Shill Marketing

Nathan Berg
University of Otago

Jeong-Yoo Kim*
Kyung Hee University

May 4, 2023

Abstract

Shill marketing occurs whenever sellers pay influencers to transmit positive messages about the seller's product and this arrangement remains hidden from consumers. We present two-stage signaling models with low- or high-quality sellers, influencers, and second-period consumers who condition purchase decisions on influencer messages. In this paper, we raise the question of why sellers of low-quality products are more willing to engage in shill advertising, despite the Nelson Effect. We argue that high-quality sellers engage in less shill advertising in order to maintain the high reliability of word-of-mouth communication. The different incentives are not due to the Nelson Effect but due to what we call credibility effect.

JEL Classification Code: D82, L15

Keywords: Advertising, Experience Good, Influencer, Followers, Signal, Shill Marketing, Word-of-Mouth (WoM) Communication

*This research was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2022S1A5A2A0304932311). We are grateful to Jay Pil Choi and seminar participants at KAIST and University at Albany-SUNY for helpful comments. Corresponding author: Department of Economics, Kyung Hee University, 1 Hoegidong, Dongdaemunku, Seoul 130-701, Korea, (Tel) +822-961-0986, (Email) jyookim@khu.ac.kr

1 Introduction

Shills can be defined as those who are paid to promote a product (brand, organization or idea) while portraying themselves as independent from the seller of that product or failing to disclose otherwise hidden conflicts of interest linking the shill to the seller. "Shills promote companies, products, public figures and viewpoints for profit, while pretending to have no motivation for doing so other than personal belief" (Wigmore, 2013).[1] Also known as incentivized or paid influencers, shills pretend not to work for the organization that pays them and is ultimately the responsible party for undertaking *shill marketing*, the essence of which is that consumers do not know they are being marketed to. Despite being illegal (in some instances) or possibly constituting fraud (when financial losses are incurred by customers under the influence of influencer messaging), shill marketing can be impossible to detect (Balough, 2012) and takes place frequently, with top influencers receiving large payments (Shifferaw, 2019) and influencer-promoted products comprising substantial revenue flows across many industries.[2]

Among the earliest cases of shill marketing in the US courts was the plastic surgery company, Lifestyle Lift, which instructed its employees to post positive reviews that were fake. A civil court found in favor of the Office of the New York State Attorney General (2009), ordering the company to pay a fine and stop its employees across 40 US locations from illegally posing as consumers.

Kim Kardashian is an interesting example because she exhibits well-documented cases of shilling as well as non-shilling influencer messaging (and in-between cases, too). Kardashian promoted garments, skin products, pharmaceuticals and a cryptocurrency without disclosing that her messages were paid advertisements (examples of her shilling). Responding to regulatory pressure, she later added notifications informing her followers, by including "#ad" or

[1] Shills may also send messages to discredit critics of the seller's product in which they have a vested interest or the seller's competitors. For simplicity, these more complex forms of shilling that involve additional game players are not directly addressed in our model, although the intuition it formalizes using the simplest possible binary message space can be interpreted more broadly to include any ordinal ranking of the influencer's message space that a seller or other party is willing to pay to manipulate.

[2] There are substantial academic and professional literatures concerning shill bidding in auction markets (Grether et al., 2015), which is not our primary focus, although shills who send positive messages about product quality (which *is* our focus) would appear to be intuitively analogous to some of word-of-mouth informational effects (but not direct the price effects) of shill bidding.

"#promo" on her online messages, as required by the Federal Trade Commission (FTC).[3] The Securities and Exchange Commission (SEC) documented one of Kardashian's most notorious instances of shilling, where she was fined $1.4mil for failing to disclose that she was paid $250,000 to promote a cryptocurrency sold by EthereumMax (Laskowski, 2022).

Similar to the FTC's regulation in the U.S., the Korean Fair Trade Committee's (KFTC) guidelines (since September 1, 2020) stipulate that if someone is paid to post information about a firm's product, then explicit notification is required ("#advertisement"). Such notifications (common across many countries' jurisdictions) can be expected to reduce a seller's chances of influencing other consumers' purchase decisions, however. Laws requiring such notification negatively affect the seller's expected return on ad expenditure, which leads to an obvious motive for sellers to employ influencers while keeping their contracted relationship hidden.[4]

Sellers increasingly involve consumers in the process of developing advertising and other marketing actions.[5] One of the main reasons why sellers utilize consumers in marketing is because many consumers dislike ads and expend money and effort to avoid them. Consumers are likely to subjectively discount or ignore messages transparently notified as paid

[3]Evans (2016) notes product-promoting photos posted by Kardashian: "In none of those snaps did any of them admit to having been paid to promote the product. But we've since learned that they were, because posts have been recently updated with captions that include the hashtag #ad." Roettgers (2016) cites an investigation and complaint by the non-profit organization, Truth In Advertising, followed by regulatory pressure from the FTC and FDA, against Kardashian's product endorsements that were undisclosed at first (shilling), followed up by disclosures of paid influencing on those same online posts (no longer shilling). He writes: "FTC regulations require journalists, celebrities and other media personalities to disclose their relationship with companies if they are reimbursed for a review or endorsement. These regulations were initially put in place to make sure TV show hosts wouldn't advertise products without proper disclosures on air, but have since been adapted to the social media age." Notices that are acceptable to the FTC include #Spon (sponsored), #Ad or the seven-character "Paid-ad" or nine-character "Sponsored" or "Promotion."

[4]According to an investigation by the Korea Consumer Agency (Chang, 2022), only 174 out of 583 advertisements in their sample of incentivized influencers had posted the required notifications, meaning that 70% of ads were shills. (available at https://www.dailian.co.kr/news/view/1079170) A well-known example in Korea (following controversial debates over the regulation of sponsored advertisements that received prominent media coverage) is stylist/YouTuber Hyeyeon Han's channel, *I Bought It With My Own Money*, famous for introducing shoes and cosmetics to her followers. She later admitted that she was paid $20,000 to $30,000 per show, despite claiming otherwise in the show's title.

[5]See Thompson and Malaviya (2013).

advertisement. For these reasons, marketers turned to the goal of involving unpaid real-world consumers to generate more credible marketing messages (Thompson, 2013) and the question of whether sellers and the marketers they employ could orchestrate word-of-mouth (WoM) communication consisting of (at least a strictly positive proportion of) voluntary messages about the products people consume. Advice from other consumers is often considered to be among the most powerful influences on purchasing decisions (Hung et al., 2007). This is the main reason why viral marketing and WoM communication recently receive considerable attention in marketing. However, **consumers have no good reason to engage in WoM communication nor to be honest if they do, because they are just indifferent. This implies that WoM communication does not necessarily need to convey truthful information about product quality,** although most of the literature implicitly or explicitly assumes that consumers can learn the **true** quality of the product by WoM communication.[6]

The more fundamental reason for a seller to need the consumer/influencer product evaluations—whether paid (shill) or unpaid (non-shill)—is that consumers are not perfectly informed about product quality before purchasing, referred to in the literature as *experience goods* (Nelson, 1970).[7] As such, shill marketing can be used only for selling experience goods. It is well known that there are two distinct effects in pricing an experience good; Nelson Effect and Schmalansee Effect, depending on whether the seller engages in strategic under-pricing (or too much dissipative advertising) or strategic over-pricing (or too little advertising) for a signaling purpose. Then, it is natural to ask whether the insight from the economics of advertising literature, which suggests that ad spend can serve as a genuine signal of otherwise unobservable quality, extends to shill marketing. The Nelson Effect occurs when high-quality sellers have a stronger incentive to attract consumers by a low introductory price or dissipative advertising expenditures, because a high-quality product can generate more repeat purchases. On the other hand, the Schmalansee Effect occurs when low-quality sellers have a stronger incentive to set a lower price, because low-quality

[6]See, for example, Satterthwaite (1979), Milgrom and Roberts (1986), Ellison and Fudenberg (1995), and p.118 of Tirole (1988).

