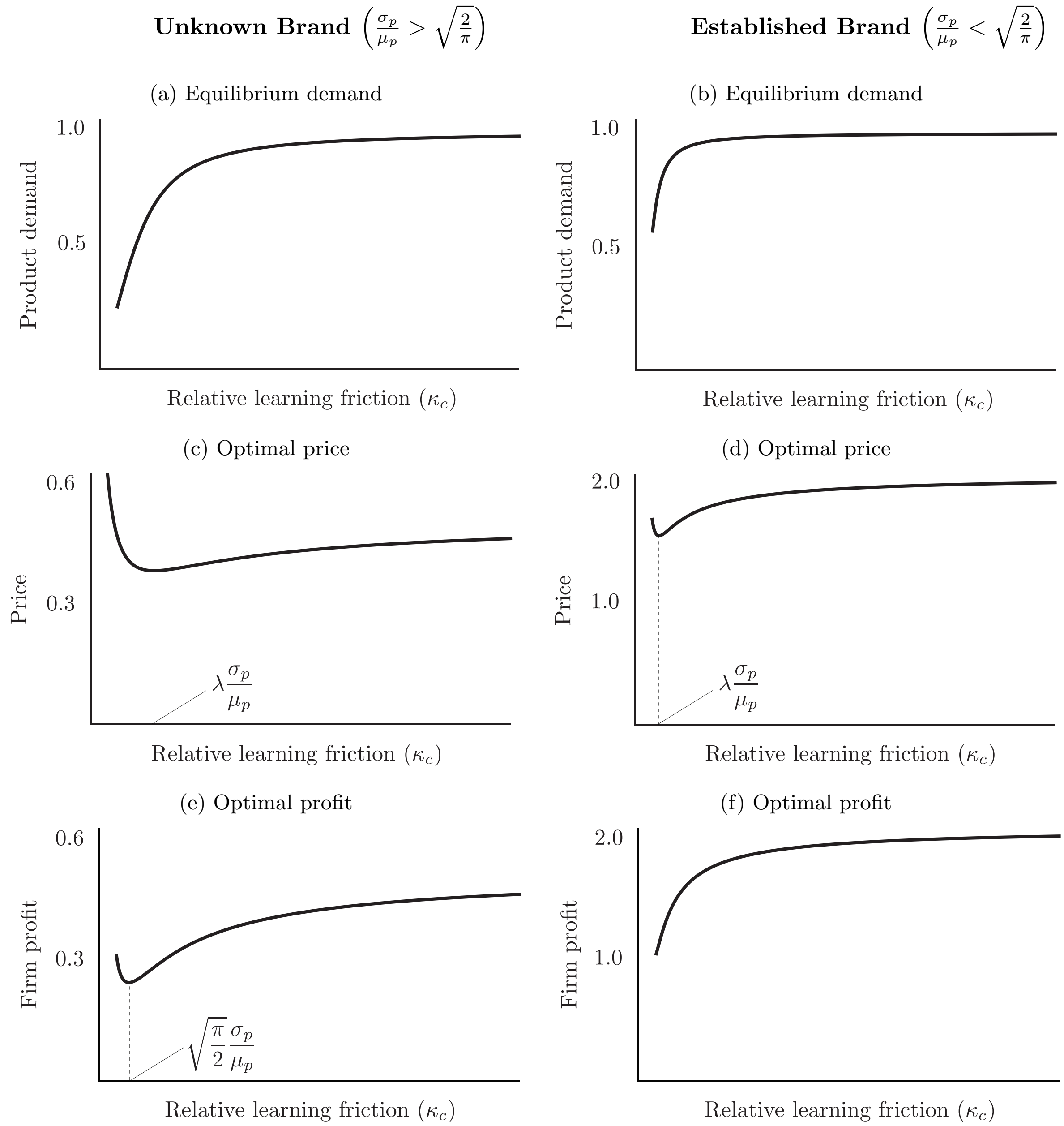
Corollary 2.3 (Comparative statics of profits with respect to κ_c)**.**

The optimal profit decreases in κ_c if (9) *is satisfied, and otherwise increases in κ_c.*

Extremely high learning frictions allow the marketer to extract all consumer surplus by charging $m = \mu_p$ absent any learning because buyers have a common prior about the product quality μ_p. Given that some learning occurs, the degree to which the posteriors about product quality disperse depend on the information friction and ex-ante product quality uncertainty. For an established brand with low product quality uncertainty, posteriors about quality remain relatively concentrated, while for a new brand with high product uncertainty, posteriors about product quality are always dispersed as we have seen in Proposition 1. This intuition results in the above comparative statics which we discuss below.

Figure 2 illustrates these different effects of the learning friction κ_c, where the left panels illustrate the case for less established brands with high $\frac{\sigma_p}{\mu_p} > \sqrt{\frac{2}{\pi}}$ and the right panels illustrate the case with low $\frac{\sigma_p}{\mu_p} < \sqrt{\frac{2}{\pi}}$, i.e., for established brands. For established brands, κ_c is always higher than $\sqrt{\frac{\pi}{2}}\frac{\sigma_p}{\mu_p} < 1$, and by Proposition 2, the price m_c^* is always lower than μ_p. In this case, demand increases with learning frictions, because learning more about established products leads consumers to buy less if the price is not adjusted downwards. As a result, the marketer would like to run a campaign only to make consumers aware of the product without them learning much about its quality. Mathematically, this follows from the envelope theorem: the effect of learning friction on

Figure 2: Equilibrium demand, price and profit at the optimal price as a function of κ_c



Notes: The left panels display the outcomes for less-established brands with $\mu_p = 1/2$ and $\sigma_p = 1$ resulting in high σ_p/μ_p. The right panels display the outcomes for established brands with $\mu_p = 2$ and $\sigma_p = 1$ resulting in low σ_p/μ_p. The demand functions are per-person demands given by the probability of purchase.

profit only depends on how demand is directly affected by learning frictions, which implies that for established brands learning frictions always increase profits.

For less established brands, profits are decreasing in κ_c for $\kappa_c < \lambda \frac{\sigma_p}{\mu_p}$ and increasing in κ_c for $\kappa_c > \lambda \frac{\sigma_p}{\mu_p}$. This is because for small κ_c condition (9) is satisfied, yielding an optimal price $m_c^* > \mu_p$ that is decreasing in κ_c. Intuitively, as discussed above, this effect is driven by the option value of learning if consumers know little about product quality ex-ante, which the firm can exploit if the learning friction is small: Lower frictions result in higher demand for any given price, allowing the firm to increase the equilibrium price significantly. This price effect dominates the effect of decreased equilibrium demand from lower κ_c. For large κ_c, the option value cannot be exploited causing both equilibrium demand and prices to increase with learning frictions.

4 Profitability of advertising versus influencer campaigns

In this section, we use the equilibrium characterization of Section 3 in order to compare the profitability of influencer and ad campaigns from an information aggregation perspective. As highlighted in the equilibrium characterization, the differences in the information aggregation between influencer and ad campaigns is captured by the differences in the relative learning friction κ_c^2. The degree of targeting of an ad $\frac{1}{\tau_{\text{ad}}}$ and the variability of content quality σ_{inf}^2 of an influencer affect the profit function in Equation (10) only through κ_c^2. However, due to the social learning role of influencers, the number of followers N_{inf} affects both the information friction κ_c^2 and the number of consumers who are being made aware of the product. The ad reach N_{ad}, by contrast, only affects the number of consumers made aware.

Hence, we first compare ad and influencer campaigns if the reach/follower bases are equal and fixed, i.e., if $N_{\text{ad}} = N_{\text{inf}}$. Then, we explore the profitability of many micro-versus one mega-influencer relative to an ad, where the profit with many micro-influencers can be thought of as the sum of profit functions of the form (10).

4.1 Advertising targeting versus influencer consistency

Since the degree of targeting of an ad $\frac{1}{\tau_{\text{ad}}}$, and the creativity of an influencer σ_{inf} only enter the profit function through the information friction κ_c, we can use the comparative statics result from Corollary 2.3 to compare ad campaigns versus influencer campaigns along these two dimensions.

We summarize this comparison in Proposition 3 and illustrate them in Figure 3.

Proposition 3 (Advertising vs. influencer marketing). *Suppose that $N_{\text{inf}} = N_{\text{ad}}$. Then:*

1. *For an established brand $\left(\frac{\sigma_p}{\mu_p} < \sqrt{\frac{2}{\pi}}\right)$ an influencer campaign is more profitable than an advertising campaign if and only if $\kappa_{\text{inf}} > \kappa_{\text{ad}}$.*

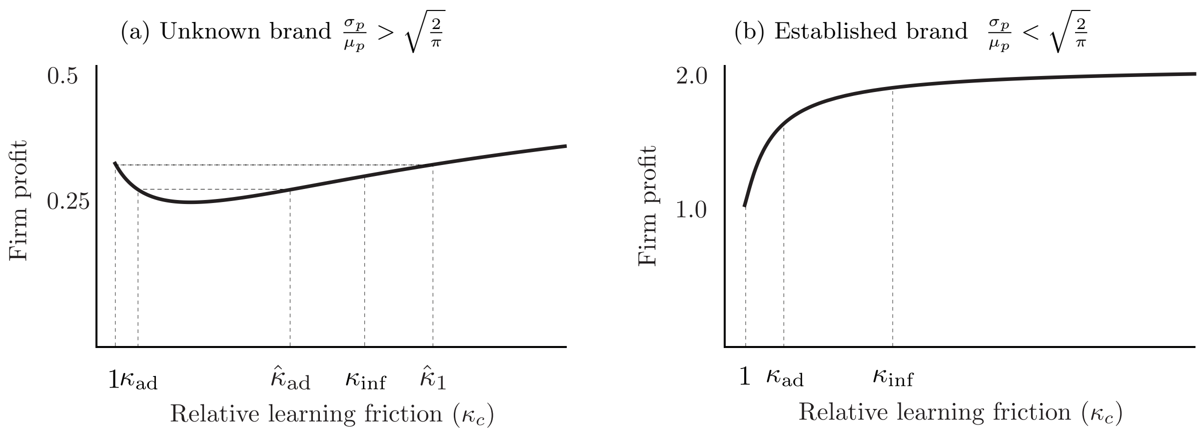
2. *For an unknown brand $\left(\frac{\sigma_p}{\mu_p} > \sqrt{\frac{2}{\pi}}\right)$ we distinguish between two cases:*

   (a) Compared to a broad ad technology with $\kappa_{\text{ad}} > \hat{\kappa}_1$, an influencer campaign is more profitable than an advertising campaign when $\kappa_{\text{inf}} > \kappa_{\text{ad}}$;

   (b) Compared to a targeted ad technology with $\kappa_{\text{ad}} < \hat{\kappa}_1$, an influencer campaign is more profitable than an advertising campaign when either (i) $\kappa_{\text{inf}} < \min\{\kappa_{\text{ad}}, \hat{\kappa}_{\text{ad}}\}$ or (ii) $\kappa_{\text{inf}} > \max\{\kappa_{\text{ad}}, \hat{\kappa}_{\text{ad}}\}$.

$\hat{\kappa}_{\text{ad}} \neq \kappa_{\text{ad}}$ is the friction level that yields the same profit as κ_{ad}, all else being equal, and $\hat{\kappa}_1$ is the friction level $\hat{\kappa}_1 > 1$ that yields the same profit as $\kappa = 1$.

Figure 3: Optimal per-consumer profits as a function of κ_c



Notes: Panels (a) and (b) correspond to the profit charts in panels (e) and (f) of Figure 2, where we add the cutoffs for κ_{inf} and κ_{ad} to illustrate Proposition 3. Panel (a) uses a shorter range of κ_c on the horizontal axis to better illustrate the effects of κ_c. Both panels depict the profit per customer of campaigns as a function of learning friction κ_c. The left panel shows the effect for less established brands with high σ_p/μ_p ($\mu_p = 1/2$, $\sigma_p = 1$) while the right panel shows it for established brands with low σ_p/μ_p ($\mu_p = 2$, $\sigma_p = 1$). κ_{ad} and κ_{inf} indicate example values for frictions of ad and influencer campaigns. $\hat{\kappa}_{\text{ad}}$ is the level of κ not equal to κ_{ad} that generates the same profit as κ_{ad}. $\hat{\kappa}_1$ is the level of $\kappa \neq 1$ that generates the same profit as $\kappa = 1$.

For established brands (Figure 3(b)), higher frictions increase profits by Corollary 2.3. Thus an influencer campaign is more profitable than an ad campaign if $\kappa_{\text{inf}} > \kappa_{\text{ad}}$. This condition can be interpreted using Lemma 1. It is satisfied for influencer campaigns with large attribution frictions driven by the variability of content quality σ^2_{inf}, e.g., due to creativity of the influencer. It is also satisfied if the ad campaign is very targeted (small τ_{ad}), as targeting decreases the profitability of ad campaigns for established brands. Although it might seem counter-intuitive that more targeted campaigns are less profitable, this follows from the fact that less learning about the product's quality allows the marketer to charge a higher price. This effect occurs with low product quality uncertainty $\frac{\sigma_p}{\mu_p}$, i.e., when the option value from learning about the product quality is low. In fact, in this case, the firm would ideally want to run an ad campaign that is completely uninformative and broad, and thereby only creates awareness for the product.

With unknown brands (Figure 3(a)), a similar logic applies if the advertising campaign is broad (case 2(a)). In that case, an influencer campaign is more profitable if the influencer is sufficiently creative or inconsistent. Compared to a targeted ad, the option value of learning from an influencer campaign is high. Thus, an influencer campaign with very low learning friction (case 2(b)) is more profitable. This is the case if the influencer has low variability of content quality σ^2_{inf}. With such low creativity the demand for any given price might be lower, but the increase in learning increases the willingness to pay of many consumers, allowing the marketer to charge a higher price. In this way, the marketer can exploit the option value from learning for unknown brands.

To summarize, the marketer has a variety of options to exploit influencer campaigns to either increase awareness without losing much pricing power, or to achieve higher profits by facilitating social learning, which allows the marketer to charge a higher price.

4.2 Micro versus mega influencers

A common decision for marketers when designing influencer campaigns is whether to engage micro- or mega-influencers in their campaigns. The reach N_{ad} of an ad campaign is irrelevant for per-consumer profit since the learning friction is not affected by the reach of ads. However, the per-consumer profit is different for micro-influencers with few followers versus mega-influencers with many followers. Hence, the marketer can create the same awareness with many micro-influencers or a few mega-influencers, but the degree of learning changes in these two scenarios.[7] Figure 4 compares per-customer ad campaign and influencer campaign profits as a function of the campaign's

[7]For simplicity we assumed that the micro-influencers have non-overlapping follower bases and that a follower of one micro-influencer cannot learn from the followers of another.

reach N. One can see that for ad campaigns the profit per consumer does not depend on N_{ad}, but for influencer campaigns, the effect of N_{inf} on profit depends on the type of product being promoted. This is because more followers always reduce the learning friction, but as Corollary 2.3 showed, learning frictions have a nuanced effect on profits.

For established brands (Figure 4(a)) we already showed that higher learning frictions are better. As the followership of influencers only decreases learning frictions, this means that established brands would generally benefit from influencers without too many followers, ceteris paribus.