[7]Nelson (1970) defined , experience goods as those whose true quality is learned by consumers only after purchasing, while *search goods* are those whose quality can be ascertained before purchasing. Later, Darby and Karni (1973) extended Nelson's taxonomy by introducing *credence goods*, defined as those whose quality cannot be observed by consumers even after purchasing.

products generate greater profits **(per unit sale)** because of lower production costs.

In this paper, we raise the following issues. According to the Nelson Effect, a high-quality seller seems to have a stronger incentive to attract more influencers by more generous shill payments because it could generate more second-period consumers through WoM communication, but anecdotal evidence indicates that shill advertising is more frequently used by sellers of low-quality products or products whose quality is uncertified.[8] Then, why are sellers of low-quality products more willing to engage in shill advertising, contrary to the Nelson effect and the theoretical result by Milgrom and Roberts (1986) that supports Nelson's insight? Secondly, it is clear that a high rate of shill advertising degrades the credibility of WoM communication. Then, why do sellers engage in an enormous amount of shill advertising despite the negative credibility effect? In this paper, we argue that sellers of high-quality products engage in less shill advertising in order to maintain the high reliability of WoM communication, whereas sellers of low-quality products engage in more shill advertising because reliability of WoM communication is not so valuable to them as to those selling high-quality products. This difference in incentives is not due to the Nelson Effect but due to what we call credibility effect. Although consumers purchase repeatedly in the sense that there are consumers in both periods, the Nelson Effect is not the main driving force in this model, because WoM communication on which it relies heavily is not always credible. It is not due to the Schmalensee Effect, either, insofar as costs of production are identical across high- and low-quality-type sellers.

In our models, the seller secretly makes a monetary offer to an influencer for their service of providing a shill advertisement in the first period. The influencer may or may not accept the seller's offer, however. In the second period, consumers are therefore unsure about whether any positive messages they encounter are shill ads or genuine messages from unpaid influencers. After observing the influencer's product evaluation[9] in the second period,

[8] According to KFTC (2023), products of health supplements have the highest proportion of 25.5% among 21,037 violations that they detected from April 2022 to December 2022.

[9]Insofar as uninformed future consumers believe that there is information content in seeing, hearing or reading about an influencer having used the seller's product, then our references to "product evaluations" by influencers can be substituted with "product placement" of a very general kind, which includes both online and traditional marketing channels as well as government "white papers" and research reports (even peer-reviewed literature) that are presented as independent and impartial but with less than fully disclosed conflicts of interest.

consumers update their beliefs about the truthfulness of the influencer's message and the quality of the product. Based on these updated beliefs, consumers finally decide whether to purchase.

In our baseline model (Section 3), the seller's offer (i.e., advertisement fee) is not observable to second-period consumers but is, of course, observable to the first-period influencer who base their decision of whether to shill, in part, on the seller's offer. Thus, ours is a signaling game in which signaling can occur in two stages sequentially. In the first signaling stage, the seller, who is privately informed about the quality of the product, moves first by choosing the shill offer. Next, the influencer, who is uninformed about product quality, responds by choosing whether or not to accept the shill contract, and then choose their messages based on their experiences or shill contracts. In this second signaling stage, the first-period influencer sends a message to his or her followers, after experiencing the product and obtaining a noisy signal of the product quality. The message may be authentic or not, depending on whether he or she made a shill contract in the first period. Finally, second-period consumers, who are uninformed about the seller's quality type, make purchasing decisions with the possibility of conditioning on updated beliefs about product quality and the accuracy (or, equivalently, honesty) of the influencer's message. We can then examine (Sections 3 and 4) whether high or low offers (from sellers to influencers for their supply of shilling services) signal high product quality.

We consider two models to address this issue. In the first model the seller's choice space is unlimited so that their shill offers reach all potential influencers, who, in turn, base their **decisions** of whether to accept based solely on a pecuniary motive. In the second model the seller's information about available influencers and the resulting choice are limited so that not all potential influencers get shill offers from the seller.

In our first model, we show that an influencer who accepts the shill contract must purchase the product, while any influencers rejecting the seller's offer do not buy the product, since any influencer who wants to buy the product will strictly prefer to accept the shill contract. This is simply because they receive the same consumer surplus as non-shill influencers do from consuming the product, in addition to the seller's monetary shill offer (or, equivalently, advertising fee). We then show that sellers can signal their quality type by making different shill offers, but only as a knife-edge case in our first model. This is counterintuitive. Since good quality of a product is more likely to be realized in the second period, it seems that

the Nelson Effect appears. However, the result in this paper shows that it is not possible in equilibrium, because of Lemma 1 **implying that all the WoM communications are shill advertising messages**. **Although the Nelson effect depends on the truthfulness of WoM communication, all the WoM communications come from dishonest influencers** who made shill contracts, so they never convey truthful messages about the product quality. Only when messages convey some truthful information of the noisy signals that the influencer receives, does the second-period profit depend on the seller's actual type. Insofar as all the messages are fake, profit does not depend on the seller's true type, but on consumers' beliefs about the seller's quality type. The cost of signaling is therefore the same, regardless of the seller's true quality type.

In the second model, the seller chooses a shill offer (as payment for influencers to provide shilling services, just as in the previous model), but due to search costs incurred finding and/or contracting with them, some influencers may purchase and evaluate the product honestly without a shill contract. Since messages of some influencers can be truthful in this second model (Section 4), the seller's profit depends on the product's true quality type as well as consumer beliefs about the product's true quality. Due to the dependence of profits on true product quality, the seller's choice of shill offers can now signal product quality. We show (Section 4) that high-quality sellers make lower shill offers, which translates to higher effective product pricing offered to first-period influencers that, in turn, increases second-period consumer demand (via signaling), which resembles **a Schmalansee effect insofar as high type is signaled by high price.** Our analysis demonstrates that **it is not due to Schmalansee effect nor Nelson effect, however.** Instead, we show the reason why high-quality sellers can signal their quality by offering a smaller payment to shills is what we refer to as the *credibility effect*, which means that high-quality sellers prefer to reduce sales to shills, thereby increasing the proportion of non-shill influencers sending messages about the seller's product (i.e., in equilibrium, increasing influencer credibility as measured by the signal-to-noise ratio of messages sent by influencers). Improved credibility among a smaller number of influencers, a greater proportion of which are honest (i.e., non-shill influencers) more than offsets their reduced number of messages. Why do sellers engage in shill marketing even if shill marketing degrades the credibility of WoM communication? Low quality sellers are less concerned about credibility, i.e., it is more important to attract more first-period consumers by giving up second-period consumers. So, they choose more shill marketing

by high advertising payments. On the other hand, high quality sellers are more concerned about credibility, i.e., they would rather give up attracting more first-period consumers for more second-period consumers by credible WoM communication. So, they choose less shill marketing by low advertising payments. Therefore, low shill advertising payments signal high quality.