For unknown brands, a larger follower base might increase or decrease profits, depending on whether a higher or a lower learning friction is desirable. The three cases are illustrated in Figures 4(b) – 4(d). First, by Corollary 2.3, π^*_{inf} is decreasing in N_{inf} everywhere (Figure 4(b)) for brands with little quality uncertainty:

$$\sqrt{\frac{\pi}{2}}\frac{\sigma_p}{\mu_p} < \sqrt{\frac{\sigma^2_{\text{inf}}}{\sigma^2_p}+1} = \min_n \underbrace{\sqrt{\frac{\sigma^2_{\text{inf}}+\frac{1}{n}}{\sigma^2_p}+1}}_{=\kappa_{\text{inf}}}\,.$$

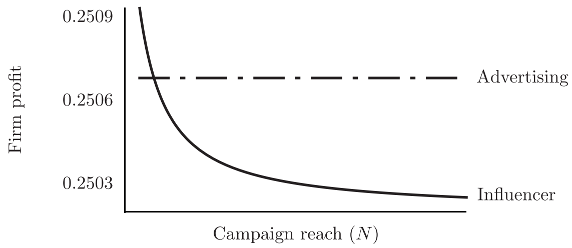
π^*_{inf} is increasing in N_{inf} everywhere (Figure 4(d)) for brands with high quality uncertainty:

$$\sqrt{\frac{\pi}{2}}\frac{\sigma_p}{\mu_p} > \sqrt{\frac{\sigma^2_{\text{inf}}+1}{\sigma^2_p}+1}.$$

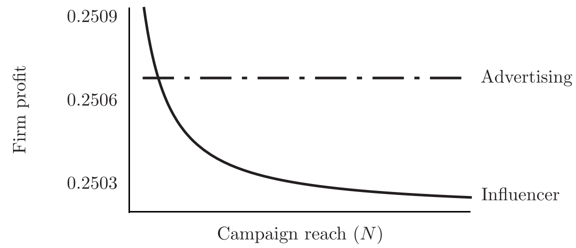
Consequently, for brands with little quality uncertainty micro-influencers are always better because they increase the per-customer learning friction. For brands with a lot of quality uncertainty, mega-influencers are always better because they allow the consumers to learn the most and utilize the option value of learning. Finally, for brands with intermediate levels of uncertainty (Figure 4(c)), both micro- and mega-influencers can be more profitable than ad campaigns on a per-customer basis. Micro-influencers tend to be more profitable if they are also creative, and mega-influencers tend to be more profitable if they are more consistent. Influencers with intermediate follower base size and creativity, and hence intermediate relative learning friction κ_{inf}, tend to have lower profits.

Figure 4: Per-consumer profits as a function of campaign reach N_c and product uncertainty $\frac{\sigma_p}{\mu_p}$

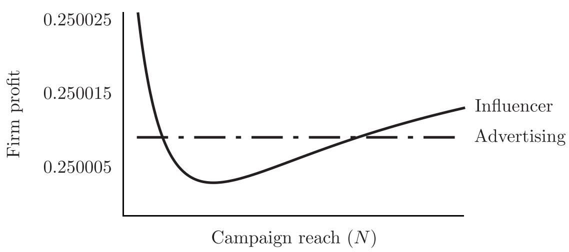
(a) $\frac{\sigma_p^2}{\mu_p^2} < \frac{2}{\pi}$



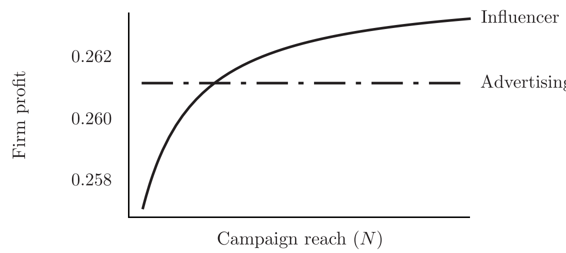
(b) $\frac{2}{\pi} < \frac{\sigma_p^2}{\mu_p^2} < \frac{2}{\pi}\left(\frac{\sigma_{\text{inf}}^2}{\sigma_p^2} + 1\right)$



(c) $\frac{2}{\pi}\left(\frac{\sigma_{\text{inf}}^2}{\sigma_p^2} + 1\right) < \frac{\sigma_p^2}{\mu_p^2} < \frac{2}{\pi}\left(\frac{\sigma_{\text{inf}}^2+1}{\sigma_p^2} + 1\right)$



(d) $\frac{2}{\pi}\left(\frac{\sigma_{\text{inf}}^2+1}{\sigma_p^2} + 1\right) < \frac{\sigma_p^2}{\mu_p^2}$



Notes: Panels (a)-(d) illustrate profits as a function of campaign reach N_c for various values $\frac{\sigma_p}{\mu_p}$ in an increasing order. Panel (a) and (b) depict an established brand and a less established brand, where influencer campaign profits are decreasing in the reach N_{inf}. Panel (d) depicts very unknown brands, where influencer campaign profits are increasing in the reach N_{inf}. Panel (c) shows a mildly unknown brand with intermediate range of $\frac{\sigma_p}{\mu_p}$ where influencer campaign profits first decrease and then increase in N_{inf}.

5 Discussion of results and consumer surplus

Our results show an interaction between the type of product promoted in terms of quality uncertainty (established or unknown brand), and the learning friction of the campaign. The learning frictions, in turn, are determined by the creativity of the influencers, the level of ad targeting, and the size of the follower base of influencers.

For unknown brands there is a high option value from having consumers learn about the quality of the product, which allows the marketer to increase prices and realize higher demand. In this case the marketer would like to engage in campaigns with low learning frictions, which entail using a highly targeted ad technology, or mega-influencers which are consistent in their posts and yield a lower attribution friction. Thus, influencer marketing in this case allows consumers to learn about the product's quality.

For established brands, quality uncertainty is low, and the impact of learning about product quality has a potential downside of lowering demand for the product. In this case a marketer benefits the most from campaigns that increase attention to the product but do not add too much information about the product's quality. These campaigns are the ones that use generic, non-targeted ads, or many micro-influencers who are very creative. These micro-influencers generate high attribution frictions and make learning about product quality ineffective. In this way, the marketers inhibit learning among consumers.

One might think that consumers always benefit from learning, i.e., from facing lower learning frictions and observing more signals from other followers of an influencer. However, similar to the profit effects for marketers the impact is more nuanced. The following proposition characterizes the ex-ante expected consumer surplus (CS) in social media campaigns:

Proposition 4 (Consumer Surplus)**.** *The ex-ante expected consumer surplus is given by*

$$\mathbb{E}_{\mathbf{v}_i^c, q_p, q_c}\left[\mathbb{1}_{\{\mu_p(\mathbf{v}_i^c)\geq m_c^*\}}\cdot(q_p - m_c^*)\right] =$$

$$\frac{\sigma_p}{\kappa_c}\left[1-\Phi\left(\frac{m_c^*-\mu_p}{\sigma_p/\kappa_c}+\frac{\mu_p}{\kappa_c\sigma_p}\right)\right]\cdot\left[\frac{h\left(\frac{m_c^*-\mu_p}{\sigma_p/\kappa_c}+\frac{\mu_p}{\kappa_c\sigma_p}\right)}{\sigma_p}-\frac{m_c^*-\mu_p}{\sigma_p/\kappa_c}\right]$$

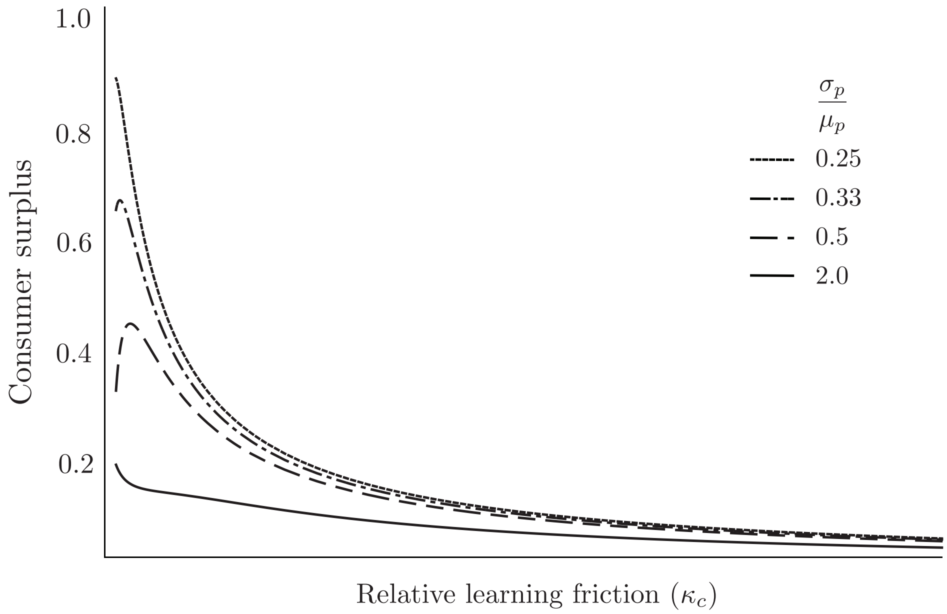
The effect of κ_c on consumer surplus can be non-monotonic. To show this effect, we decompose the consumer surplus as follows:

$$\underbrace{\left(1-\Phi\left(m_c^0\right)\right)}_{\substack{\text{average per customer}\\ \text{probability of purchase}}}\cdot\underbrace{\underbrace{\frac{1}{\kappa_c}}_{\substack{\text{direct benefit}\\ \text{from learning}}}\cdot\underbrace{\frac{1-\Phi\left(m_c^0+\frac{\mu_p}{\kappa_c\sigma_p}\right)}{1-\Phi\left(m_c^0\right)}\cdot\left[h\left(m_c^0+\frac{\mu_p}{\kappa_c\sigma_p}\right)-m_c^0\sigma_p\right]}_{\substack{\text{effect of}\\ \text{consumer surplus extraction}}}}_{\substack{\text{average per purchase}\\ \text{consumer surplus}}} \tag{11}$$

For high levels of learning frictions κ_c, the CS is decreasing in learning frictions κ_c. The reasons are as follows. First, the direct benefit from learning $\frac{1}{\kappa_c}$ decreases CS. Second, by Corollary 2.2, the average per customer probability of purchase is increasing in κ_c and $1-\Phi\left(m_c^0\right)\to_{\kappa_c\to\infty}1$. This is because consumers are more homogeneous and the firm can therefore set a price that incentivizes most consumers to buy. However, this marginal benefit from an increase in the learning friction vanishes as κ_c grows. Third, note that $\frac{1-\Phi\left(m_c^0+\frac{\mu_p}{\kappa_c\sigma_p}\right)}{1-\Phi(m_c^0)} \to_{\kappa_c\to\infty} 1$, and $h\left(m_c^0+\frac{\mu_p}{\kappa_c\sigma_p}\right) \to_{\kappa_c\to\infty} 0$.

By contrast, for small κ_c and low $\frac{\sigma_p}{\mu_p}$, the firm may charge a lower price with more information frictions as illustrated in Figure 2, extracting less surplus with higher frictions. All in all, the higher demand and higher average surplus may dominate the direct benefit from learning. Thus, the CS can be increasing in κ_c for small κ_c. This is illustrated in Figure 5.

Figure 5: Consumer surplus as a function of consumers' learning frictions.



Notes: The Figure depicts the consumer surplus as a function of κ_c for different level of product quality uncertainty. For high values and low values of σ_p/μ_p the consumer surplus always decreases in learning frictions. For intermediate values, consumers might benefit from some level of learning inefficiency.

6 Consumer heterogeneity and influencer choice

The baseline model assumed homogeneous consumption utility that depends only on product quality. Realistically, many consumers have heterogeneous tastes for product consumption. This section achieves two goals: (i) we show that many of the previous insights carry over to this more general case, and (ii) we show that for niche products consistent micro-influencers may be preferred to creative mega-influencers, although the same mega-influencers will be preferred for promoting mass market products. This reversal in profitability stems from how learning frictions interact with consumer heterogeneity. Hence, a marketer might prefer to change the type of influencers they engage with for different products even when the overall brand characteristics do not change.

6.1 Impact of heterogeneity in consumption preferences

Idiosyncratic consumption utilities add novel trade-offs in the following dimensions. First, the marketer can no longer extract all surplus without consumer learning, so the demand effects will be more nuanced compared to the homogeneous case. Second, the information set in the social learning stage is not uniform across consumers for influencer marketing campaigns, which increases the overall information frictions.

Formally, we assume that the consumption utility is given by $u_i = q_p + \epsilon_i^p$, where as before, $q_p \sim \mathcal{N}(\mu_p, \sigma_p^2)$ represents the quality of the product, but now $\epsilon_i^p \sim \mathcal{N}(0, \tau_p^2)$ is an idiosyncratic taste component of consumer i. Accordingly, a consumer's content utility is given by

$$v_i = q_c + q_p + \epsilon_i^c + \epsilon_i^p. \tag{12}$$

τ_p^2 captures the heterogeneity in consumption taste, while the variances of ϵ_i^c, $\tau_{\text{inf}}^2 = 1$ and τ_{ad}^2, capture the heterogeneity in content taste. As before, we can interpret $\frac{1}{\tau_{\text{ad}}^2}$ to be the level of targeting of an ad.

Each marketer's product is characterized by two parameters: the product's taste heterogeneity τ_p^2 and the uncertainty about the product quality σ_p/μ_p. We call products with a small τ_p^2 mass market products and products with a large τ_p^2 niche products. We previously analyzed the homogeneous case with $\tau_p^2 = 0$, which can be considered a mass market case.

We first extend the analysis of Section 3 by showing that the qualitative results remain unchanged. Unlike in the homogeneous case, now even for an influencer campaign the information set $(\mathbf{v}_i^{\text{inf}}, \epsilon_i^p)$ is different for every consumer i, but has an overlapping $q_c + q_p + \epsilon_i^c$ component for all i. Thus, there is still a force that makes the product either "go viral" or "go bust," but it is less of a dichotomous (bang-bang) outcome than without ϵ_i^p. We derive the belief distribution $q_p \big| \mathbf{v}^c, \epsilon_i^p \sim \mathcal{N}\left(\mu_p(\mathbf{v}^c, \epsilon_i^p), \sigma_p^2(\mathbf{v}^c, \epsilon_i^p)\right)$ in Lemma OA.1 in the Online Appendix,[8] and analogous to Proposition 1, we characterize the demand faced by the firm in Proposition OA.1 in the Online Appendix. Using these results, we characterize the firm's profit maximizing price:

[8]The proofs for results in this Section appear in the Online Appendix.

Proposition 5 (Optimal price with consumption heterogeneity)**.** *The unique profit-maximizing price $m_c^{\text{het},*}$ is given by*

$$m_c^{\text{het},*} = \mu_p + m_c^{\text{het},0} \frac{\sigma_p}{\tilde{\kappa}_c^{\text{het}}}. \tag{13}$$

Moreover, the price $m_c^{\text{het},*}$ exceeds the ex-ante expected product quality if and only if

$$\sqrt{\frac{2}{\pi}} \tilde{\kappa}_c^{\text{het}} < \frac{\sigma_p}{\mu_p}. \tag{14}$$

Proposition 5 is the heterogeneous analog of Proposition 2. In its definition, $m_c^{\text{het},0}$ is the heterogeneity equivalent of m_c^0 from the solution to Equation (7) where the learning friction κ_c was replaced by the heterogeneity equivalent learning friction $\tilde{\kappa}_c^{\text{het}}$ which is defined as::

$$\tilde{\kappa}_c^{\text{het}} = \frac{\kappa_c^{\text{het}}}{\sqrt{1 + (\kappa_c^{\text{het}})^2 \frac{\tau_p^2}{\sigma_p^2}}}$$

with

$$(\kappa_{\text{inf}}^{\text{het}})^2 = \frac{\sigma_{\text{inf}}^2 + \dfrac{1+\tau_p^2}{N_{\text{inf}} + \tau_p^2}}{\sigma_p^2} + 1 \qquad (\kappa_{\text{ad}}^{\text{het}})^2 = \frac{1+\tau_{\text{ad}}^2}{\sigma_p^2} + 1.$$

The expressions for $\kappa_{\text{inf}}^{\text{het}}$ and $\kappa_{\text{ad}}^{\text{het}}$ illustrate how heterogeneity affects the learning frictions differently for the two campaign types. First, $\kappa_{\text{inf}}^{\text{het}} > \kappa_{\text{inf}}$ because consumption heterogeneity adds another information friction. Second, when $n = 1$, which is the case for ad campaigns, then $\kappa_{\text{ad}}^{\text{het}} = \kappa_{\text{ad}}$ because without social learning from others, the consumption heterogeneity friction is not present. Hence, consumer taste heterogeneity has the same effect for ad campaigns regardless of their reach, but will have a different effect for influencer campaigns depending on their level of creativity and the number of followers they reach. The profit from a campaign c is then given by

$$\Pi_c^{\text{het},*} = N_c \cdot \pi_c^{\text{het},*}\left(\tilde{\kappa}_c^{\text{het}}\right) = N_c \left(1 - \Phi\left(m_c^{\text{het},0}\right)\right) m_c^{\text{het},*}. \tag{15}$$

Corollary 2.2 implies that equilibrium demand is increasing in $\tilde{\kappa}_c^{\text{het}}$. Similarly, the optimal profits decrease in $\tilde{\kappa}_c^{\text{het}}$, when (14) is satisfied, and they increase otherwise, analogous to Corollary 2.3. Consequently, previous results from the homogeneous baseline model generalize, and we can examine how heterogeneity affects the heterogeneous learning frictions $\tilde{\kappa}_c^{\text{het}}$. The following result shows that consumer heterogeneity generally decrease learning frictions:

Lemma 2.

1. *For any level of consumer heterogeneity* $\tilde{\kappa}_c^{\text{het}} \leq \kappa_c$.

2. *The heterogeneous learning friction is decreasing with consumer heterogeneity. Moreover,*

$$\lim_{\tau_p \to 0} \tilde{\kappa}_c^{\text{het}} = \kappa_c \qquad \text{and} \qquad \lim_{\tau_p \to \infty} \tilde{\kappa}_c^{\text{het}} = 0. \tag{16}$$

An increase in consumer heterogeneity therefore monotonically decreases the heterogeneous learning friction $\tilde{\kappa}_c^{\text{het}}$ even though consumers benefit less from social learning. It might seem surprising that an increase in heterogeneity lowers learning frictions. This is because the option value of learning is also smaller for the firm. Put differently, social learning is less important if consumption preferences are heterogeneous.

6.2 Impact of heterogeneity on influencer choice

The effect of decreasing heterogeneous learning frictions with increasing consumer heterogeneity has important implications for the comparison of influencer marketing with ad campaigns. For ad campaigns $\tilde{\kappa}_{\text{ad}}^{\text{het}}$, which is smaller then κ_{ad}, is unaffected by the ad reach N_{ad}, and is only affected by heterogeneity τ_p. However, for influencer campaigns, $\tilde{\kappa}_{\text{inf}}^{\text{het}}$ is decreasing in the reach N_{inf}. Importantly, the larger τ_p is, the smaller the negative impact of N_{inf} on $\tilde{\kappa}_{\text{inf}}^{\text{het}}$, i.e., $\frac{\partial}{\partial \tau_p} \frac{\partial}{\partial N_{\text{inf}}} \tilde{\kappa}_{\text{inf}}^{\text{het}} > 0$, where in the limit as $\tau_p \to \infty$, N_{inf} does not affect $\tilde{\kappa}_{\text{inf}}^{\text{het}} = 0$ and therefore per-consumer profits at all. In that sense, for a niche product, the influencer follower base N_{inf} is relatively more relevant than the influencer's creativity when choosing an influencer. In turn, for a mass market product, the creativity σ_{inf} or consistency of posts are relatively more relevant compared to the follower base.

Because of the interaction between preference heterogeneity and the importance of follower base and creativity, for niche products with large τ_p, different types of influencers can be more profitable than for mass market products with small τ_p. This can be seen by inspection of the definition of κ_c^{het}, which monotonously affects $\tilde{\kappa}_c^{\text{het}}$. Consider two influencers $j \in \{1, 2\}$ with $\sigma_{\text{inf},1} < \sigma_{\text{inf},2}$ and

$2 = N_{\text{inf},1} < N_{\text{inf},2}$, i.e., influencer 2 is more creative and has a greater follower base than influencer 1. The following proposition identifies conditions where a marketer might find it profitable to engage with mega-influencers for mass market products, but switch to using micro-influencers for niche products.

Proposition 6. *Suppose that the follower size of influencer 2 is sufficiently large.*[9] *Consistent micro-influencers yield the firm higher profits per consumer for niche products (τ_p large), and a creative mega-influencer yields the firm higher profits per consumer for mass market products (τ_p small) when:*

(a) For established brands: Either $0 < \sigma^2_{\text{inf},1} \leq \sigma^2_{\text{inf},2} - 1$ *or* $\max\{\sigma^2_{\text{inf},2} - 1, 0\} < \sigma^2_{\text{inf},1} \leq \sigma^2_{\text{inf},2} - \frac{1}{2}$.

(b) For unknown brands: $\max\{\sigma^2_{\text{inf},2} - \frac{1}{2}, 0\} < \sigma^2_{\text{inf},1} < \sigma^2_{\text{inf},2}$.

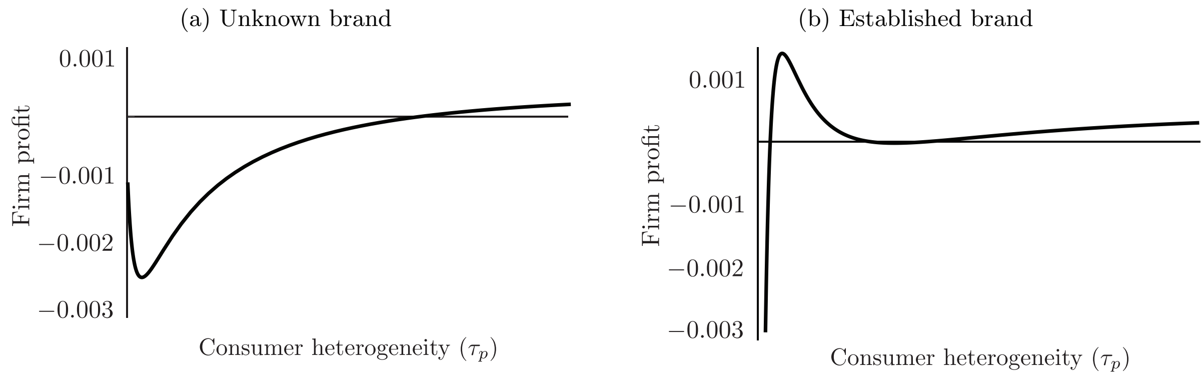
Intuitively, for a mass market product from an established brand (low product uncertainty), we have shown that the marketer prefers a campaign with high information friction (Figure 3(b)). An increase in heterogeneity τ_p reduces the heterogeneous learning friction $\tilde{\kappa}_c^{\text{het}}$, and to counteract this effect, the marketer will need to switch to influencers who have a lower follower base. We illustrate this crossing effect in Figure 6(b), where the profit using mega-influencers is higher for mass market products, and lower with niche products. The figure also shows that there might be cases where there are multiple reversals of profitability between micro- and mega-influencers, and this is because for intermediate levels of τ_p, the level of creativity still has a large effect. For unknown brands, lower or higher information frictions might be more profitable (Figure 3(a)) in general, but given the constraint that influencer 2 has more followers and is more creative, switching to using micro-influencers for niche products will increase learning frictions and be more profitable, as illustrated in Figure 6(a).

7 Conclusion

What contributes to the popularity of influencer marketing campaigns? And when should marketers consider using them instead of targeted ad campaigns? We studied these questions using a model that focuses on the effect of social learning that is enabled by social media influencers, but that is

[9] We provide lower bounds on $N_{\text{inf},2}$ and thresholds for $\frac{\sigma_p}{\mu_p}$ in the proof in the Online Appendix.

Figure 6: Effect of consumer heterogeneity on difference of per-consumer profits for two influencers



Notes: Panels (a) and (b) illustrate the difference in profits $\pi^*_{\text{inf},1} - \pi^*_{\text{inf},2}$ between campaigns with consistent micro-influencers and a creative mega-influencer as a function of consumer heterogeneity τ_p for various values of $\frac{\sigma_p}{\mu_p}$. Consistent micro-influencers lead to higher profit for a niche product with a higher τ_p. Panel (a) depicts an unknown brand: $\frac{\sigma_p}{\mu_p} = 4$, $\sigma^2_{\text{inf},1} = 0.7$. Panel (b) depicts an established brand: $\frac{\sigma_p}{\mu_p} = 0.4$, $\sigma^2_{\text{inf},1} = 0.5$. Both panels: $N_{\text{inf},1} = 2$, $N_{\text{inf},2} = 25$, $\sigma^2_{\text{inf},2} = 1.1$.

rarely possible using targeted ad campaigns. First, we highlight that a marketer should evaluate whether the product has high product quality uncertainty (e.g., because it belongs to an unknown brand) or low quality uncertainty (e.g., because it belongs to an established brand). For established brands, the marketer should focus on campaigns that focus on making consumers aware of a product without providing too much information. Instead, for unknown brands, learning can increase profits by allowing the marketer to charge a higher price for the product. Second, we show how information frictions may operate differently for ad campaigns and influencer campaigns. The attribution friction—that stems from the inability to separate enjoying the social media post for its entertainment content from the quality of the product promoted—can be high if the influencer is very creative or if the follower base has heterogeneous content taste. The finite information friction is always present for ad campaigns, but is low for influencers with a large follower base. Since the number of followers affect the information frictions in this way, mega-influencers do not only make more consumers aware, but also lead to a lot of social learning. Hence, for established brands, it can be valuable to engage many micro-influencers with high content creativity instead of mega-influencers, to achieve high awareness of the product while minimizing social learning.

Another notable finding is that influencer campaigns—especially those with many followers—concentrate beliefs of consumers and allow the marketer to charge a price that entices all consumers to buy the product. This “one price fits all” effect is not feasible for ad campaigns. As a result, influencer campaigns generate a “go viral” or “go bust” effect, which affect the benefit that different types of promoted products can gain. One implication of the demand concentration of influencer campaigns is that a marketer can use a discount after the campaign to make it even more profitable. The strategy would operate as follows: before the campaign the marketer would set a relatively high price for the product, and then depending on the success of the campaign (whether the response from consumers was positive or negative), the marketer can offer a discount code through the influencer to extract the potential consumer surplus. This strategy is not possible with an ad campaign, because one price will not extract much surplus after an ad campaign that does not concentrate consumer beliefs. In essence, influencer marketing with many followers allows a marketer to use the network effect to price discriminate based on the consumers’ responses to the campaign.

Because we focused on the impact of social learning in this paper, we abstracted away from other considerations that affect the efficacy of influencer and advertising campaigns. Particularly, we did not consider the compensation contract between the influencer and the marketer, nor did we allow the influencer to strategically choose which products they promote. These choices were made to compare influencer and ad campaigns using an apples-to-apples comparison that removes the effect of market equilibria in ad prices and influencer campaign costs. Our comparison can provide guidance to marketers on when they should expect to pay influencers more, and what returns they should see from these campaigns. Endogenizing the incentive contracts for influencers and considering competitive ad markets are two promising future directions for additional research.

Two other forces that we did not model can potentially come from the consumer side. Consumers might consider signals from other followers of the influencer with different weights than their own signal, while our model assumes that all signals have equal weight. In this case similar forces to our model should apply, but such a framework would add yet another learning friction of influencer marketing. Another force that is of interest revolves around the dynamics of consumers. As consumers do not all simultaneously observe the social media posts by the influencers, there is a potential for observational learning and other word-of-mouth effects that have been previously analyzed in the literature (Banerjee 1992, Bikhchandani et al. 1992, Zhang 2010, Kamada and Öry 2020, Fainmesser et al. 2021). Analyzing the effects of such dynamics is another promising avenue for future work.

The results we uncovered are useful for both practitioners and researchers. For practitioners, we uncovered important features of influencer and ad campaigns that may be relevant for campaign effectiveness, and that shhould be used when planning campaigns. We also provide normative findings about the effects of influencer campaigns on overrall learning, consumer surplus and firm profits. Measuring and identifying the effects of influencer campaigns is notoriously difficult because of the endogenous nature of the data being observed. Thus, for researachers, our results provide predictions about outcomes that are generated by influencer campaigns in the presence of social learning, and also provide a benchmark that can be used to compare ad and influencer campaigns beyond simple ROI metrics. We also provide predictions on which products we expect to be promoted by influencer campaigns. Our framework uncovers new mechanisms that make influencer marketing succeed or fail, which can strengthen the conclusions from empirical analysis that focuses on estimating the effectiveness of social media promotion campaigns.

Funding and Competing Interests

Partial financial support was received from the NET Institute. All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

References

Banerjee, A. V. (1992). A simple model of herd behavior. *Quarterly Journal of Economics 107*(3), 992–1026.

Berman, R. and Z. Katona (2020). Curation algorithms and filter bubbles in social networks. *Marketing Science 39*(2), 296–316.

Bikhchandani, S., D. Hirshleifer, and I. Welch (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy 100*(5), 992–1026.

Campbell, A. (2013). Word-of-mouth communication and percolation in social networks. *American Economic Review 103*(6), 2466–98.

Campbell, A., D. Mayzlin, and J. Shin (2017). Managing buzz. *The RAND Journal of Economics 48*(1), 203–229.

Carroni, E., P. Pin, and S. Righi (2020). Bring a friend! privately or publicly? *Management Science 66*(5), 2269–2290.

Crapis, D., B. Ifrach, C. Maglaras, and M. Scarsini (2017). Monopoly pricing in the presence of social learning. *Management Science 63*(11), 3586–3608.

Fainmesser, I. P. and A. Galeotti (2021). The market for online influence. *American Economic Journal: Microeconomics 13*(4), 332–72.

Fainmesser, I. P., D. Olié Lauga, and E. Ofek (2021). Ratings, reviews, and the marketing of new products. *Management Science 67*(11), 7023–7045.

Galeotti, A., C. Ghiglino, and F. Squintani (2013). Strategic information transmission networks. *Journal of Economic Theory 148*(5), 1751–1769.

Ifrach, B., C. Maglaras, M. Scarsini, and A. Zseleva (2019). Bayesian social learning from consumer reviews. *Operations Research 67*(5), 1209–1221.

Jiang, B., O. Turut, and T. Zou (2021). A one-sentence tweet or a one-hour video? influencing the influencer's recommendations with discounts. Available at SSRN 3922188.

Kamada, Y. and A. Öry (2020). Contracting with word-of-mouth management. *Management Science 66*(11), 5094–5107.

Kuksov, D. and C. Liao (2019). Opinion leaders and product variety. *Marketing Science 38*(5), 812–834.

Mitchell, M. (2021). Free ad(vice): internet influencers and disclosure regulation. *RAND Journal of Economics 52*(1), 3–21.

Nistor, C. and M. Selove (2022). Influencers: The power of comments. Available at SSRN 4118010.

Pei, A. and D. Mayzlin (2022). Influencing social media influencers through affiliation. *Marketing Science 41*(3), 593–615.

Shin, J. and J. Yu (2021). Targeted advertising and consumer inference. *Marketing Science 40*(5), 900–922.

Zhang, J. (2010). The sound of silence: Observational learning in the u.s. kidney market. *Marketing Science 29*(2), 315–335.

Appendix – Proofs

Proof of Proposition 1. For the posterior belief about the product's quality, we define the following vector notation:

$$\mathbf{q} = \begin{pmatrix} q_c \\ q_p \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_c \\ \mu_p \end{pmatrix}, \begin{pmatrix} \sigma_c^2 & 0 \\ 0 & \sigma_p^2 \end{pmatrix}\right).$$

Since q_c and q_p are independent, their sum is distributed $q_c + q_p \sim \mathcal{N}(\mu_c + \mu_p, \sigma_c^2 + \sigma_p^2)$. Consumer i's content utility is distributed $v_i \sim \mathcal{N}(q_c + q_p, \tau_c^2)$.

We denote the density of the distribution of a random vector $\mathbf{a}$ by $f_{\mathbf{a}}$, and with a slight abuse of notation we denote the argument of the density functions by the same letter as the random variables, i.e., $f_{\mathbf{a}}(\mathbf{a})$.