In the case of an experience good **or a credence good**, consumers must make purchasing decisions without certainty about product quality, because quality is not directly observable to them before purchasing. Such consumers therefore have a clear motive to seek information that might help identify goods with quality characteristics they prefer, and sellers of experience goods also have a clear motive to find ways to signal quality. Price is one such signaling device (Wolinsky, 1983; Bagwell and Riordan, 1991; Judd and Riordan, 1994; Daughety and Reinganum, 1995). Strategic advertising used by sellers of experience goods is another (Nelson, 1974; Schmalensee, 1978; Kihlstrom and Riordan, 1984). Milgrom and Roberts (1986) and Hertzendorf (1993) considered price and advertising as a jointly chosen pair of signaling devices, and Grossman (1980) analyzed warranties as yet another means of signaling quality. Berg, Kim and Seon (2022) identified conditions under which the seller's refund policy could be used as a signaling device for sellers of credence goods.

This literature alerts us to two conflicting effects: the Schmalensee and Nelson effects we described above. In Milgrom and Roberts (1986) that allow a difference in the production costs, both effects appear. If the production costs are identical, a seller signals his high quality by low price or too much dissipative advertising. This is contrasted with our result that a high quality seller spends a low amount in advertising. This difference comes from the feature that in Milgrom and Roberts (1986) who implicitly assume that WoM communication is credible, there is no credibility effect whereby a high quality seller spends little in advertising to maintain a high degree of credibility of WoM communication. It is only the high type that can recover the first-period loss due to such a wasteful expenditure by a high second-period profit through credible WoM communication in their model.

Also, the information transmission mechanism in our model of influencer marketing appears to be closer to the Nelson effect, in the sense that high payments to shills lower the effective price required to induce greater numbers of first-period influencers to purchase—who, one may conjecture, generate increased demand among second-period consumers by WoM communication (product evaluation) for high-quality products. This effect does not

occur in our model, **because WoM communication is not always credible.**

To the best of our knowledge, there are no previous studies of shill marketing, but there is a literature on consumer reviews. Chevalier and Mayzlin (2006) found that online book reviews significantly boost sales. Also, both volume and valence (which can be interpreted as a proxy for quality) of consumer reviews at Movies.Yahoo.com were found to significantly increase future box-office revenue (Liu, 2006; Chintagunta et al., 2010).

In their formal model of consumer reviews, Chen and Xie (2008) argued that online consumer reviews help consumers identify products that best match their preferences and, thereby, provide a valuable coordination service by facilitating horizontal matches between sellers and consumers. In contrast to Chen and Xie (2008), our model is vertical, because product evaluations in our model signal product quality to consumers whose preferences are assumed to be uniform (high- preferred to low-quality). They also distinguish online consumer reviews from offline WoM communication, arguing that WoM communication is limited to local social networks. Our model permits a broader range of interpretation that extends well beyond online consumer reviews to any contexts in which sellers have an incentive to hide incentive payments to shills they employ to generate favorable messages (online or offline). Our paper complements the marketing literature on WoM communication and consumer reviews by endogenizing WoM communication (which includes consumer decisions about posting online reviews) as a function of the incentive scheme offered by sellers.

In the context of crowdfunding, Chen et al. (2019) showed that platforms whose consumer-review systems provide more information earn more profit for the crowdfunding platform. In their model, however, none of the crowdfunding platform's payment decisions, signaling effects or consumer decisions about posting reviews are analyzed strategically. Liu (2017) also considered a two-period signaling model similar to ours with advance selling in the first period followed by spot selling in the second. In her model, however, neither the seller's decision about how much to offer as payment nor the consumer/influencer's decision about posting reviews was endogenized as strategic variables (although the effect of negative reviews was discussed). Instead, Liu's model focused on the seller's optimal choice of a two-dimensional signal consisting of price and the credibility of the e-commerce platform.

Li, Tadelis, and Zhou (2020) argued that rewarding buyers for product ratings, referred to as Rebate-for-Feedback (RFF), could signal sellers' quality, although no formal signaling model was provided. The separating equilibria we identify in this paper contradicts their

claim that "the monetary value of the RFF is not what supports the separation of good versus bad types, but instead it is the quality of feedback that a buyer will leave." As shown in our model, the monetary value of the RFF is precisely what enables first-period consumer/influencers to distinguish good from bad sellers, while the evaluations that first-period consumer/influencers transmit is what enables second-period consumers to distinguish good- from bad-types sellers probabilistically.

The paper we found most closely related to ours was Kim and Xu (2022), which examined (i) the seller's decision about offering refunds for consumers who take the trouble to post online product reviews and (ii) the consumer's decisions about purchasing and reviewing. In their model, lenient cashback offers for consumer reviews can signal high quality, because high-quality sellers who expect a larger proportion of reviews to be positive can afford to pay more per review. They also considered the possibility that sellers manipulate reviews by paying only for positive reviews, arguing that such conditional offers are not possible in equilibrium, because other consumers understand that the messages they generate would be nothing but babbling or cheap talk.[10] If a seller's conditional cashback strategy is known to all consumers, then there is no non-babbling equilibrium in Kim and Xu's (2022) model. Moreover, consumer/influencers who accept publicly posted conditional cashback offers (only for positive reviews) do not fit the definition of shilling, because their incentivized relationship with the seller is not hidden. We extend Kim and Xu's (2022) model to the case in which the contractual arrangement between sellers and first-period influencers is hidden (i.e., unobservable to second-period consumers).

There are also some studies on fake product reviews and the credibility of product reviews in general. Anonymous online reviews are essentially cheap talk. Therefore, such reviews may not be credible, and there is substantial evidence that fake reviews are widespread (Xu et al., 2015; He et al., 2022). Luca and Zervas (2016) found that roughly 16% of restaurant reviews on Yelp were fake, and these reviews tended to be more extremely favorable or unfavorable (higher "valence") than other reviews, also trending higher (as a proportion of all reviews in a given market) over time. They also reported that restaurants were more likely to generate

[10]Review messages are cheap talk because they are unverifiable and costless. Cheap-talk games always have a babbling equilibrium in which the sender babbles (by sending random messages from which no payoff-relevant information can be learned) and the receiver ignores those messages regardless of their content (Crawford and Sobel, 1982).

fraudulent reviews when their reputations were weak (i.e., after receiving bad reviews, or at times when the restaurant had few reviews overall). Schuckert et al. (2016) suggested a method for ascertaining whether online reviews posted on TripAdvisor were suspicious (i.e., likely to be fraudulent). Relatedly, Jin et al. (2022) considered the so-called *brushing strategy* by online merchants who place fake orders for their own products on e-commerce platforms to boost their rankings in search results. This literature shares with our paper the common insight that sellers manipulate evaluations by offering monetary rewards to induce more positive evaluations, thereby distorting the field of available information available to uninformed consumers regarding product quality.