By Bayes' rule, the probability density of the vector $\mathbf{q}$ conditional on the observations in $\mathbf{v}$ is given by:

$$f_{\mathbf{q}|\mathbf{v}}(\mathbf{q}) = \frac{f_{\mathbf{v}|\mathbf{q}}(\mathbf{v}) f_{\mathbf{q}}(\mathbf{q})}{f_{\mathbf{v}}(\mathbf{v})} \propto f_{\mathbf{v}|\mathbf{q}}(\mathbf{v}) f_{\mathbf{q}}(\mathbf{q}). \tag{17}$$

Since the v_i's are i.i.d from $\mathcal{N}(q_c + q_p, \tau_c^2)$, the likelihood $f_{\mathbf{v}|\mathbf{q}}(\mathbf{v})$ equals:

$$f_{\mathbf{v}|\mathbf{q}}(\mathbf{v}) = \prod_{i=1}^{n} f_{v_i|\mathbf{q}}(v_i) \propto \prod_{i=1}^{n} \exp\left(-\frac{(v_i - q_p - q_c)^2}{2\tau_c^2}\right) = \exp\left(-\frac{\sum_{i=1}^{n}(v_i - q_p - q_c)^2}{2\tau_c^2}\right). \tag{18}$$

Further, the prior $f_{\mathbf{q}}(\mathbf{q})$ equals:

$$f_{\mathbf{q}}(\mathbf{q}) = \exp\left[-\frac{1}{2}\begin{pmatrix} q_c - \mu_c & q_p - \mu_p \end{pmatrix} \begin{pmatrix} \frac{1}{\sigma_c^2} & 0 \\ 0 & \frac{1}{\sigma_p^2} \end{pmatrix} \begin{pmatrix} q_c - \mu_c \\ q_p - \mu_p \end{pmatrix}\right]. \tag{19}$$

We expand the exponents in (18) and (19) to rewrite (17) in the standard form of a multivariate normal density. First, (18) can be written as:

$$
\begin{aligned}
-\frac{\sum_{i=1}^n (v_i - q_p - q_c)^2}{2\tau_c^2} &= -\frac{1}{2\tau_c^2}\left(\underbrace{nq_c^2 + nq_p^2 + 2nq_cq_p}_{\text{(I)}} \underbrace{-2\sum_{i=1}^n v_iq_c - 2\sum_{i=1}^n v_iq_p}_{\text{(II)}} + \underbrace{\sum_{i=1}^n (v_i)^2}_{\text{(III)}}\right) \\
&= \underbrace{-\frac{n}{2\tau_c^2}\mathbf{q}'\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\mathbf{q}}_{(\text{I}_1)} + \underbrace{\frac{\sum_{i=1}^n v_i}{\tau_c^2}\begin{pmatrix} 1 & 1 \end{pmatrix}\mathbf{q}}_{(\text{II}_1)} \underbrace{-\frac{\sum_{i=1}^n (v_i)^2}{2\tau_c^2}}_{(\text{III}_1)},
\end{aligned} \tag{20}
$$

Next, we expand the exponent in (19) as follows

$$
\begin{aligned}
&-\frac{1}{2}\begin{pmatrix} q_c - \mu_c & q_p - \mu_p \end{pmatrix}\begin{pmatrix} \frac{1}{\sigma_c^2} & 0 \\ 0 & \frac{1}{\sigma_p^2} \end{pmatrix}\begin{pmatrix} q_c - \mu_c \\ q_p - \mu_p \end{pmatrix} \\
&= -\frac{1}{2}\mathbf{q}'\begin{pmatrix} \frac{1}{\sigma_c^2} & 0 \\ 0 & \frac{1}{\sigma_p^2} \end{pmatrix}\mathbf{q} + \begin{pmatrix} \mu_c & \mu_p \end{pmatrix}\begin{pmatrix} \frac{1}{\sigma_c^2} & 0 \\ 0 & \frac{1}{\sigma_p^2} \end{pmatrix}\mathbf{q} - \frac{1}{2}\begin{pmatrix} \mu_c & \mu_p \end{pmatrix}\begin{pmatrix} \frac{1}{\sigma_c^2} & 0 \\ 0 & \frac{1}{\sigma_p^2} \end{pmatrix}\begin{pmatrix} \mu_c \\ \mu_p \end{pmatrix} \\
&= \underbrace{-\frac{1}{2}\mathbf{q}'\begin{pmatrix} \frac{1}{\sigma_c^2} & 0 \\ 0 & \frac{1}{\sigma_p^2} \end{pmatrix}\mathbf{q}}_{(\text{I}_2)} + \underbrace{\begin{pmatrix} \frac{\mu_c}{\sigma_c^2} & \frac{\mu_p}{\sigma_p^2} \end{pmatrix}\mathbf{q}}_{(\text{II}_2)} \underbrace{-\frac{1}{2}\left(\frac{\mu_c^2}{\sigma_c^2} + \frac{\mu_p^2}{\sigma_p^2}\right)}_{(\text{III}_2)}.
\end{aligned} \tag{21}
$$

To derive the expression in (17) we add the exponents in (20) and in (21):

$$
\begin{aligned}
\text{I}_1 + \text{I}_2 &= -\frac{1}{2}\mathbf{q}'\left[\begin{pmatrix} \frac{n}{\tau_c^2} & \frac{n}{\tau_c^2} \\ \frac{n}{\tau_c^2} & \frac{n}{\tau_c^2} \end{pmatrix} + \begin{pmatrix} \frac{1}{\sigma_c^2} & 0 \\ 0 & \frac{1}{\sigma_p^2} \end{pmatrix}\right]\mathbf{q} \\
&= -\frac{1}{2}\mathbf{q}'\begin{pmatrix} \frac{n}{\tau_c^2} + \frac{1}{\sigma_c^2} & \frac{n}{\tau_c^2} \\ \frac{n}{\tau_c^2} & \frac{n}{\tau_c^2} + \frac{1}{\sigma_p^2} \end{pmatrix}\mathbf{q}, &(22) \\
\text{II}_1 + \text{II}_2 &= \left[\begin{pmatrix} \frac{\sum_{i=1}^n v_i}{\tau_c^2} & \frac{\sum_{i=1}^n v_i}{\tau_c^2} \end{pmatrix} + \begin{pmatrix} \frac{\mu_c}{\sigma_c^2} & \frac{\mu_p}{\sigma_p^2} \end{pmatrix}\right]\mathbf{q} \\
&= \begin{pmatrix} \frac{\sum_{i=1}^n v_i}{\tau_c^2} + \frac{\mu_c}{\sigma_c^2} & \frac{\sum_{i=1}^n v_i}{\tau_c^2} + \frac{\mu_p}{\sigma_p^2} \end{pmatrix}\mathbf{q}, &(23) \\
\text{III}_1 + \text{III}_2 &= -\frac{\sum_{i=1}^n (v_i)^2}{2\tau_c^2} - \frac{1}{2}\left(\frac{\mu_c^2}{\sigma_c^2} + \frac{\mu_p^2}{\sigma_p^2}\right).
\end{aligned}
$$

When $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is drawn from a bivariate normal distribution, its probability density function is given by

$$f_{\mathbf{X}}(\mathbf{X}) = \frac{1}{2\pi \det(\mathbf{\Sigma})^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{X}-\boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu})\right] =$$
$$\frac{1}{2\pi \det(\mathbf{\Sigma})^{\frac{1}{2}}} \exp\left[-\frac{1}{2}\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X} + \boldsymbol{\mu}'\mathbf{\Sigma}^{-1}\mathbf{X} - \frac{1}{2}\boldsymbol{\mu}'\mathbf{\Sigma}^{-1}\boldsymbol{\mu}\right].$$

By comparing the quadratic form $\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X}$ to (22) we find $\Sigma_{\mathbf{q}}(\mathbf{v})$ and then we recover the mean $\mu_{\mathbf{q}}(\mathbf{v^c})$ by post-multiplying the linear coefficients of (23) with $\Sigma_{\mathbf{q}}(\mathbf{v})$.

The posterior variance-covariance results in:

$$\Sigma_{\mathbf{q}}(\mathbf{v}) = \begin{pmatrix} \frac{1}{\sigma_c^2} + \frac{n}{\tau_c^2} & \frac{n}{\tau_c^2} \\ \frac{n}{\tau_c^2} & \frac{1}{\sigma_p^2} + \frac{n}{\tau_c^2} \end{pmatrix}^{-1} = \begin{pmatrix} \frac{(\sigma_p^2 + \frac{\tau_c^2}{n})\sigma_c^2}{\sigma_c^2 + \sigma_p^2 + \frac{\tau_c^2}{n}} & -\frac{\sigma_c^2\sigma_p^2}{\sigma_c^2 + \sigma_p^2 + \frac{\tau_c^2}{n}} \\ -\frac{\sigma_c^2\sigma_p^2}{\sigma_c^2 + \sigma_p^2 + \frac{\tau_c^2}{n}} & \frac{(\sigma_c^2 + \frac{\tau_c^2}{n})\sigma_p^2}{\sigma_c^2 + \sigma_p^2 + \frac{\tau_c^2}{n}} \end{pmatrix}$$
$$= \begin{pmatrix} \frac{(\sigma_p^2 + \frac{\tau_c^2}{n})\sigma_c^2}{\kappa_c^2\sigma_p^2} & -\frac{\sigma_c^2}{\kappa_c^2} \\ -\frac{\sigma_c^2}{\kappa_c^2} & \sigma_p^2\left(1 - \frac{1}{\kappa_c^2}\right) \end{pmatrix}$$

and the posterior mean of $\mathbf{q}|\mathbf{v}$ is

$$\mu_{\mathbf{q}}(\mathbf{v})' = \begin{pmatrix} \frac{\sum_{i=1}^n v_i}{\tau_c^2} + \frac{\mu_c}{\sigma_c^2} & \frac{\sum_{i=1}^n v_i}{\tau_c^2} + \frac{\mu_p}{\sigma_p^2} \end{pmatrix} \Sigma_{\mathbf{q}}(\mathbf{v})$$
$$= \begin{pmatrix} \frac{n\bar{\mathbf{v}}^c}{\tau_c^2} + \frac{\mu_c}{\sigma_c^2} & \frac{n\bar{\mathbf{v}}^c}{\tau_c^2} + \frac{\mu_p}{\sigma_p^2} \end{pmatrix} \begin{pmatrix} \frac{(\sigma_p^2 + \frac{\tau_c^2}{n})\sigma_c^2}{\sigma_c^2 + \sigma_p^2 + \frac{\tau_c^2}{n}} & -\frac{\sigma_c^2\sigma_p^2}{\sigma_c^2 + \sigma_p^2 + \frac{\tau_c^2}{n}} \\ -\frac{\sigma_c^2\sigma_p^2}{\sigma_c^2 + \sigma_p^2 + \frac{\tau_c^2}{n}} & \frac{(\sigma_c^2 + \frac{\tau_c^2}{n})\sigma_p^2}{\sigma_c^2 + \sigma_p^2 + \frac{\tau_c^2}{n}} \end{pmatrix}$$
$$= \begin{pmatrix} \frac{(\sigma_p^2 + \frac{\tau_c^2}{n})\mu_c + (\bar{\mathbf{v}}^c - \mu_p)\sigma_c^2}{\sigma_c^2 + \sigma_p^2 + \frac{\tau_c^2}{n}} & \frac{(\sigma_c^2 + \frac{\tau_c^2}{n})\mu_p + (\bar{\mathbf{v}}^c - \mu_c)\,\sigma_p^2}{\sigma_c^2 + \sigma_p^2 + \frac{\tau_c^2}{n}} \end{pmatrix}$$
$$= \begin{pmatrix} \mu_c + \frac{\sigma_c^2}{\sigma_p^2}\frac{\bar{\mathbf{v}}^c - \mu_c - \mu_p}{\kappa_c^2} & \mu_p + \frac{\bar{\mathbf{v}}^c - \mu_c - \mu_p}{\kappa_c^2} \end{pmatrix}.$$

The expression for $\mu_p(\mathbf{v}^c)$ equals $\mu_p + \dfrac{\bar{\mathbf{v}}^c - \mu_c - \mu_p}{\kappa_c^2}$, with expectation

$$\mathbb{E}[\mu_p(\mathbf{v})] = \mu_p + \frac{q_c + q_p - \mu_c - \mu_p}{\kappa_c^2}.$$

The independence across the v_i's implies that the variance of $\mu_p(\mathbf{v})$ equals

$$\mathrm{Var}(\mu_p(\mathbf{v})) = \frac{\tau_c^2}{n\kappa_c^4}.$$

Next, let $\Phi(\cdot)$ denote the CDF of the standard normal distribution and $\phi(\cdot)$ its PDF. A consumer i's individual demand with price m is

$$\Pr(\mu_p(\mathbf{v})) \geq m) = 1 - \Phi\left(\frac{m - \mathbb{E}[\mu_p(\mathbf{v})]}{\sqrt{\mathrm{Var}(\mu_p(\mathbf{v}))}}\right).$$

In particular, the demand is determined by the realization of $\mathbf{v}_i^c$. Then the demand faced by the firm is

$$D(m; q_p, q_c) = N \cdot \int \mathbb{1}\left(\frac{1}{\kappa_c^2}\left(x\sqrt{\frac{\tau_c^2}{n}} + q_p + q_c - \mu_c\right) + \left(1 - \frac{1}{\kappa_c^2}\right)\mu_p \geq m\right)\phi(x)\,\mathrm{d}x.$$

Integrating over x, we get

$$D(m; q_p, q_c) = N\left(1 - \Phi\left(\frac{(m - \mu_p)\kappa_c^2 + \mu_p + \mu_c - q_p - q_c}{\sqrt{\frac{\tau_c^2}{n}}}\right)\right).$$

Next, we can calculate the firm's expected demand $\mathbb{E}_{q_p,q_c}[D(m; q_p, q_c)]$. Normalizing $q_c + q_p$ as $w = \frac{q_c + q_p - \mu_c - \mu_p}{\sqrt{\sigma_c^2 + \sigma_p^2}} \sim \mathcal{N}(0, 1)$, we get

$$\mathbb{E}_{q_p,q_c}[D(m; q_p, q_c)] = N\left(1 - \Phi\left(\frac{m - \mu_p}{\sigma_p}\kappa_c\right)\right).$$

□

Proof of Lemma 1.

$$\kappa_{\mathrm{ad}} > \kappa_{\mathrm{inf}} \iff \kappa_{\mathrm{ad}}^2 > \kappa_{\mathrm{inf}}^2 \iff \tau_{\mathrm{ad}}^2 > \sigma_{\mathrm{inf}}^2 + \frac{\tau_{\mathrm{inf}}^2}{N_{\mathrm{inf}}}.$$

□

Proof of Proposition 2. The marketer's expected revenue is $\pi(m) = m \cdot \mathbb{E}_{q_p,q_c}[D(m; q_p, q_c)]$. Taking derivative of the expected revenue with respect to m, we obtain

$$\frac{\partial \pi(m)}{\partial m} = n\left(1-\Phi\left(\frac{m-\mu_p}{\sigma_p}\kappa_c\right)\right)\cdot\left(1-\frac{m}{\sigma_p}\kappa_c\cdot h\left(\frac{m-\mu_p}{\sigma_p}\kappa_c\right)\right),$$

where $h(\cdot)=\frac{\phi(\cdot)}{1-\Phi(\cdot)}$ is the hazard function of the standard normal distribution. To show that m_c^* is the unique profit maximizing price, we will show that (1) the necessary first-order condition is also sufficient, and (2) m_c^* satisfies the first-order condition for profit maximization.

The sign of $\frac{\partial \pi(m)}{\partial m}$ is determined by the sign of the second factor $1-\frac{m}{\sigma_p}\kappa_c\cdot h\left(\frac{m-\mu_p}{\sigma_p}\kappa_c\right)$ as the first factor is always positive. Since the hazard function $h(x)$ is strictly increasing, $1-\frac{m}{\sigma_p}\kappa_c\cdot h\left(\frac{m-\mu_p}{\sigma_p}\kappa_c\right)$ is strictly decreasing in m. When m is sufficiently small, $1-\frac{m}{\sigma_p}\kappa_c\cdot h\left(\frac{m-\mu_p}{\sigma_p}\kappa_c\right)>0$, whereas when m is sufficiently large, $1-\frac{m}{\sigma_p}\kappa_c\cdot h\left(\frac{m-\mu_p}{\sigma_p}\kappa_c\right)<0$. It follows that $\frac{\partial \pi(m)}{\partial m}>0$ when m is sufficiently small and $\frac{\partial \pi(m)}{\partial m}<0$ when m is sufficiently large, and that the sign of $\frac{\partial \pi(m)}{\partial m}$ only changes once. Therefore, there is a unique m_c^* such that $\left.\frac{\partial \pi(m)}{\partial m}\right|_{m=m_c^*}=0$.

To complete the proof of the first part of the proposition, we verify that the expression of $m_c^*=\mu_p+m_c^0\frac{\sigma_p}{\kappa_c}$ satisfies $1-\frac{m}{\sigma_p}\kappa_c\cdot h\left(\frac{m-\mu_p}{\sigma_p}\kappa_c\right)=0$. Since m_c^* is a linear transformation from m_c^0, substitution of m_c^0 into m_c^* satisfies the equation $1-\frac{m}{\sigma_p}\kappa_c\cdot h\left(\frac{m-\mu_p}{\sigma_p}\kappa_c\right)=0$ if and only if m_c^0 satisfies the equation

$$1-\frac{1}{\sigma_p}\left(\mu_p+m_c^0\frac{\sigma_p}{\kappa_c}\right)\kappa_c\cdot h\left(m_c^0\right)=0.$$

Simplifying the equation above, m_c^0 satisfies $m_c^0-\frac{1}{h(m_c^0)}=-\frac{\mu_p}{\sigma_p}\kappa_c$, which is precisely the definition of m_c^0. Therefore, the revenue-maximizing price is m_c^*.

To prove the second part of the proposition, we note that $m_c^*>\mu_p$ if and only if $m_c^0>0$, which is the condition in (9). □

Proof of Corollary 2.1. Differentiating the optimal price m_c^* from (8) with respect to κ_c yields:

$$\frac{\mathrm{d}m_c^*}{\mathrm{d}\kappa_c}=-\frac{\sigma_p}{\kappa_c^2}m_c^0+\frac{\sigma_p}{\kappa_c}\frac{\mathrm{d}m_c^0}{\mathrm{d}\kappa_c}, \tag{24}$$

where by (7):

$$-\frac{\mu_p}{\sigma_p}\frac{\mathrm{d}\kappa_c}{\mathrm{d}m_c^0}=1+\frac{h'(m_c^0)}{(h(m_c^0))^2}\implies\frac{\mathrm{d}m_c^0}{\mathrm{d}\kappa_c}=-\frac{\sigma_p}{\mu_p}\frac{1}{1+\frac{h'(m_c^0)}{(h(m_c^0))^2}}. \tag{25}$$

Plugging (25) into (24) yields:

$$\frac{\mathrm{d}m_c^*}{\mathrm{d}\kappa_c} = -\frac{\sigma_p}{\kappa_c^2}\frac{h(m_c^0)}{(h(m_c^0))^2 + h'(m_c^0)}\left(1 + \frac{m_c^0 h'(m_c^0)}{h(m_c^0)}\right). \tag{26}$$

Since the hazard function for the standard normal distribution is strictly increasing, the last factor $1+\frac{m_c^0 h'(m_c^0)}{h(m_c^0)}$ above determines the sign of the derivative $\frac{\mathrm{d}m_c^*}{\mathrm{d}\kappa_c}$. In particular, $\frac{\mathrm{d}m_c^*}{\mathrm{d}\kappa_c} > 0$, i.e., the firm's optimal price is increasing in κ_c if and only if $1 + \frac{m_c^0 h'(m_c^0)}{h(m_c^0)} < 0$, and it is decreasing in κ_c if and only if $1 + \frac{m_c^0 h'(m_c^0)}{h(m_c^0)} > 0$. Since $\frac{m_c^0 h'(m_c^0)}{h(m_c^0)}$ is increasing in m_c^0 then

$$1 + \frac{m_c^0 h'(m_c^0)}{h(m_c^0)} < 0 \iff \quad m_c^0 < \underline{m}_c^0,$$

where $\underline{m}_c^0$ is the unique number such that $1 + \frac{m_c^0 h'(m_c^0)}{h(m_c^0)} = 0$. Consequently, $\lambda = \frac{1}{h(\underline{m}_c^0)} - \underline{m}_c^0$. □

Proof of Corollary 2.2. When the firm sets the price at m_c^*, the equilibrium demand is $1 - \Phi(m_c^0)$. m_c^0 is decreasing in κ_c. Since the CDF for the standard normal distribution is increasing, $\Phi(m_c^0)$ is decreasing in κ_c. Consequently, the equilibrium demand is increasing in κ_c. □

Proof of Corollary 2.3. Totally differentiation Equation (7) with respect to κ_c yields:

$$\left(1 + \frac{h'(m_c^0)}{h(m_c^0)^2}\right)\frac{\mathrm{d}m_c^0}{\mathrm{d}\kappa_c} = -\frac{\mu_p}{\sigma_p}.$$

Since the hazard function $h(\cdot)$ is increasing, it follows that $\frac{\mathrm{d}m_c^0}{\mathrm{d}\kappa_c} < 0$.

Next, totally differentiating the marketer's profit per consumer at the profit-maximizing price (either m_{inf}^* or m_{ad}^*) with respect to κ_c yields:

$$\begin{aligned}\frac{\mathrm{d}\pi_c^*}{\mathrm{d}\kappa_c} &= -\phi(m_c^0)\frac{\mathrm{d}m_c^0}{\mathrm{d}\kappa_c}\left(\frac{\sigma_p m_c^0}{\kappa_c} + \mu_p\right) + \left(1 - \Phi(m_c^0)\right)\frac{\sigma_p \frac{\mathrm{d}m_c^0}{\mathrm{d}\kappa_c}\kappa_c - \sigma_p m_c^0}{\kappa_c^2}\\ &= \left(1 - \Phi(m_c^0)\right)\left[\frac{\sigma_p}{\kappa_c} - h(m_c^0)\left(\frac{\sigma_p m_c^0}{\kappa_c} + \mu_p\right)\right]\frac{\mathrm{d}m_c^0}{\mathrm{d}\kappa_c} - \left(1 - \Phi(m_c^0)\right)\frac{\sigma_p m_c^0}{\kappa_c^2}.\end{aligned}$$

The factor $\frac{\sigma_p}{\kappa_c} - h(m_c^0)\left(\frac{\sigma_p m_c^0}{\kappa_c} + \mu_p\right)$ in the first summand equals zero by the definition of m_c^0. Hence, $\frac{\mathrm{d}\pi_c^*}{\mathrm{d}\kappa_c} = -\left(1 - \Phi(m_c^0)\right)\frac{\sigma_p m_c^0}{\kappa_c^2}$ and the sign of $\frac{\mathrm{d}\pi_c^*}{\mathrm{d}\kappa_c}$ is determined by the sign of m_c^0. To conclude, π_c^* is increasing with κ_c iff $m_c^0 < 0$, that is, $\kappa_c > \frac{\sigma_p}{\mu_p}\sqrt{\frac{\pi}{2}}$, and it is decreasing with κ_c iff $m_c^0 > 0$, that is, $\kappa_c < \frac{\sigma_p}{\mu_p}\sqrt{\frac{\pi}{2}}$. □

Proof of Proposition 3. As both campaigns reach the same population size we can focus on the firm's revenue per consumer at the campaign-specific revenue-maximizing price $\pi_c^* = (1-\Phi(m_c^0))m_c^*$.

For an established brand, $\frac{\sigma_p}{\mu_p}\sqrt{\frac{\pi}{2}} < 1$, and κ_c is always larger than 1, by Corollary 2.3, π_c^* is increasing with κ_c iff $\kappa_c > \frac{\sigma_p}{\mu_p}\sqrt{\frac{\pi}{2}}$. Hence $\pi_{\text{inf}}^* > \pi_{\text{ad}}^*$ iff $\kappa_{\text{inf}} > \kappa_{\text{ad}} > \frac{\sigma_p}{\mu_p}\sqrt{\frac{\pi}{2}}$.

For an unknown brand, if $\kappa_{\text{ad}} > \hat{\kappa}_1 > 1$, then an influencer campaign is more profitable only when the influencer campaign has an even higher level of learning friction, i.e., $\kappa_{\text{inf}} > \kappa_{\text{ad}}$, leading to the first case in the proposition, because the profit is increasing in κ_c, similarly to the case of an established brand.

If $\kappa_{\text{ad}} < \hat{\kappa}_1$, then an influencer campaign yields a higher profit than an advertising campaign if the learning friction κ_{inf} is outside the range between κ_{ad} and $\hat{\kappa}_{\text{ad}}$, as by Corollary 2.3, the profit is decreasing in κ_c for low values and increasing in κ_c for high values. If κ_{inf} is below this range, we are in case (ii), and if it is above, we are in case (iii).

□

Proof of Proposition 4. Recall that $m_c^* = \mu_p + \frac{\sigma_p}{\kappa_c} m_c^0$. The ex-ante consumer surplus is:

$$
\begin{aligned}
\mathbb{E}[CS] &= \mathbb{E}_{\mathbf{v},q_p,q_c}\left[\mathbb{1}_{\{\mu_p(\mathbf{v}) \geq m_c^*\}} \cdot (q_p - m_c^*)\right] = \mathbb{E}_{q_p,q_c}\left[D(m_c^*; q_p, q_c)\,(q_p - m_c^*)\right] \\
&= \mathbb{E}_{q_p,q_c}\left[\left(1 - \Phi\left(\frac{(m_c^* - \mu_p)\kappa_c^2 + \mu_p + \mu_c - q_p - q_c}{\sqrt{\frac{\tau_c^2}{n}}}\right)\right)(q_p - m_c^*)\right] \\
&= \mathbb{E}_{q_p}\left[\left(1 - \Phi\left(\frac{(m_c^* - \mu_p)\kappa_c^2 + \mu_p - q_p}{\sqrt{\sigma_c^2 + \frac{\tau_c^2}{n}}}\right)\right)(q_p - m_c^*)\right] \\
&= \left[1 - \Phi\left(\frac{m_c^*\kappa_c - \left(\kappa_c - \frac{1}{\kappa_c}\right)\mu_p}{\sigma_p}\right)\right] \cdot \left[\frac{1}{\kappa_c} h\left(\frac{m_c^*\kappa_c - \left(\kappa_c - \frac{1}{\kappa_c}\right)\mu_p}{\sigma_p}\right) - (m_c^* - \mu_p)\right] \\
&= \frac{\sigma_p}{\kappa_c}\left[1 - \Phi\left(\frac{m_c^* - \mu_p}{\sigma_p/\kappa_c} + \frac{\mu_p}{\kappa_c\sigma_p}\right)\right] \cdot \left[\frac{h\left(\frac{m_c^* - \mu_p}{\sigma_p/\kappa_c} + \frac{\mu_p}{\kappa_c\sigma_p}\right)}{\sigma_p} - \frac{m_c^* - \mu_p}{\sigma_p/\kappa_c}\right].
\end{aligned} \tag{27}
$$

□

Online Appendix – Equilibrium analysis with consumer heterogeneity

We extend the results from Section 3 to the case when consumers have idiosyncratic consumption utility for the product. We assume that the heterogeneity in consumption utility ϵ_i^p and heterogeneity in follower content utility ϵ_i^c are independently distributed. For ease of comparison, we keep the structure of the analysis as close as possible to Section 3.

Product demand A consumer i buys the product if $\mathbb{E}[q_p|\mathbf{v}, \epsilon_i^p] + \epsilon_i^p \geq m$. Thus, while a consumer understands her idiosyncratic taste for a product, she only observes other followers' overall consumption utilities. We denote the average consumption utility of all items in $\mathbf{v}$ other than follower i's utility by $\bar{\mathbf{v}}_{-i} = \frac{1}{n-1} \sum_{j \neq i} v_i$. Then, analogous to Proposition 1, we characterize the distribution of $q_p \big| \mathbf{v}, \epsilon_i^p$. The derivation below focuses on an influencer campaign with variance τ_{inf}^2 for ϵ_i^c. The derivation for an ad campaign (omitted) follows a similar approach, but without any social learning from other users:

Lemma OA.1 (Social learning)**.** *Consumer i's posterior belief about q_p given $\mathbf{v}$ and ϵ_i^p is*

$$q_p \big| \mathbf{v}, \epsilon_i^p \sim \mathcal{N}\left(\mu_p(\mathbf{v}, \epsilon_i^p), \sigma_p^2(\mathbf{v}, \epsilon_i^p)\right)$$

where

$$\begin{aligned}
\mu_p(\mathbf{v}, \epsilon_i^p) &= \mu_p + \frac{\frac{n-1}{n} \frac{\tau_{\text{inf}}^2}{\tau_{\text{inf}}^2 + \frac{\tau_p^2}{n}} \bar{\mathbf{v}}_{-i} + \frac{1}{n} \frac{\tau_{\text{inf}}^2 + \tau_p^2}{\tau_{\text{inf}}^2 + \frac{\tau_p^2}{n}} (v_i - \epsilon_i^p) - \mu_c - \mu_p}{(\kappa_c^{\text{het}})^2} \\
\sigma_p(\mathbf{v}, \epsilon_i^p) &= \sigma_p^2 \left(1 - \frac{1}{(\kappa_c^{\text{het}})^2}\right).
\end{aligned}$$

Proof. We calculate the posterior belief of the vector (q_c, q_p) after observing $\mathbf{v}^c$ given the prior

$$\mathbf{q} = \begin{pmatrix} q_c \\ q_p \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_c \\ \mu_p \end{pmatrix}, \begin{pmatrix} \sigma_c^2 & 0 \\ 0 & \sigma_p^2 \end{pmatrix}\right).$$

Consumer i's content utility net of the product taste shock is normally distributed $v_i - \epsilon_i^p \sim \mathcal{N}(q_c + q_p, \tau_{\text{inf}}^2)$. Also, each consumer i's observation of any other consumer j's content utility $v_j|\mathbf{q}$

is normally distributed with $\mathbb{E}[v_j] = q_p + q_c$, and $\text{Var}(v_j) = \tau_p^2 + \tau_{\text{inf}}^2$ for $j = 1, \ldots, N$, $j \neq i$. By Bayes' rule, the probability density of the vector $\mathbf{q}$ conditional on the observations $\mathbf{v}^c$ and ϵ_i^p is given by

$$f_{\mathbf{q}|\mathbf{v}^c,\epsilon_i^p}(\mathbf{q}) = \frac{f_{\mathbf{v}_{-i}^c|\mathbf{q}}(\mathbf{v}_{-i}^c) f_{\mathbf{q}|v_i,\epsilon_i^p}(\mathbf{q})}{f_{\mathbf{v}_{-i}^c}(\mathbf{v}_{-i}^c)} = \frac{f_{\mathbf{v}_{-i}^c|\mathbf{q}}(\mathbf{v}_{-i}^c)}{f_{\mathbf{v}_{-i}^c}(\mathbf{v}_{-i}^c)} \frac{f_{v_i,\epsilon_i^p|\mathbf{q}}(v_i - \epsilon_i^p) f_{\mathbf{q}}(\mathbf{q})}{f_{v_i,\epsilon_i^p}(v_i - \epsilon_i^p)}$$
$$\propto f_{\mathbf{v}_{-i}^c|\mathbf{q}}(\mathbf{v}_{-i}^c) \cdot f_{v_i,\epsilon_i^p|\mathbf{q}}(v_i - \epsilon_i^p) \cdot f_{\mathbf{q}}(\mathbf{q}). \tag{28}$$

By the fact that the v_j's are independently drawn from $\mathcal{N}(q_c + q_p, \tau_p^2 + \tau_{\text{inf}}^2)$, we know that

$$f_{\mathbf{v}_{-i}^c|\mathbf{q}}(\mathbf{v}_{-i}^c) = \prod_{j \neq i} f_{v_j|\mathbf{q}}(v_j) \propto \prod_{j \neq i} \exp\left(-\frac{(v_j - q_p - q_c)^2}{2(\tau_p^2 + \tau_{\text{inf}}^2)}\right) = \exp\left(-\frac{\sum_{j \neq i}(v_j - q_p - q_c)^2}{2(\tau_p^2 + \tau_{\text{inf}}^2)}\right).$$