The article is organized as follows. In Section 2, we introduce the basic model. In Section 3, we analyze equilibria in the basic model. In Section 4, we extend the basic model to allow for the simultaneous presence of some influencers who are not offered a shill contract, whose proportion (unpaid influencers who voluntarily send messages about the seller's product without ever being offered a shill contract) drives our main result. In Section 5, we briefly discuss the possibility that a seller can use both price and secret advertising as signals. Concluding remarks follow in Section 6. Proofs are provided in the Appendix.

2 Basic Model

We consider the following two-period model. There is a firm or seller; a representative first-period consumer (hereafter, "he"); and second-period consumers whose numbers or market size is represented by $\gamma > 0$. The first-period consumer's role is that of the (online) influencer, and second-period consumers are those who follow, possibly conditioning their purchasing decision on having observed the first-period influencer's message. If we interpret the representative influencer as a continuum of influencers with size one, the corresponding interpretation is that each influencer has followers of size γ.

The firm sells an experience good whose quality is unknown to consumers. For simplicity, product quality is assumed to take on only two types, H or L ($H > L > 0$). The quality type, which is private information of the firm, is denoted as $q \in \{H, L\}$. The prior probability that $q = H$ is denoted $\lambda \in (0, 1)$, which is common knowledge. Firms of either type (q) face identical marginal costs of production, which is assumed to be constant and equal to zero.

Consumers can buy at most one unit of the product. If a consumer purchases the good

of quality q, he receives a noisy signal s (after consuming the product), whose probability distribution depends on q as follows. If $q = H$, then the signal received is:

$$s = \begin{cases} H & \text{with probability } \alpha \\ L & \text{with probability } 1-\alpha; \end{cases}$$

and if $q = L$, then the signal received is:

$$s = \begin{cases} L & \text{with probability } \alpha \\ H & \text{with probability } 1-\alpha, \end{cases}$$

where $1 > \alpha > \frac{1}{2}$. The accuracy of the signal the consumer receives is increasing in α and approaches perfect accuracy as $\alpha \to 1$.

The firm can make a confidential advertisement contract with the influencer in which the influencer agrees to evaluate the product favorably in return for a monetary reward $A \geq 0$. In other words, the firm is buying the favorable evaluation from the influencer at the price of A. After observing the offer A, the first-period consumer updates his posterior belief that the product is high-quality, denoted $\hat{\lambda}_1(A)$, and, based on that belief, decides whether to buy the product.

Although A is observable to the first-period consumer, it is not observable to second-period consumers because the contract is confidential. Two possibilities follow from this setup depending on whether the influencer accepts or declines the firm's offer: the influencer will become a shill or a non-shill. These non-shill influencers who evaluate the product non-strategically based on their own experience are referred to as the *honest* type of influencers.

The utility function of the first-period consumer (influencer) who buys the product is given by $U(p, q) = \theta q - p$ if he does not accept the contract (i.e., conditional on the influencer being honest) and $U(p, q) = \theta q - p + A$ if he accepts the contract (conditional on the influencer being a shill), where θ is the measure of the consumer's sensitivity to product quality (i.e., how much a consumer cares about quality), assumed to be uniformly distributed on the unit interval $[0, 1]$, and $p > 0$ is the product price.[11] We assume that θ is private information of the consumer. Thus, the firm cannot price discriminate conditional on θ. The utility of consumers (of any type in either period) who choose not to buy the product is normalized

[11] Shill influencers are often offered freebie promotional products without need to buy them. We can regard them as included in the amount of A.

to zero. We assume that there is no penalty for shill advertising, that is, shill advertising is legal.

After consuming the good, the influencer transmits his message (e.g., uploads a photo on Instagram or a video on Youtube—or is interviewed as an impartial expert in the traditional news media or in any type of document, public-facing or otherwise, that might influence the purchase decisions of second-period consumers) about his experience of the product by choosing a message m from two possible messages about product quality: "high-quality (H)" or "low-quality (L)". This message sent by the first-period influencer may or may not be used by second-period consumers, who are also *a priori* uninformed about the quality of the product, to condition second-period consumers' binary purchase decisions.

A non-shill influencer (i.e., who did not accept the seller's shill contract offer) is assumed to send a truthful message (i.e., perfectly matching the noisy signal about product quality received by the non-shill influencer).[12] In contrast, a shill influencer (who accepted the seller's shill contract offer) always sends the message that the product is high-quality (message "H"), regardless of the information received from consuming the product (i.e., regardless of whether $s = H$ or L).

The second-period consumers cannot tell whether the influencer's message is truthful, because they know they that any possible shill contract between the seller and the first-period influencer is unobservable. Second-period consumers, therefore, do not know whether the influencer they are considering conditioning their purchase decision on is a shill.

After observing the message of the influencer, second-period consumers update their beliefs about: (i) the seller's type (or, equivalently, the true product-quality type) and (ii) the influencer's type (shill or non-shill—or, equivalently, whether the influencer is honest or not). Let $\hat{\lambda}_2(m)$ represent second-period consumers' posterior belief that the product is high-quality, and let $\mu(m)$ represent their posterior belief representing the probability that the influencer is an honest non-shill type. Both of these beliefs that are updated in the second period depend functionally on the observed value of the influencer's message $m \in \{H, L\}$. Based on these updated beliefs, second-period consumers decide whether to buy the product

[12]Since the message is essentially cheap talk, other less informative equilibria may also exist. However, a fully revealing equilibrium can be reasonably selected if there is a direct lying cost or the influencer must take the indirect reputation cost due to his lying into account. For more on lying costs, see Chen et al. (2008) and Kartik (2009).

or not. By assumption, second-period consumers are not offered a shill contract. To cleanly isolate the effect of product quality on the pervasiveness of shilling in this two-period market structure with three distinct roles, we assume that second-period consumers become aware of new products only through influencers. In other words, our model makes a strong assumption here—appropriate for markets in which influencers wield significant power or are employed to a significant extent as a primary pathway to market—that second-period consumers can only buy the product conditional on having observed the first-period consumer buying the product.

Interaction between the seller and consumers of each of two time-period-specific types proceeds as follows. In the first period, the seller chooses the continuously valued offer $A > 0$ and the product price p is assumed to be exogenously given. After observing the seller's offer A, the first-period influencer updates his belief about product quality to $\hat{\lambda}_1(A)$. Based on this updated belief, the influencer then decides whether to purchase the good. If he accepts the seller's offer for generating the shill contract, then the influencer must buy the product. If the shill offer is rejected, then the influencer may or may not buy the product. If the product is purchased, then the noisy signal s about q is received. Finally (still in period 1), the influencer must decide which message to send, H or L, based jointly on the acceptance or rejection of the shill contract and the observed value of the signal. In the second period, the mass of γ (quantifying the number of second-period consumers) makes a purchasing decision based on the message sent by the first-period influencer.

3 Analysis

To analyze the two-period game under incomplete information, we will employ the weak Perfect Bayesian Equilibrium (wPBE) as our main equilibrium concept and Cho and Kreps (1987)'s Intuitive Criterion as a refinement whenever necessary. The set of weak Perfect Bayesian Equilibria can be found by backward induction.

3.1 Purchasing Decisions

Second-period consumers make their purchasing decisions based on their two beliefs about product quality and the influencer's honesty, $\hat{\lambda}_2(m)$ and $\mu(m)$, respectively. Given their

updated belief about product quality $\hat{\lambda}_2(m)$, second-period consumers choose to purchase the product if and only if their net expected utility is non-negative:

$$V_2(m) \equiv E[U(q;p);\hat{\lambda}_2(m)] = \theta E(q|m) - p \geq 0, \tag{1}$$

where $E(q|m) = \hat{\lambda}_2(m)H + (1-\hat{\lambda}_2(m))L$. Thus, an individual second-period consumer purchases one unit of the product if $\theta \geq \frac{p}{E(q|m)} \equiv \hat{\theta}_2(m)$, and total demand in the second period (conditional on first-period demand) is $\gamma D^{t=2}(m)$, where $D^{t=2}(m) = 1 - \hat{\theta}_2(m)$. Because second-period transactions are possible only when the first-period consumer purchases, second-period total demand is conditional on the first-period purchase by the influencer.

The influencer's purchasing decision takes place only in the first period. There are two types of influencers: shills who have a contract with the seller (denoted C for "contract with the seller") and non-shills (denoted N for "no contract" or "non-shill"). Let U^C and U^N represent the utility functions of shills and non-shills, respectively. Let $V_1^C(A)$ and $V_1^N(A)$ represent the net utilities of these two types of influencers in the first period, which functionally depend on the seller's shill offer A. The shill influencer will accept a shill contract (which also obligates him to buy the product) if and only if the sum of the consumer surplus from the purchase and payment received are weakly positive:

$$V_1^C(A) \equiv E[U^C(q|A)] = \theta E(q|A) - p + A \geq 0, \tag{2}$$

where $E(q|A) = \hat{\lambda}_1(A)H + (1-\hat{\lambda}_1(A))L$ or, equivalently, if $\theta \geq \hat{\theta}_{1,C}(A) \equiv \frac{p-A}{E(q|A)}$.[13] A non-shill influencer will purchase the product if and only if:

$$V_1^N(A) \equiv E[U^N(q|A)] = \theta E(q|A) - p \geq 0, \tag{3}$$

or, equivalently, if $\theta \geq \hat{\theta}_{1,N}(A) \equiv \frac{p}{E(q|A)}$. It is clear that $\hat{\theta}_{1,C}(A) < \hat{\theta}_{1,N}(A)$, i.e., the dishonest influencer is more likely to buy than the honest influencer.

The two inequalities above partition the support of θ into three profiles of observed purchasing behaviour by the two influencer types: (i) if $\theta \geq \hat{\theta}_{1,N}(A)$, both the shill and non-shill influencer buy; (ii) if $\theta \leq \hat{\theta}_{1,C}(A)$, neither type buys, and (iii) if $\theta \in (\hat{\theta}_{1,C}(A), \hat{\theta}_{1,N}(A))$, only the dishonest type buys.

The following lemma implies that case (iii) cannot occur.

[13] If the amount they receive A is greater than the purchase price p or the product is given to the influencer as a freebie promotional one, it is possible that $\hat{\theta}_{1,C}(A) \leq 0$. In this case, all the influencers will make shill contracts. Considering the reality that there are quite a few influencers who are honest, we exclude this case.

Lemma 1. *For first-period influencers, the strategy of purchasing without making a shill contract is strictly dominated by the strategy of purchasing after making a shill contract.*

This lemma implies that any influencer who buys must be dishonest, and that any consumer who does not buy must be honest because he would make no shill contract. There cannot be an honest consumer who buys, because if a consumer wants to buy, it is always better to do so after making a shill contract. Any consumer with θ in the intermediate range ($\theta \in (\hat{\theta}_{1,C}(A), \hat{\theta}_{1,N}(A))$) would rather choose to be dishonest by making a shill contract than to remain honest by refusing the seller's shill offer, which leads to the following proposition.

Proposition 1. *(i) An influencer chooses to make a shill contract and buys one unit of the product if $\theta \geq \hat{\theta}_1(A)$, where $\hat{\theta}_1(A) = \hat{\theta}_{1,C}(A) = \frac{p-A}{E(q|A)}$, and (ii) chooses not to make a shill contract and buys none if $\theta < \hat{\theta}_1(A)$.*

Here, $\hat{\theta}_1(A)$ is the cutoff value of θ_1 that dichotomizes influencers' purchasing behavior. All influencers who purchase should rationally choose to be shills. Proposition 1 implies that an influencer becomes a shill and makes a first-period purchase only if θ is sufficiently high; and he chooses to remain honest as a non-shill influencer and chooses not to purchase if θ is low. The cutoff value of θ_1 that dichotomizes the shilling and purchasing behaviour of influencers (which is one in the same) is denoted by $\hat{\theta}_1(A)$ and given in the proposition. The proof follows from Lemma 1. Since the influencer with $\theta \in (\hat{\theta}_{1,C}(A), \hat{\theta}_{1,N}(A))$ strictly prefers to purchase and receive the contracted shilling payment A rather than not purchase and receive no contracted payment, it follows that $\hat{\theta}_{1,N}(A)$ is meaningless and, therefore, $\hat{\theta}_1(A) = \hat{\theta}_{1,C}(A)$.

3.2 Belief Updating

In the second period, consumers update two beliefs, $\mu(m)$ and $\hat{\lambda}_2(m)$ based on the observed message m that the first-period influencer sent. We first compute the posterior belief that the influencer is honest, $\mu(m)$.

If $m = L$, it is clear that $\mu(L) = 1$, because the dishonest influencer never uses the message $m = L$. Unfortunately, however, this ($m = L$) cannot occur in equilibrium, because an influencer who buys the product and sends a message cannot be honest. Therefore, the only possible equilibrium message is $m = H$ which is sent by a dishonest influencer. This implies that $\mu(H) = 0$.

The posterior belief that product-quality is high ($q = H$) conditional on observing the message $m = H$ can be computed as follows:

$$\hat{\lambda}_2(H) = \frac{\lambda D(H) \times 1}{\lambda D(H) \times 1 + (1-\lambda) D(L) \times 1} = \frac{\lambda}{\lambda + (1-\lambda)\frac{D(L)}{D(H)}}, \tag{4}$$

where $D(q) = D^{t=1}(q)$ can be interpreted as the probability that the influencer buys the product of quality q. We note that the denominator of the expression on the right-hand side of (4) is the probability that the message H is observed, which is the sum of two probabilities of disjoint events. The first product in this denominator is the probability that the actual quality is H (λ) and the influencer purchases the product ($D(H)$) and sends the message H (1). The second product in the denominator is the probability that the actual quality is L ($1 - \lambda$) and the influencer purchases the product ($D(L)$) and sends a message H (1). The probability of the latter joint event also appears in the numerator of (4).