Then the likelihood function in (28) becomes

$$f_{\mathbf{v}_{-i}^c|\mathbf{q}}(\mathbf{v}_{-i}^c) \cdot f_{v_i,\epsilon_i^p|\mathbf{q}}(v_i - \epsilon_i^p) \propto \exp\left(-\frac{\sum_{j \neq i}(v_j - q_p - q_c)^2}{2(\tau_p^2 + \tau_{\text{inf}}^2)} - \frac{(v_i - \epsilon_i^p - q_p - q_c)^2}{2\tau_{\text{inf}}^2}\right). \tag{29}$$

Also, by the assumption on the prior distribution of $\mathbf{q}$,

$$f_{\mathbf{q}}(\mathbf{q}) \propto \exp\left[-\frac{1}{2}\begin{pmatrix} q_c - \mu_c & q_p - \mu_p \end{pmatrix} \begin{pmatrix} \frac{1}{\sigma_c^2} & 0 \\ 0 & \frac{1}{\sigma_p^2} \end{pmatrix} \begin{pmatrix} q_c - \mu_c \\ q_p - \mu_p \end{pmatrix}\right]. \tag{30}$$

We expand the exponents in (29) and (30) and collect terms:

$$-\frac{\sum_{j \neq i}(v_j - q_p - q_c)^2}{2(\tau_p^2 + \tau_{\text{inf}}^2)} = -\frac{1}{2(\tau_p^2 + \tau_{\text{inf}}^2)} \left(\underbrace{\sum_{j \neq i}(q_c + q_p)^2}_{\text{(I)}} \underbrace{-2\sum_{j \neq i} v_j (q_c + q_p)}_{\text{(II)}} + \underbrace{\sum_{j \neq i} v_j^2}_{\text{(III)}} \right)$$
$$= -\frac{n-1}{2(\tau_p^2 + \tau_{\text{inf}}^2)} \mathbf{q}' \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \mathbf{q} + \frac{\sum_{j \neq i} v_j}{\tau_p^2 + \tau_{\text{inf}}^2} \begin{pmatrix} 1 & 1 \end{pmatrix} \mathbf{q} - \frac{\sum_{j \neq i} v_j^2}{2(\tau_p^2 + \tau_{\text{inf}}^2)}, \tag{31}$$

Hence, the exponent in (29) is reduced to

$$
\begin{aligned}
&-\frac{\sum_{j\neq i}(v_j-q_p-q_c)^2}{2(\tau_p^2+\tau_{\text{inf}}^2)}-\frac{(v_i-\epsilon_i^p-q_p-q_c)^2}{2\tau_{\text{inf}}^2}\\
&=\underbrace{-\left(\frac{n-1}{2(\tau_p^2+\tau_{\text{inf}}^2)}+\frac{1}{2\tau_{\text{inf}}^2}\right)\mathbf{q}'\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix}\mathbf{q}}_{(\text{I}_1)}+\underbrace{\left(\frac{\sum_{j\neq i}v_j}{\tau_p^2+\tau_{\text{inf}}^2}+\frac{v_i-\epsilon_i^p}{\tau_{\text{inf}}^2}\right)\begin{pmatrix}1 & 1\end{pmatrix}\mathbf{q}}_{(\text{II}_1)}-\underbrace{\left(\frac{\sum_{j\neq i}v_j^2}{2(\tau_p^2+\tau_{\text{inf}}^2)}+\frac{(v_i-\epsilon_i^p)^2}{2\tau_{\text{inf}}^2}\right)}_{(\text{III}_1)}.
\end{aligned}
\tag{32}
$$

As the summand (III_1) is deterministic, we can omit it.

Next, we expand the exponent in (30) as follows

$$
\begin{aligned}
&-\frac{1}{2}\begin{pmatrix}q_c-\mu_c & q_p-\mu_p\end{pmatrix}\begin{pmatrix}\frac{1}{\sigma_c^2} & 0\\ 0 & \frac{1}{\sigma_p^2}\end{pmatrix}\begin{pmatrix}q_c-\mu_c\\ q_p-\mu_p\end{pmatrix}\\
&=-\frac{1}{2}\mathbf{q}'\begin{pmatrix}\frac{1}{\sigma_c^2} & 0\\ 0 & \frac{1}{\sigma_p^2}\end{pmatrix}\mathbf{q}+\begin{pmatrix}\mu_c & \mu_p\end{pmatrix}\begin{pmatrix}\frac{1}{\sigma_c^2} & 0\\ 0 & \frac{1}{\sigma_p^2}\end{pmatrix}\mathbf{q}-\frac{1}{2}\begin{pmatrix}\mu_c & \mu_p\end{pmatrix}\begin{pmatrix}\frac{1}{\sigma_c^2} & 0\\ 0 & \frac{1}{\sigma_p^2}\end{pmatrix}\begin{pmatrix}\mu_c\\ \mu_p\end{pmatrix}\\
&=\underbrace{-\frac{1}{2}\mathbf{q}'\begin{pmatrix}\frac{1}{\sigma_c^2} & 0\\ 0 & \frac{1}{\sigma_p^2}\end{pmatrix}\mathbf{q}}_{(\text{I}_2)}+\underbrace{\begin{pmatrix}\frac{\mu_c}{\sigma_c^2} & \frac{\mu_p}{\sigma_p^2}\end{pmatrix}\mathbf{q}}_{(\text{II}_2)}\underbrace{-\frac{1}{2}\left(\frac{\mu_c^2}{\sigma_c^2}+\frac{\mu_p^2}{\sigma_p^2}\right)}_{(\text{III}_2)}.
\end{aligned}
\tag{33}
$$

We then add the exponents calculated above:

$$
\text{I}_1+\text{I}_2=-\frac{1}{2}\mathbf{q}'\begin{pmatrix}\frac{n-1}{\tau_p^2+\tau_{\text{inf}}^2}+\frac{1}{\tau_{\text{inf}}^2}+\frac{1}{\sigma_c^2} & \frac{n-1}{\tau_p^2+\tau_{\text{inf}}^2}+\frac{1}{\tau_{\text{inf}}^2}\\ \frac{n-1}{\tau_p^2+\tau_{\text{inf}}^2}+\frac{1}{\tau_{\text{inf}}^2} & \frac{n-1}{\tau_p^2+\tau_{\text{inf}}^2}+\frac{1}{\tau_{\text{inf}}^2}+\frac{1}{\sigma_p^2}\end{pmatrix}\mathbf{q}, \tag{34}
$$

$$
\text{II}_1+\text{II}_2=\left(\frac{\sum_{j\neq i}v_j}{\tau_p^2+\tau_{\text{inf}}^2}+\frac{v_i-\epsilon_i^p}{\tau_{\text{inf}}^2}+\frac{\mu_c}{\sigma_c^2},\quad \frac{\sum_{j\neq i}v_j}{\tau_p^2+\tau_{\text{inf}}^2}+\frac{v_i-\epsilon_i^p}{\tau_{\text{inf}}^2}+\frac{\mu_p}{\sigma_p^2}\right)\mathbf{q}, \tag{35}
$$

In order to determine the mean and variance-covariance matrix of $\mathbf{q}|\mathbf{v}^c,\epsilon_i^p$, we first calculate the inverse of the matrix in (34) to derive $\Sigma_{\mathbf{q}|\mathbf{v}^c,\epsilon_i^p}$ and then rewrite the vector in (35) as $\mu'_{\mathbf{q}|\mathbf{v}^c,\epsilon_i^p}\Sigma^{-1}_{\mathbf{q}|\mathbf{v}^c,\epsilon_i^p}$

in order to derive the mean $\mu_{\mathbf{q}|\mathbf{v}^c,\epsilon_i^p}$. The posterior's variance-covariance matrix equals:

$$\Sigma_{\mathbf{q}|\mathbf{v}^c} = \begin{pmatrix} \frac{n-1}{\tau_p^2+\tau_{\text{inf}}^2} + \frac{1}{\tau_{\text{inf}}^2} + \frac{1}{\sigma_c^2} & \frac{n-1}{\tau_p^2+\tau_{\text{inf}}^2} + \frac{1}{\tau_{\text{inf}}^2} \\ \frac{n-1}{\tau_p^2+\tau_{\text{inf}}^2} + \frac{1}{\tau_{\text{inf}}^2} & \frac{n-1}{\tau_p^2+\tau_{\text{inf}}^2} + \frac{1}{\tau_{\text{inf}}^2} + \frac{1}{\sigma_p^2} \end{pmatrix}^{-1}$$

$$= \left[\frac{1}{\sigma_c^2\sigma_p^2} + \left(\frac{n-1}{\tau_p^2+\tau_{\text{inf}}^2} + \frac{1}{\tau_{\text{inf}}^2}\right)\left(\frac{1}{\sigma_c^2} + \frac{1}{\sigma_p^2}\right)\right]^{-1} \begin{pmatrix} \frac{n-1}{\tau_p^2+\tau_{\text{inf}}^2} + \frac{1}{\tau_{\text{inf}}^2} + \frac{1}{\sigma_p^2} & -\frac{n-1}{\tau_p^2+\tau_{\text{inf}}^2} - \frac{1}{\tau_{\text{inf}}^2} \\ -\frac{n-1}{\tau_p^2+\tau_{\text{inf}}^2} - \frac{1}{\tau_{\text{inf}}^2} & \frac{n-1}{\tau_p^2+\tau_{\text{inf}}^2} + \frac{1}{\tau_{\text{inf}}^2} + \frac{1}{\sigma_c^2} \end{pmatrix}$$

$$= \frac{\sigma_c^2\sigma_p^2\tau_{\text{inf}}^2(\tau_p^2+\tau_{\text{inf}}^2)}{(n\tau_{\text{inf}}^2+\tau_p^2)(\sigma_c^2+\sigma_p^2)+\tau_{\text{inf}}^2(\tau_p^2+\tau_{\text{inf}}^2)} \begin{pmatrix} \frac{n-1}{\tau_p^2+\tau_{\text{inf}}^2} + \frac{1}{\tau_{\text{inf}}^2} + \frac{1}{\sigma_p^2} & -\frac{n-1}{\tau_p^2+\tau_{\text{inf}}^2} - \frac{1}{\tau_{\text{inf}}^2} \\ -\frac{n-1}{\tau_p^2+\tau_{\text{inf}}^2} - \frac{1}{\tau_{\text{inf}}^2} & \frac{n-1}{\tau_p^2+\tau_{\text{inf}}^2} + \frac{1}{\tau_{\text{inf}}^2} + \frac{1}{\sigma_c^2} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{[n\sigma_p^2\tau_{\text{inf}}^2+\tau_{\text{inf}}^2(\tau_p^2+\tau_{\text{inf}}^2)+\tau_p^2\sigma_p^2]\sigma_c^2}{(n\tau_{\text{inf}}^2+\tau_p^2)(\sigma_c^2+\sigma_p^2)+\tau_{\text{inf}}^2(\tau_p^2+\tau_{\text{inf}}^2)} & -\frac{(n\tau_{\text{inf}}^2+\tau_p^2)\sigma_c^2\sigma_p^2}{(n\tau_{\text{inf}}^2+\tau_p^2)(\sigma_c^2+\sigma_p^2)+\tau_{\text{inf}}^2(\tau_p^2+\tau_{\text{inf}}^2)} \\ -\frac{(n\tau_{\text{inf}}^2+\tau_p^2)\sigma_c^2\sigma_p^2}{(n\tau_{\text{inf}}^2+\tau_p^2)(\sigma_c^2+\sigma_p^2)+\tau_{\text{inf}}^2(\tau_p^2+\tau_{\text{inf}}^2)} & \frac{[n\sigma_c^2\tau_{\text{inf}}^2+\tau_{\text{inf}}^2(\tau_p^2+\tau_{\text{inf}}^2)+\tau_p^2\sigma_c^2]\sigma_p^2}{(n\tau_{\text{inf}}^2+\tau_p^2)(\sigma_c^2+\sigma_p^2)+\tau_{\text{inf}}^2(\tau_p^2+\tau_{\text{inf}}^2)} \end{pmatrix},$$

The transpose of the posterior mean of $\mathbf{q}|\mathbf{v}^c, \epsilon_i^p$ is

$$\mu'_{\mathbf{q}|\mathbf{v}^c,\epsilon_i^p} = \left(\frac{\sum_{j\neq i} v_j}{\tau_p^2+\tau_{\text{inf}}^2} + \frac{v_i-\epsilon_i^p}{\tau_{\text{inf}}^2} + \frac{\mu_c}{\sigma_c^2} \quad \frac{\sum_{j\neq i} v_j}{\tau_p^2+\tau_{\text{inf}}^2} + \frac{v_i-\epsilon_i^p}{\tau_{\text{inf}}^2} + \frac{\mu_p}{\sigma_p^2}\right)\left(\Sigma^{-1}_{\mathbf{q}|\mathbf{v}^c,\epsilon_i^p}\right)^{-1}$$

$$= \left(\frac{\sum_{j\neq i} v_j}{\tau_p^2+\tau_{\text{inf}}^2} + \frac{v_i-\epsilon_i^p}{\tau_{\text{inf}}^2} + \frac{\mu_c}{\sigma_c^2} \quad \frac{\sum_{j\neq i} v_j}{\tau_p^2+\tau_{\text{inf}}^2} + \frac{v_i-\epsilon_i^p}{\tau_{\text{inf}}^2} + \frac{\mu_p}{\sigma_p^2}\right)$$

$$\cdot \begin{pmatrix} \frac{[n\sigma_p^2\tau_{\text{inf}}^2+\tau_{\text{inf}}^2(\tau_p^2+\tau_{\text{inf}}^2)+\tau_p^2\sigma_p^2]\sigma_c^2}{(n\tau_{\text{inf}}^2+\tau_p^2)(\sigma_c^2+\sigma_p^2)+\tau_{\text{inf}}^2(\tau_p^2+\tau_{\text{inf}}^2)} & -\frac{(n\tau_{\text{inf}}^2+\tau_p^2)\sigma_c^2\sigma_p^2}{(n\tau_{\text{inf}}^2+\tau_p^2)(\sigma_c^2+\sigma_p^2)+\tau_{\text{inf}}^2(\tau_p^2+\tau_{\text{inf}}^2)} \\ -\frac{(n\tau_{\text{inf}}^2+\tau_p^2)\sigma_c^2\sigma_p^2}{(n\tau_{\text{inf}}^2+\tau_p^2)(\sigma_c^2+\sigma_p^2)+\tau_{\text{inf}}^2(\tau_p^2+\tau_{\text{inf}}^2)} & \frac{[n\sigma_c^2\tau_{\text{inf}}^2+\tau_{\text{inf}}^2(\tau_p^2+\tau_{\text{inf}}^2)+\tau_p^2\sigma_c^2]\sigma_p^2}{(n\tau_{\text{inf}}^2+\tau_p^2)(\sigma_c^2+\sigma_p^2)+\tau_{\text{inf}}^2(\tau_p^2+\tau_{\text{inf}}^2)} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{\left[n\sigma_p^2\tau_{\text{inf}}^2+\tau_{\text{inf}}^2(\tau_p^2+\tau_{\text{inf}}^2)+\tau_p^2\sigma_p^2\right]\mu_c + \left[\left(\sum_{j\neq i} v_j\right)\tau_{\text{inf}}^2 + (v_i-\epsilon_i^p)(\tau_p^2+\tau_{\text{inf}}^2) - (n\tau_{\text{inf}}^2+\tau_p^2)\mu_p\right]\sigma_c^2}{(n\tau_{\text{inf}}^2+\tau_p^2)(\sigma_c^2+\sigma_p^2)+\tau_{\text{inf}}^2(\tau_p^2+\tau_{\text{inf}}^2)} \\ \frac{\left[n\sigma_c^2\tau_{\text{inf}}^2+\tau_{\text{inf}}^2(\tau_p^2+\tau_{\text{inf}}^2)+\tau_p^2\sigma_c^2\right]\mu_p + \left[\left(\sum_{j\neq i} v_j\right)\tau_{\text{inf}}^2 + (v_i-\epsilon_i^p)(\tau_p^2+\tau_{\text{inf}}^2) - (n\tau_{\text{inf}}^2+\tau_p^2)\mu_c\right]\sigma_p^2}{(n\tau_{\text{inf}}^2+\tau_p^2)(\sigma_c^2+\sigma_p^2)+\tau_{\text{inf}}^2(\tau_p^2+\tau_{\text{inf}}^2)} \end{pmatrix}'.$$

□

Comparing these expressions to the ones in Proposition 1, where consumers have homogeneous tastes for the product, we see that consumers learn differently from other followers' signals, leading to a quality estimate of $\bar{\mathbf{v}}_{-i}^c - \mu_c$, and her own signal, leading to a quality estimate of $v_i - \epsilon_i^p - \mu_c$. The signals from other consumers' content utility are more noisy because they have different tastes of consumption. This is reflected in the fact that $\frac{\tau_{\text{inf}}^2 n+\tau_p^2}{\tau_{\text{inf}}^2(n-1)}(\sigma_c^2+\sigma_p^2)+\frac{\tau_{\text{inf}}^2+\tau_p^2}{n-1} > \frac{n}{n-1}(\sigma_c^2+\sigma_p^2)+\frac{\tau_{\text{inf}}^2}{n-1}$. Similarly, the consumer's own signal is stronger which is reflected in the fact that $\frac{\tau_{\text{inf}}^2 n+\tau_p^2}{\tau_{\text{inf}}^2+\tau_p^2}(\sigma_c^2+\sigma_p^2)+\tau_{\text{inf}}^2 < n(\sigma_c^2+\sigma_p^2)+\tau_{\text{inf}}^2$.