The probability that the influencer sends $m = H$ is always one, because a shill influencer always chooses $m = H$ regardless of whether $s = H$ or L. If $D(H) = D(L)$, it is clear that $\hat{\lambda}_2(H) = \lambda$, because the influencer always sends $m = H$ regardless of s and the purchasing decision itself does not imply that the type who sent the message is more likely to be a high type or a low type, so the message will be never informative.

Note that receiving a message itself increases the probability that $q = H$, because a high-quality product sells more in the first period than a low-quality one in a separating equilibrium. Also, note that $\hat{\lambda}_2(H)$ does not depend on α, because there is no honest influencer in this model.

We are now in a position to analyze whether the influencer's message causes the second-period consumer's belief that the product is high-quality to rise or fall by comparing the right-hand side of Equation (4) to λ. If $D(H) > D(L)$, the updated belief that $q = H$ after observing the message H rises relative to λ. The reason for this upward belief updating regarding product quality is that a high-quality product is more likely to be purchased by the first-period influencer. Therefore, the probability of seeing the message $m = H$ is greater if the product type is in fact high ($q = H$).

Is it then guaranteed that a product of high quality sells more than a product of low quality (i.e., $D(H) > D(L)$)? The answer to this question depends on the first-period consumer's belief about the quality. In a separating equilibrium, the true quality is revealed: $q^e = q$, where q^e is the perceived type of the seller. Thus, a high-quality product sells more

in the first period. In a pooling equilibrium, however, the true quality is not revealed and both types are perceived to be the same type; hence, $D(H) = D(L)$.

3.3 Shill Contract Decision

The seller offers payment A to make a shill contract that obligates the influencer, if he accepts the seller's offer, to send a high-quality signal about the quality of the seller's product. Because the influencer does not have complete information about the seller's product quality, the offered amount for shilling, A, may signal product quality due to H- and L-type sellers' different signaling costs or different expected future profits.

One may conjecture that offering a larger amount would signal low-quality because the low-quality seller would benefit more from manipulating the influencer's message, thereby leading to separation. We show that this conjecture turns out to be untrue.

Let (A_H, A_L) represent a pair of equilibrium shill offers by H- and L-type sellers, respectively. An equilibrium is separating if $A_H \neq A_L$ and pooling if $A_H = A_L = A_\lambda$.

Let $\hat{\lambda}_1(A)$ denote the first-period consumer's posterior belief that $q = H$ after observing the seller's choice of A. Then $\hat{\lambda}_1(A_H) = 1$ and $\hat{\lambda}_1(A_L) = 0$ in a separating equilibrium, and $\hat{\lambda}_1(A_\lambda) = \lambda$ in a pooling equilibrium.

We consider the possibility of separating equilibria. Because the seller's private information is revealed to the influencer with probability one in a separating equilibrium, we can use the influencer's posterior belief about the seller's type interchangeably with his perceived type q^e to describe any separating equilibrium.

Let $\Pi(A, q, q^e)$ be the sum of the seller's first- second-period profits when the seller's actual type is q, when his first-period perceived type in the eyes of the influencer is q^e, and when the seller chooses to offer advertising fee A. We assume zero discounting.

Each seller type's profit is computed as follows. If $(q, q^e) = (H, H)$, the seller's total profit is:

$$\Pi(A, H, H) = D^{t=1}(H)[(p - A) + \gamma p D^{t=2}(H)] \tag{5}$$

where

$$D^{t=1}(q^e) = 1 - \hat{\theta}_1(A; q^e), \tag{6}$$

$$D^{t=2}(m = H) = 1 - \hat{\theta}_2(H), \tag{7}$$

$$\hat{\theta}_1(A; q^e) = \frac{p - A}{q^e}, \tag{8}$$

$$\hat{\theta}_2(H) = \frac{p}{E(q|H)} = \frac{p}{\hat{\lambda}_2(H)H + (1 - \hat{\lambda}_2(H))L}, \tag{9}$$

$$\hat{\lambda}_2(H) = \frac{\lambda D(H)}{\lambda D(H) + (1-\lambda)D(L)}, \tag{10}$$

$$\text{and } D(q^e) = 1 - \frac{p - A}{q^e}. \tag{11}$$

If $(q, q^e) = (L, L)$, the low-quality seller's profit function in a separating equilibrium is computed similarly:

$$\Pi(A, L, L) = D^{t=1}(L)\left[(p - A) + \gamma p D^{t=2}(H))\right]. \tag{12}$$

Note that $D^{t=1}(q^e)$ is a function of q^e, while $D^{t=2}(m)$ is a function of m, because the first-period consumer can infer q^e from observing A but the second-period consumers only observe m, without observing A or q^e.

Comparing (5) and (12), one notices that only the first-period demands differ by seller type ($q = H$ versus $q = L$), whereas the second-period conditional demands are the same. The H-type seller enjoys greater first-period demand in a separating equilibrium, because we are considering the case that the seller's true type is revealed (i.e., $q = q^e$). Second-period conditional demands are the same because second-period consumers observe the same message, $m = H$, regardless of the seller's true type or perceived type. Because the high-quality seller's first-period demand is greater, its expected profit is higher in both periods: $\Pi(A, H, H) > \Pi(A, L, L)$ for any $A \geq 0$.

We can also calculate the profits when either type of seller pretends to be the other type (i.e., $q \neq q^e$):

$$\Pi(A, H, L) = D^{t=1}(L)\left[(p - A) + \gamma p D^{t=2}(H)\right], \tag{13}$$

$$\Pi(A, L, H) = D^{t=1}(H)\left[(p - A) + \gamma p D^{t=2}(H)\right]. \tag{14}$$

Only the first-period demand matters in determining the total profit in the two expressions above. Because first-period demand depends only on perceived (and not actual) type, so, too, does total profit only depend on perceived type. Let $\hat{\Pi}(A, q^e) \equiv \Pi(A, q, q^e)$ for any q. Because $\hat{\Pi}(A, H) > \hat{\Pi}(A, L)$, a low-type seller wants to pretend to be a high-type. However, a high-type seller cannot separate himself from a low-type seller by preventing the low-type seller's incentive to imitate the high-type seller, because the incentive to prevent L from

imitating H and the incentive for H to separate himself from L contradict each other, with the exception of knife-edge cases of payoff indifference.

There is no difference in signaling costs across actual types. Total profit does not depend on actual type. Therefore, separation is not possible except for the case in which a low-type seller is indifferent between imitating a high-type seller and not and, simultaneously, a high-type seller is indifferent between separating himself and not, which leads to the following proposition.

Proposition 2. *There exists no separating equilibrium except for a knife-edge case. In the knife-edge separating equilibrium, a high-type seller signals its high quality either by a high payment or by a low payment.*

As we argued before, a low-type seller always improves total expected profit by imitating a high-type seller, regardless of A, because there is no actual cost difference between the two types. Proposition 2 implies that there cannot be a separating equilibrium in which a high-type seller strictly prefers signaling their quality by either a low payment or a high payment. Any payment could be imitated by the low-type seller. As a result, the low-type seller's expected profit will always be the same as high-type seller's in equilibrium. Separation is possible only in knife-edge cases in which both types of sellers are indifferent between adhering to their respective equilibrium advertising fee and pretending to be the other type.

Figure 1 illustrates the two knife-edge separating equilibria labeled. As shown in Figure 1, the high-type seller can signal his type either by offering a lower or higher payment (than the low-type seller's equilibrium payment), so long as it is chosen to make himself indifferent between signaling his type and most profitably deviating (i.e., pretending), thereby being perceived to be a low type.[14]

There is no Schmalansee effect in this model, because production costs are assumed to be the same across the two seller types. There is no Nelson effect in this model either,

[14]One may wonder whether a stronger refinement could eliminate one of the two separating equilibria. However, it turns out that applying Cho and Kreps (1987)' Intuitive Criterion will not help (defined formally in Section 4 after introducing an extended version of the model in this section). A deviating payment must be in the region of Figure 1 labeled ICL (incentive-compatible shill offers for low-type sellers) to be equilibrium-dominated for a low type. But such a shill-offer payment cannot be profitable for a high type. The interior of the region labeled ICH (incentive-compatible for high types) shows where the high type's profitability condition is satisfied.

even though the presence of second-period consumers might appear to have a similar effect of increasing repeat purchases in the future. The difference in our model is that the influencer's communication (playing the role of word-of-mouth communication in Nelson (1970)) is not informative at all. In Nelson (1970, 1974), once a first-period consumer buys, more second-period consumers buy when the true quality of the product is high, because they learn the high quality through information transmission via word-of-mouth communication. In contrast, the influencer's message in the model of shill influencers as presented in this section is purely strategic rather than informative with respect to product quality. More accurately, the true quality does not affect the first-period message at all. Therefore, the second-period consumers cannot observe the true product quality of the product but only observe one message ($m = H$) from the first-period influencer. No meaningful information about product quality is conveyed to second-period consumers by the message. Hence, there is no Nelson effect.

3.4 Discussion on Social Welfare

As we mentioned in the introductory section, shill marketing is illegal in many jurisdictions. In particular, it is *per se* illegal in Korea according to the Korean Disclosure law and Korean Electric Commerce Law. It is similar in U.S. According to FTC disclosures 101 for Social Media Influencers and Creators, the FTC mandates that an influencer tells followers anytime something of value is exchanged that could impact his or her recommendation of a product or service. Also, Section 5 of FTC Act prohibits "unfair or deceptive acts or practices in or affecting commerce."

To see how social welfare can be affected by shill contracts, we need to analyze the case that shill contracts are forbidden by law and compare it with the results we previously obtained.

Without the possibility of shill contracts, consumers buy the product if

$$V = \theta E(q) - p \geq 0, \tag{15}$$

that is,

$$\theta \geq \frac{p}{E(q)} = \frac{p}{M} = \hat{\theta}, \tag{16}$$

where $M \equiv \lambda H + (1-\lambda)L$, because they receive no signal from which they can infer the product quality. This cutoff value $\hat{\theta}$ is valid for both first-period consumers and second-period consumers.

Since the influencer is always honest without shill contracts, his message m can be informative. Bayesian updating leads to

$$\hat{\lambda}_2(H) = \frac{\lambda\alpha}{\lambda\alpha + (1-\lambda)(1-\alpha)} = \frac{\lambda}{\lambda + (1-\lambda)\frac{1-\alpha}{\alpha}} > \lambda, \tag{17}$$

$$\hat{\lambda}_2(L) = \frac{\lambda(1-\alpha)}{\lambda(1-\alpha) + (1-\lambda)\alpha} = \frac{\lambda}{\lambda + (1-\lambda)\frac{\alpha}{1-\alpha}} < \lambda, \tag{18}$$

since $\alpha > \frac{1}{2}$. Thus, the profits of each type seller without shill contracts are calculated as

$$\Pi^N(H) = p(1-\hat{\theta})[1 + \gamma\{\alpha D^{t=2}(H) + (1-\alpha)D^{t=2}(L)\}], \tag{19}$$

$$\Pi^N(L) = p(1-\hat{\theta})[1 + \gamma\{(1-\alpha) D^{t=2}(H) + \alpha D^{t=2}(L)\}], \tag{20}$$

because H-type (L-type, respectively) seller receives $s = H$ with probability α ($1-\alpha$, respectively) and $s = L$ with probability $1-\alpha$ (α, respectively). Note that $\Pi^N(H) > \Pi^N(L)$ because $D^{t=2}(H) > D^{t=2}(L)$ insofar as $\alpha > \frac{1}{2}$. Since the prior probability of high-quality sellers is λ, the expected profit of a seller is

$$\mathbb{E}(\Pi^N(q)) = p(1-\hat{\theta})[1+\gamma\{(\lambda\alpha+(1-\lambda)(1-\alpha))D^{t=2}(H)+(\lambda(1-\alpha)+(1-\lambda)\alpha)D^{t=2}(L)\}]. \tag{21}$$

We now compare this profit with the profit when shill contracts are legally allowed. Since a separating equilibrium can hardly exist except in a knife-edge case, we only compare it with the outcome in a pooling equilibrium.

In a pooling equilibrium, both types of the seller use the same advertising payment $A_H = A_L = A$, so the advertising payment cannot convey any information about product quality. Then, each type seller's profit is given by

$$\Pi(H) = \Pi(L) = (p-A)(1-\hat{\theta}_{1,C})\left[1 + \gamma\left(1 - \frac{p}{M}\right)\right], \tag{22}$$

where $\hat{\theta}_{1,C} = \frac{p-A}{M}$. The reason why the profits of the two types are the same in a pooling equilibrium is clear. First, the first-period sales are the same, because they choose the same amount of advertising expenditures, so the influencer's perception on the product quality on

which the first-period profit relies is the same in pooling equilibrium. Second, the second-period profit also depends only on the belief about the quality, but since all the purchasers in the first period are dishonest, the second-period consumers do not update their beliefs by ignoring the message they receive ($m = H$). Therefore, the information about the quality can never be transmitted to any consumer, regardless of $q = L, H$.

Intuitively, allowing shills has an advantage as well as a disadvantage from a social point of view. On one hand, it reduces the credibility of WoM communication of the influencer due to shills. This is the advantage of forbidding shills. On the other hand, it increases the probability that the first-period influencer buys the product because of the lower effective price due to shill payments. However, if $\alpha \approx \frac{1}{2}$ in an extreme as in the case of credence goods, the advantage of prohibiting shills vanishes, because the signal itself is not informative at all. So, prohibiting shills will have only the disadvantage of reducing the first-period demand. Therefore, allowing shills can be socially better than prohibiting them, unless $L < 0$.

To see this intuition clearly, (21), which is the expected profit when shills are forbidden, is reduced to

$$\Pi^N(H) = \Pi^N(L) = p(1 - \hat{\theta}) \left[1 + \gamma \left(1 - \frac{p}{M}\right)\right], \tag{23}$$

where $\hat{\theta} = \frac{p}{M}$, if $\alpha = \frac{1}{2}$, because $D^{t=2}(H) = D^{t=2}(L) = 1 - \frac{p}{M}$. So, we can see from comparing (22) and (23) that if the seller can choose A strategically, the profit can be higher than that when shills are forbidden. This is the case particularly when the price is not so flexible in the short run.