Using the result on the posterior belief, and noting that $D(m;q_p,q_c)=\sum_i \Pr\left(\mu_p(\mathbf{v},\epsilon_i^p)+\epsilon_i^p \geq m|q_p,q_c\right)$, we characterize the product's demand analogously to Proposition 1:

Proposition OA.1 (Demand)**.** *When the firm sets its price at m, the expected market demand is*

$$\mathbb{E}_{q_p,q_c}[D(m;q_p,q_c)] = N\left(1-\Phi\left(\frac{m-\mu_p}{\sigma_p/\tilde{\kappa}_c^{\text{het}}}\right)\right). \tag{36}$$

Proof. By Lemma OA.1, the demand is determined by these three random variables:

$$(i)\ \bar{\mathbf{v}}_{-i}|q_p,q_c \sim \mathcal{N}\left(q_p+q_c,\frac{\tau_p^2+\tau_{\text{inf}}^2}{n-1}\right),\ (ii)\ (v_i-\epsilon_i^p)|q_p,q_c\sim\mathcal{N}(q_p+q_c,\tau_{\text{inf}}^2),\ (iii)\ \epsilon_i^p\sim\mathcal{N}\left(0,\tau_p^2\right).$$

which are mutually independent conditional on q_p and q_c. We define $q=q_p+q_c$, $x=\frac{\bar{\mathbf{v}}_{-1}-q}{\sqrt{\frac{\tau_p^2+\tau_{\text{inf}}^2}{N-1}}}$, $y=\frac{v_i-\epsilon_i^p-q}{\tau_{\text{inf}}}$, $z=\frac{\epsilon_i^p}{\tau_p}$. We write the product's demand given price m and qualities q_c and q_p as:

$$\begin{aligned} D(m;q_p,q_c) = N\cdot \iiint \mathbb{1}\left(\Sigma_x\left(x\sqrt{\frac{\tau_p^2+\tau_{\text{inf}}^2}{n-1}}+q-\mu_c\right)+\Sigma_y\left(y\tau_p+q-\mu_c\right)+\Sigma_p+z\tau_p\geq m\right) \\ \cdot\,\phi(x)\phi(y)\phi(z)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z \end{aligned} \tag{37}$$

where

$$\begin{aligned}
\Sigma_x &= \frac{\tau_{\text{inf}}^2\sigma_p^2}{(\sigma_c^2+\sigma_p^2)\tau_{\text{inf}}^2+(\tau_{\text{inf}}^2+\sigma_c^2+\sigma_p^2)\frac{\tau_p^2+\tau_{\text{inf}}^2}{n-1}} \\
\Sigma_y &= \frac{\frac{\tau_{\text{inf}}^2+\tau_p^2}{n-1}\sigma_p^2}{(\sigma_c^2+\sigma_p^2)\tau_{\text{inf}}^2+(\tau_{\text{inf}}^2+\sigma_c^2+\sigma_p^2)\frac{\tau_p^2+\tau_{\text{inf}}^2}{n-1}} \\
\Sigma_p &= \frac{\sigma_c^2\tau_{\text{inf}}^2+(\tau_{\text{inf}}^2+\sigma_c^2)\frac{\tau_p^2+\tau_{\text{inf}}^2}{n-1}}{(\sigma_c^2+\sigma_p^2)\tau_{\text{inf}}^2+(\tau_{\text{inf}}^2+\sigma_c^2+\sigma_p^2)\frac{\tau_p^2+\tau_{\text{inf}}^2}{n-1}}\mu_p.
\end{aligned}$$

Since

$$\mathbb{1}\left(\Sigma_x\left(x\sqrt{\frac{\tau_p^2+\tau_{\text{inf}}^2}{n-1}}+q-\mu_c\right)+\Sigma_y\left(y\tau_p+q-\mu_c\right)+\Sigma_p+z\tau_p\geq m\right)$$
$$=Pr\left(x\geq\frac{m-\Sigma_p-z\tau_p-\Sigma_y\left(y\tau_p+q-\mu_c\right)-\Sigma_x(q-\mu_c)}{\Sigma_x\sqrt{\frac{\tau_p^2+\tau_{\text{inf}}^2}{n-1}}}\right),$$

integrating over x yields

$$D(m;q_p,q_c)=N\iint\left(1-\Phi\left(\frac{m-\Sigma_p-z\tau_p-\Sigma_y\left(y\tau_p+q-\mu_c\right)-\Sigma_x(q-\mu_c)}{\Sigma_x\sqrt{\frac{\tau_p^2+\tau_{\text{inf}}^2}{n-1}}}\right)\right)\phi(y)\phi(z)\,\mathrm{d}y\,\mathrm{d}z.$$

Integrating over y, by the formula $\int\Phi(a_1+b_1y)\phi(y)\,\mathrm{d}y=\Phi\left(\frac{a_1}{\sqrt{1+b_1^2}}\right)$, we get

$$D(m;q_p,q_c)=N\int\left(1-\Phi\left(\frac{a_1}{\sqrt{1+b_1^2}}\right)\right)\phi(z)\,\mathrm{d}z,$$

with

$$a_1=\frac{m-\Sigma_p-z\tau_p-(\Sigma_x+\Sigma_y)(q-\mu_c)}{\Sigma_x\sqrt{\frac{\tau_p^2+\tau_{\text{inf}}^2}{n-1}}},$$
$$b_1=-\frac{\Sigma_y\tau_p}{\Sigma_x\sqrt{\frac{\tau_p^2+\tau_{\text{inf}}^2}{n-1}}}=-\frac{\sqrt{\frac{\tau_p^2+\tau_{\text{inf}}^2}{n-1}}}{\tau_{\text{inf}}}\Rightarrow\sqrt{1+b_1^2}=\sqrt{\frac{\tau_{\text{inf}}^2+\frac{\tau_p^2+\tau_{\text{inf}}^2}{n-1}}{\tau_{\text{inf}}^2}}.$$

Using the formula $\int\Phi(a_2+b_2z)\phi(z)\,\mathrm{d}z=\Phi\left(\frac{a_2}{\sqrt{1+b_2^2}}\right)$ again, we can further simplify

$$D(m;q_p,q_c)=N\left(1-\Phi\left(\frac{a_2}{\sqrt{1+b_2^2}}\right)\right),\tag{38}$$

where

$$a_2 = \frac{m - \Sigma_p - (\Sigma_x + \Sigma_y)(q - \mu_c)}{\Sigma_x \sqrt{\frac{\tau_p^2 + \tau_{\text{inf}}^2}{n-1}}}$$

$$= (m - \Sigma_p) \frac{(\sigma_c^2 + \sigma_p^2)\tau_{\text{inf}}^2 + (\tau_{\text{inf}}^2 + \sigma_c^2 + \sigma_p^2)\frac{\tau_p^2 + \tau_{\text{inf}}^2}{n-1}}{\tau_{\text{inf}} \sigma_p^2 \sqrt{\frac{\tau_p^2 + \tau_{\text{inf}}^2}{n-1}} \sqrt{\tau_{\text{inf}}^2 + \frac{\tau_p^2 + \tau_{\text{inf}}^2}{n-1}}} - (q - \mu_c) \sqrt{\frac{\frac{\tau_p^2 + \tau_{\text{inf}}^2}{n-1} + \tau_{\text{inf}}^2}{\tau_{\text{inf}}^2 \frac{\tau_p^2 + \tau_{\text{inf}}^2}{n-1}}},$$

$$b_2 = -\frac{\tau_p}{\tau_{\text{inf}}} \frac{(\sigma_c^2 + \sigma_p^2)\tau_{\text{inf}}^2 + (\tau_{\text{inf}}^2 + \sigma_c^2 + \sigma_p^2)\frac{\tau_p^2 + \tau_{\text{inf}}^2}{n-1}}{\sigma_p^2 \sqrt{\frac{\tau_p^2 + \tau_{\text{inf}}^2}{n-1}} \sqrt{\tau_{\text{inf}}^2 + \frac{\tau_p^2 + \tau_{\text{inf}}^2}{n-1}}},$$

$$\sqrt{1 + b_2^2} = \sqrt{1 + \frac{\tau_p^2}{\tau_{\text{inf}}^2} \frac{\left((\sigma_c^2 + \sigma_p^2)\tau_{\text{inf}}^2 + (\tau_{\text{inf}}^2 + \sigma_c^2 + \sigma_p^2)\frac{\tau_p^2 + \tau_{\text{inf}}^2}{n-1}\right)^2}{\sigma_p^4 \frac{\tau_p^2 + \tau_{\text{inf}}^2}{n-1} \left(\tau_{\text{inf}}^2 + \frac{\tau_p^2 + \tau_{\text{inf}}^2}{n-1}\right)}}.$$

Next, we can calculate the firm's expected demand $\mathbb{E}_{q_p,q_c}[D(m; q_p, q_c)]$. Define the normalized q as $w = \frac{q - \mu_c - \mu_p}{\sqrt{\sigma_c^2 + \sigma_p^2}} \sim \mathcal{N}(0, 1)$. Using the same integration formula again, we get

$$\mathbb{E}_{q_p,q_c}[D(m; q_p, q_c)] = N \left(1 - \Phi\left(\frac{a_3}{\sqrt{1 + b_3^2}}\right)\right),$$

where

$$b_3 = -\sqrt{\sigma_c^2 + \sigma_p^2} \sqrt{\frac{\frac{\frac{\tau_p^2 + \tau_{\text{inf}}^2}{n-1} + \tau_{\text{inf}}^2}{\tau_{\text{inf}}^2 \frac{\tau_p^2 + \tau_{\text{inf}}^2}{n-1}}}{1 + \frac{\tau_p^2}{\tau_{\text{inf}}^2} \frac{\left((\sigma_c^2 + \sigma_p^2)\tau_{\text{inf}}^2 + (\tau_{\text{inf}}^2 + \sigma_c^2 + \sigma_p^2)\frac{\tau_p^2 + \tau_{\text{inf}}^2}{n-1}\right)^2}{\sigma_p^4 \frac{\tau_p^2 + \tau_{\text{inf}}^2}{n-1} \left(\tau_{\text{inf}}^2 + \frac{\tau_p^2 + \tau_{\text{inf}}^2}{n-1}\right)}}},$$

$$= -\frac{\sigma_p^2 \left(\tau_{\text{inf}}^2 + \frac{\tau_p^2 + \tau_{\text{inf}}^2}{n-1}\right) \sqrt{\sigma_c^2 + \sigma_p^2}}{\sqrt{\tau_{\text{inf}}^2 \sigma_p^4 \left(\tau_{\text{inf}}^2 + \frac{\tau_p^2 + \tau_{\text{inf}}^2}{n-1}\right) \frac{\tau_p^2 + \tau_{\text{inf}}^2}{n-1} + \tau_p^2 \left((\sigma_c^2 + \sigma_p^2)\tau_{\text{inf}}^2 + (\tau_{\text{inf}}^2 + \sigma_c^2 + \sigma_p^2)\frac{\tau_p^2 + \tau_{\text{inf}}^2}{n-1}\right)^2}}$$

$$a_3 = \frac{(m - \Sigma_p)\left[(\sigma_c^2 + \sigma_p^2)\tau_{\text{inf}}^2 + (\tau_{\text{inf}}^2 + \sigma_c^2 + \sigma_p^2)\frac{\tau_p^2 + \tau_{\text{inf}}^2}{n-1}\right]}{\sqrt{\tau_{\text{inf}}^2 \sigma_p^4 \left(\tau_{\text{inf}}^2 + \frac{\tau_p^2 + \tau_{\text{inf}}^2}{n-1}\right) \frac{\tau_p^2 + \tau_{\text{inf}}^2}{n-1} + \tau_p^2 \left((\sigma_c^2 + \sigma_p^2)\tau_{\text{inf}}^2 + (\tau_{\text{inf}}^2 + \sigma_c^2 + \sigma_p^2)\frac{\tau_p^2 + \tau_{\text{inf}}^2}{n-1}\right)^2}} + \frac{b_3 \mu_p}{\sqrt{\sigma_c^2 + \sigma_p^2}}.$$

Hence,

$$\sqrt{1+b_3^2} = \sqrt{1 + \frac{\sigma_p^4(\sigma_c^2+\sigma_p^2)\left(\frac{\tau_p^2+\tau_{\inf}^2}{n-1}+\tau_{\inf}^2\right)^2}{\tau_{\inf}^2\sigma_p^4\left(\tau_{\inf}^2+\frac{\tau_p^2+\tau_{\inf}^2}{n-1}\right)\frac{\tau_p^2+\tau_{\inf}^2}{n-1}+\tau_p^2\left((\sigma_c^2+\sigma_p^2)\tau_{\inf}^2+(\tau_{\inf}^2+\sigma_c^2+\sigma_p^2)\frac{\tau_p^2+\tau_{\inf}^2}{n-1}\right)^2}}.$$

and

$$\frac{a_3}{\sqrt{1+b_3^2}} = (m-\mu_p)\sqrt{\frac{(\sigma_c^2+\sigma_p^2)\tau_{\inf}^2+(\tau_{\inf}^2+\sigma_c^2+\sigma_p^2)\frac{\tau_p^2+\tau_{\inf}^2}{n-1}}{\sigma_p^4\left(\frac{\tau_p^2+\tau_{\inf}^2}{n-1}+\tau_{\inf}^2\right)+\tau_p^2\left[(\sigma_c^2+\sigma_p^2)\tau_{\inf}^2+(\tau_{\inf}^2+\sigma_c^2+\sigma_p^2)\frac{\tau_p^2+\tau_{\inf}^2}{n-1}\right]}}.$$

□

One can verify that the expression above equals $(m-\mu_p)\frac{\tilde{\kappa}_c^{\text{het}}}{\sigma_p}$ with

$$\tilde{\kappa}_c^{\text{het}} = \frac{\kappa_c^{\text{het}}}{\sqrt{1+(\kappa_c^{\text{het}})^2\frac{\tau_p^2}{\sigma_p^2}}} \qquad (\kappa_c^{\text{het}})^2 = \frac{\sigma_c^2+\tau_c^2\dfrac{1+\frac{\tau_p^2}{\tau_c^2}}{n+\frac{\tau_p^2}{\tau_c^2}}}{\sigma_p^2}+1.$$

Proof of Proposition 5. We derive the optimal price by comparing the expected demands in Equations (5) and (36). When the normalized learning friction $\tilde{\kappa}_c^{\text{het}}$ in a heterogeneous market substitutes for the relative learning friction in a homogeneous one, the marketer expects identical demands. It follows that the marketer would set its price by optimizing the same profit function while replacing κ_c with $\tilde{\kappa}_c^{\text{het}}$. The unique profit-maximizing price in a heterogeneous market has the same form as in the homogeneous market. □

Proof of Lemma 2. For part 1, $\tilde{\kappa}_c^{\text{het}} \leq \kappa_c$ iff

$$\frac{\frac{\sigma_{\mathrm{c}}^2+\tau_{\mathrm{c}}^2\frac{1+\frac{\tau_p^2}{\tau_{\mathrm{c}}^2}}{n+\frac{\tau_p^2}{\tau_{\mathrm{c}}^2}}}{\sigma_p^2}+1}{1+\left(\frac{\sigma_{\mathrm{c}}^2+\tau_{\mathrm{c}}^2\frac{1+\frac{\tau_p^2}{\tau_{\mathrm{c}}^2}}{n+\frac{\tau_p^2}{\tau_{\mathrm{c}}^2}}}{\sigma_p^2}+1\right)\frac{\tau_p^2}{\sigma_p^2}} \leq \frac{\sigma_{\mathrm{c}}^2+\frac{\tau_c^2}{n}}{\sigma_p^2}+1$$

$$\iff \sigma_{\mathrm{c}}^2+\tau_{\mathrm{c}}^2\frac{1+\frac{\tau_p^2}{\tau_{\mathrm{c}}^2}}{n+\frac{\tau_p^2}{\tau_{\mathrm{c}}^2}}+\sigma_p^2 \leq \left(\sigma_{\mathrm{c}}^2+\frac{\tau_{\mathrm{c}}^2}{n}+\sigma_p^2\right)\left[1+\left(\frac{\sigma_{\mathrm{c}}^2+\tau_{\mathrm{c}}^2\frac{1+\frac{\tau_p^2}{\tau_{\mathrm{c}}^2}}{n+\frac{\tau_p^2}{\tau_{\mathrm{c}}^2}}}{\sigma_p^2}+1\right)\frac{\tau_p^2}{\sigma_p^2}\right]$$

$$\iff \tau_{\mathrm{c}}^2\frac{1+\frac{\tau_p^2}{\tau_{\mathrm{c}}^2}}{n+\frac{\tau_p^2}{\tau_{\mathrm{c}}^2}} \leq \frac{\tau_{\mathrm{c}}^2}{n}\left[1+\left(\frac{\sigma_{\mathrm{c}}^2+\tau_{\mathrm{c}}^2\frac{1+\frac{\tau_p^2}{\tau_{\mathrm{c}}^2}}{n+\frac{\tau_p^2}{\tau_{\mathrm{c}}^2}}}{\sigma_p^2}+1\right)\frac{\tau_p^2}{\sigma_p^2}\right]+(\sigma_{\mathrm{c}}^2+\sigma_p^2)\left(\frac{\sigma_{\mathrm{c}}^2+\tau_{\mathrm{c}}^2\frac{1+\frac{\tau_p^2}{\tau_{\mathrm{c}}^2}}{n+\frac{\tau_p^2}{\tau_{\mathrm{c}}^2}}}{\sigma_p^2}+1\right)\frac{\tau_p^2}{\sigma_p^2}$$

$$\iff \tau_{\mathrm{c}}^2\left(\frac{\tau_{\mathrm{c}}^2+\tau_p^2}{n\tau_{\mathrm{c}}^2+\tau_p^2}-\frac{\sigma_p^4+\tau_p^2\left(\sigma_{\mathrm{c}}^2+\tau_{\mathrm{c}}^2\frac{\tau_{\mathrm{c}}^2+\tau_p^2}{n\tau_{\mathrm{c}}^2+\tau_p^2}+\sigma_p^2\right)}{n\sigma_p^4}\right) \leq (\sigma_{\mathrm{c}}^2+\sigma_p^2)\frac{\left(\sigma_{\mathrm{c}}^2+\tau_{\mathrm{c}}^2\frac{\tau_{\mathrm{c}}^2+\tau_p^2}{n\tau_{\mathrm{c}}^2+\tau_p^2}+\sigma_p^2\right)\tau_p^2}{\sigma_p^4}$$

$$\iff \tau_{\mathrm{c}}^2\left(\frac{\tau_{\mathrm{c}}^2+\tau_p^2}{n\tau_{\mathrm{c}}^2+\tau_p^2}-\frac{1}{n}\right) \leq \left(\sigma_{\mathrm{c}}^2+\frac{\tau_{\mathrm{c}}^2}{n}+\sigma_p^2\right)\frac{\left(\sigma_{\mathrm{c}}^2+\tau_{\mathrm{c}}^2\frac{\tau_{\mathrm{c}}^2+\tau_p^2}{n\tau_{\mathrm{c}}^2+\tau_p^2}+\sigma_p^2\right)\tau_p^2}{\sigma_p^4}.$$

The last inequality follows by expanding the two sides.

For part 2, we directly verify that

$$\frac{\partial(\tilde{\kappa}_c^{\text{het}})^2}{\partial\tau_p^2} \propto -\tau_{\mathrm{c}}^4[n^2(\sigma_{\mathrm{c}}^2+\sigma_p^2)^2+n(2(\sigma_{\mathrm{c}}^2+\sigma_p^2)\tau_{\mathrm{c}}^2-\sigma_p^4)+\tau_{\mathrm{c}}^4+\sigma_p^4]$$
$$-2\tau_{\mathrm{c}}^2(\sigma_{\mathrm{c}}^2+\sigma_p^2+\tau_{\mathrm{c}}^2)(n(\sigma_{\mathrm{c}}^2+\sigma_p^2)+\tau_{\mathrm{c}}^2)\tau_p^2-(\sigma_{\mathrm{c}}^2+\sigma_p^2+\tau_{\mathrm{c}}^2)^2\tau_p^4<0.$$

Their limiting behavior as τ_p approaches 0 follows from the definitions of the learning frictions. When τ_p approaches infinity,

$$\lim_{\tau_p\to\infty}\kappa_c^{\text{het}}=\frac{\sigma_{\mathrm{c}}^2+\tau_{\mathrm{c}}^2}{\sigma_p^2}+1.$$

In particular, it is bounded both from above and away from 0. Then as $\tau_p \to \infty$, the normalization factor as the denominator $\sqrt{1 + (\kappa_c^{\text{het}})^2 \frac{\tau_p^2}{\sigma_p^2}}$ of $\tilde{\kappa}_c^{\text{het}}$ approaches infinity. Therefore,

$$\lim_{\tau_p \to \infty} \tilde{\kappa}_c^{\text{het}} = 0.$$

□

Proof of Proposition 6. We proceed with the proof in two steps. First, we analyze the behavior of $\tilde{\kappa}_{\text{inf}}^{\text{het}}$ for the two influencers, testing whether and when $\tilde{\kappa}_{\text{inf},1}^{\text{het}} - \tilde{\kappa}_{\text{inf},2}^{\text{het}}$ changes sign as τ_p increases. Second, we use Corollary 2.3—which tells us that the monotonicity of profit-per-consumer with respect to $\tilde{\kappa}_{\text{inf}}^{\text{het}}$ (recall that $\tilde{\kappa}_{\text{inf}}^{\text{het}}$ plays the exact same role as κ_c in the homogeneous model on profitability)—to understand how brand value interacts with consumer heterogeneity to affect profitability.

To determine the sign of $\tilde{\kappa}_{\text{inf},1}^{\text{het}} - \tilde{\kappa}_{\text{inf},2}^{\text{het}}$, solving for $\tilde{\kappa}_{\text{inf},1}^{\text{het}} = \tilde{\kappa}_{\text{inf},2}^{\text{het}}$ yields the following quadratic expression denoted by $Q(\tau_p)$:

$$Q(\tau_p) = A\tau_p^2 + B\tau_p + C, \tag{39}$$

where

$$\begin{aligned}
A &= \sigma_{\text{inf},1}^2 - \sigma_{\text{inf},2}^2, \\
B &= N_{\text{inf},1}(\sigma_{\text{inf},1}^2 - \sigma_{\text{inf},2}^2 - 1) + N_{\text{inf},2}(\sigma_{\text{inf},1}^2 + 1 - \sigma_{\text{inf},2}^2), \\
C &= N_{\text{inf},2} - N_{\text{inf},1} + N_{\text{inf},1}N_{\text{inf},2}(\sigma_{\text{inf},1}^2 - \sigma_{\text{inf},2}^2).
\end{aligned}$$

It suffices to determine the sign of $Q(\tau_p)$ as $\tilde{\kappa}_{\text{inf},1}^{\text{het}} - \tilde{\kappa}_{\text{inf},2}^{\text{het}} > 0$ if and only if $Q(\tau_p) > 0$. Under the assumption that $\sigma_{\text{inf},1} < \sigma_{\text{inf},2}$ and since $\tau_p > 0$, we can focus on analyzing the following three cases:

(1) $Q(\tau_p)$ has no positive root—either it has no roots at all, or it has no positive root(s). In this case, $\tilde{\kappa}_{\text{inf},1}^{\text{het}} - \tilde{\kappa}_{\text{inf},2}^{\text{het}} < 0$ for any τ_p.

(2) $Q(\tau_p)$ has two roots of different signs. In this case, $\tilde{\kappa}_{\text{inf},1}^{\text{het}} - \tilde{\kappa}_{\text{inf},2}^{\text{het}} > 0$ when τ_p is small, and $\tilde{\kappa}_{\text{inf},1}^{\text{het}} - \tilde{\kappa}_{\text{inf},2}^{\text{het}} < 0$ when τ_p is large.

(3) $Q(\tau_p)$ has two positive roots (including the case of a double root). In this case, $\tilde{\kappa}_{\text{inf},1}^{\text{het}} - \tilde{\kappa}_{\text{inf},2}^{\text{het}} < 0$

when τ_p is either sufficiently small or sufficiently large. Otherwise, $\tilde{\kappa}^{\text{het}}_{\text{inf},1} - \tilde{\kappa}^{\text{het}}_{\text{inf},2} > 0$.

We can determine the sets of parameter values $(N_{\text{inf},1}, N_{\text{inf},2}, \sigma_{\text{inf},1}, \sigma_{\text{inf},2})$ that satisfy the conditions of each case. For example, case (1) requires either $\Delta(Q(\tau_p)) < 0$, or $B < 0$ and $C < 0$. Case (2) requires to $\Delta(Q(\tau_p)) > 0$ and $C > 0$. Case (3) requires $\Delta(Q(\tau_p)) \geq 0$, $B > 0$, and $C < 0$. Further imposing the requirements that $N_{\text{inf},2} \gg 0$ and $N_{\text{inf},1} = 2$ will not eliminate any of the three cases above, but will eliminate some solutions sets to every case. We obtain the following sets of solutions. (1') is the solution set to case (1), etc.

(1') $0 < \sigma^2_{\text{inf},1} \leq \sigma^2_{\text{inf},2} - 1$, and $N_{\text{inf},2} > 2$.

(2') $\max\{\sigma^2_{\text{inf},2} - \frac{1}{2}, 0\} < \sigma^2_{\text{inf},1} < \sigma^2_{\text{inf},2}$, and $N_{\text{inf},2} > \frac{2}{1+2\sigma^2_{\text{inf},1}-\sigma^2_{\text{inf},2}}$.

(3') $\max\{\sigma^2_{\text{inf},2} - 1, 0\} < \sigma^2_{\text{inf},1} \leq \sigma^2_{\text{inf},2} - \frac{1}{2}$, and $N_{\text{inf},2} > \frac{2+2(\sigma^2_{\text{inf},2}-\sigma^2_{\text{inf},1})^2}{(1+\sigma^2_{\text{inf},1}-\sigma^2_{\text{inf},2})^2}$.

If we impose that $N_{\text{inf},2} > \max\left\{2, \frac{2}{1+2\sigma^2_{\text{inf},1}-\sigma^2_{\text{inf},2}}, \frac{2+2(\sigma^2_{\text{inf},2}-\sigma^2_{\text{inf},1})^2}{(1+\sigma^2_{\text{inf},1}-\sigma^2_{\text{inf},2})^2}\right\}$, then we reduce these solution sets to conditions only on $\sigma_{\text{inf},1}$ and $\sigma_{\text{inf},2}$. This completes the first step of the proof. Using Corollary 2.3, which in the context of $\tilde{\kappa}^{\text{het}}_{\text{inf}}$, says that the equilibrium profit-per-consumer π^*_{inf} is increasing in $\tilde{\kappa}^{\text{het}}_{\text{inf}}$ if and only if $\tilde{\kappa}^{\text{het}}_{\text{inf}} > \sqrt{\frac{\pi}{2}}\frac{\sigma_p}{\mu_p}$, we observe that π^*_{inf} will generally exhibit a U-shape with respect to τ_p because $\tilde{\kappa}^{\text{het}}_{\text{inf}}$ is decreasing in τ_p.

We now analyze the three cases from the first step in two scenarios: (a) $\sqrt{\frac{\pi}{2}}\frac{\sigma_p}{\mu_p}$ is greater than the highest possible value of $\tilde{\kappa}^{\text{het}}_{\text{inf}}$, which occurs for very unknown brands, and (b) $\sqrt{\frac{\pi}{2}}\frac{\sigma_p}{\mu_p}$ is attainable by $\tilde{\kappa}^{\text{het}}_{\text{inf}}$ for some τ_p, which occurs for well established brands. By Lemma 2, the possible range of $\tilde{\kappa}^{\text{het}}_{\text{inf}}$ is $(0, \kappa_{\text{inf}}]$. Hence, we are in scenario (a) if and only if $\sqrt{\frac{\pi}{2}}\frac{\sigma_p}{\mu_p} > \kappa_{\text{inf}}$.

Case (1) $\tilde{\kappa}^{\text{het}}_{\text{inf},1} < \tilde{\kappa}^{\text{het}}_{\text{inf},2}$ for any τ_p: In scenario (a), π^*_{inf} is increasing in τ_p, and $\pi^*_{\text{inf},1} > \pi^*_{\text{inf},2}$ for any τ_p. In scenario (b) $\pi^*_{\text{inf},1} < \pi^*_{\text{inf},2}$ when τ_p is small and $\pi^*_{\text{inf},1} > \pi^*_{\text{inf},2}$ when τ_p is large.

Case (2) $\tilde{\kappa}^{\text{het}}_{\text{inf},1} - \tilde{\kappa}^{\text{het}}_{\text{inf},2} > 0$ when τ_p is small, and $\tilde{\kappa}^{\text{het}}_{\text{inf},1} - \tilde{\kappa}^{\text{het}}_{\text{inf},2} < 0$ when τ_p is large: In scenario (a), $\pi^*_{\text{inf},1} < \pi^*_{\text{inf},2}$ when τ_p is small and $\pi^*_{\text{inf},1} > \pi^*_{\text{inf},2}$ when τ_p is large. In scenario (b) $\pi^*_{\text{inf},1} > \pi^*_{\text{inf},2}$ when either τ_p is very low or very high.

Case (3) $\tilde{\kappa}^{\text{het}}_{\text{inf},1} - \tilde{\kappa}^{\text{het}}_{\text{inf},2} < 0$ when τ_p is either sufficiently small or sufficiently large: In scenario (a), $\pi^*_{\text{inf},1} > \pi^*_{\text{inf},2}$ when either τ_p is very low or very high. In scenario (b), $\pi^*_{\text{inf},1} < \pi^*_{\text{inf},2}$ when τ_p is very low, and $\pi^*_{\text{inf},1} > \pi^*_{\text{inf},2}$ when τ_p is very high.

Combining the different scenarios, creative mega-influencer are better under low consumer heterogeneity and consistent micro-influencer are better under high consumer heterogeneity for unknown brands when $\max\{\sigma^2_{\text{inf},2} - \frac{1}{2}, 0\} < \sigma^2_{\text{inf},1} < \sigma^2_{\text{inf},2}$; and for established brand, when $0 < \sigma^2_{\text{inf},1} \leq \sigma^2_{\text{inf},2} - 1$ or $\max\{\sigma^2_{\text{inf},2} - 1, 0\} < \sigma^2_{\text{inf},1} \leq \sigma^2_{\text{inf},2} - \frac{1}{2}$. □