Allowing shills do not reduce consumer surplus, either, in this model. If shills are not illegal, some influencers engage in shill contracts in a pooling equilibrium. Since they are always dishonest and the second-period consumers know it, they will never be deceived. In other words, they do not update their posterior belief even after they observe the WoM message from the dishonest influencer, implying that the second-period consumers will have the same posterior beliefs whether shills are legally permitted or not. However, since the seller can increase its first-period sales by choosing $A > 0$ when shills are legal, the consumer surplus will be also higher when shills are allowed.[15] Since the marginal cost is assumed to be zero, reducing the effective price to the profit-maximizing level by increasing A will improve allocation efficiency by increasing both the profit and the consumer surplus. **This result**

[15]If the seller can choose both p and A flexibly, he cannot be made better off by shill contracts, because it will choose the effective price that maximizes his profit in both cases.

has an interesting policy implication that shill marketing should be subject to a rule of reason rather than *per se* illegal.

4 Extended Model with Search Costs and a Limited Number of Shill Influencers

The analysis in the previous section suggests that the informativeness of a message requires at least some influencers who buy the product to send an authentic message. In this section, we assume that the seller has limited information about the availability of high-profile influencers and must incur search costs to find and contract with them.

We consider a simplified binary state space to represent costs incurred by the seller to find and contract with an influencer known to have a significant number of second-period followers and be willing to accept shill offers if the offered amount (A) is sufficiently large. The seller is either informed about the existence of a particular influencer with probability $\beta \in (0,1)$ or not informed with probability $1-\beta$. The seller can offer a shill contract only to an influencer he is informed of. This information (about whether the seller is informed of the particular influencer) is not known to second-period consumers.[16] Then, following the terminology in Section 3, the influencer who receives no shill offer is the "honest" or non-shill type, , whereas an offered influencer becomes a dishonest shill type.

4.1 Purchasing Decisions

The purchasing decisions of the second-period consumers are not affected except that the subjective probability that the influencer is honest ($\hat{\lambda}_2(m)$) is modified by the presence of honest non-shill influencers who buy the product and send a message without receiving payment and by the non-zero probability that messages saying that the product is high-quality may now be worth conditioning purchasing decisions on because there is a non-zero

[16] Alternatively, we can assume that an influencer is either innately honest with some probability $1-\beta \in (0,1)$ or strategic with the complementary probability. Being "innately honest" means that the influencer never accepts a shill contract even if he can benefit from it. That is, an innately honest influencer is ex ante honest and can be distinguished from one who is ex post honest (i.e., one who did not make a shill contract because it was not profitable to do so).

probability that such messages were sent by an honest non-shill influencer. This modified posterior belief (in the model below) is denoted $\tilde{\lambda}_2(m)$. Recall that, in Section 3, $m = L$ did not occur in equilibrium, which meant that $\hat{\lambda}_2(L)$ was arbitrary, because $m = L$ was an off-the-equilibrium message. In this section, $m = L$ is possible in equilibriumand, therefore, this modified posterior belief conditional on observing L ($\tilde{\lambda}_2(m)$) can be pinned down. Second, $\mu(H)$ is no longer zero, because an honest influencer may buy the product and send a non-shilling message reflecting positively on product quality ($m = H$). To incorporate the possibility of honest high-quality messages, $\tilde{\lambda}_2(m)$ must be modified accordingly.

Given $\tilde{\lambda}_2(m)$, second-period consumers purchase the product if:

$$V_2(m) \equiv E[U(q;p);\tilde{\lambda}_2(m)] = \theta E(q|m;\tilde{\lambda}_2) - p \geq 0, \tag{24}$$

or, equivalently, if $\theta \geq \frac{p}{E(q|m;\tilde{\lambda}_2)} \equiv \tilde{\theta}_2(m)$. Note that $\hat{\theta}_2(m) \neq \tilde{\theta}_2(m)$ because $E(q|m;\tilde{\lambda}_2) \neq E(q|m;\hat{\lambda}_2)$ due to the difference between $\tilde{\lambda}_2$ and $\hat{\lambda}_2$ seen by comparing (8) and (16). Conditional demand in the second period is $\gamma D^{t=2}(m)$, where $D^{t=2}(m) = 1 - \tilde{\theta}_2(m)$.

The purchasing decision of the first-period consumer remains the same as in Section 3. The conditions under which shill and non-shill influencers purchase the product are the same as (2) and (3), respectively, except that no shill offer (A) is observed by the honest influencer, which implies that $q^e = \lambda H + (1-\lambda)L \equiv M$. That is, a dishonest shill influencer buys one unit of the product if $\theta \geq \tilde{\theta}_{1,C}(A) = \frac{p-A}{E(q|A)} \equiv \hat{\theta}_{1,C}(A)$, and an honest non-shill influencer purchases one unit of the product if $\theta \geq \tilde{\theta}_{1,N} = \frac{p}{E(q)}$.

The validity of Lemma 1 is now slightly more restricted, in the sense that purchasing the product after accepting a shill offer is not an option for the non-shill influencer, who never receives or sees any shill offers in the first place.

Figure 2 shows the mapping from the range of possible values of θ into the influencer's purchasing decision for *unoffered* and *offered* influencer types (unobservable to second-period consumers). Unoffered influencers are honest non-shills. Offered influencers are those who are offered a shill contract because they are known to be willing to consider and possibly accept payment for sending positive messages about product quality. The reason for introducing new labels for influencer types in the extended model is because an offered influencer type may or may not accept the contract they are offered and therefore become a shill. They are, in principle, willing to work as shills. Thus, we shift labels by referring to unoffered (honest) and offered (willing to be dishonest) influencer types.

Unlike the baseline model in Section 3, Figure 2 shows that, in the extended model, it is possible that some unoffered influencers (i.e., without a shill contract) buy the product and send a message about its quality. The assumption from Section 3 remains, however, that offered influencers who accept a shill contract (i.e., shills) will always purchase, and otherwise (if they refuse payment A) never purchase. Due to this consideration, Proposition 1 is modified as follows.

Proposition 3. *(i) An unoffered influencer buys one unit of the product if $\theta \geq \tilde{\theta}_{1,N}$, and otherwise does not buy; and (ii) an offered influencer makes a shill contract and buys one unit of the product if $\theta > \hat{\theta}_1(A)$, and otherwise makes no shill contract and does not buy.*

This proposition implies that an influencer with $\theta \in (\hat{\theta}_{1,C}(A), \hat{\theta}_{1,N}(A))$ buys after making a shill contract if he is offered but does not necessarily buy without a contract if he is not offered.[17] High search costs represented by low values of $\beta(< 1)$ may reduce first-period demand for the product among influencers but increase second-period consumer demand by making the message of the influencer more credible.

4.2 Belief Updating

Although there are some honest influencers who actually buy the product and send a message contrary to the model in Section 2, it is still the case that $\mu(L) = 1$, because the dishonest influencer never uses the message $m = L$. Also, as we argued, this must be possible for an informative separating equilibrium to emerge, because an honest influencer who buys the product may send the message L if he learns that $s = L$.

On the other hand, if $m = H$, it is not the case that $\mu(H) = 0$, because an honest influencer may buy the product and receive the signal $s = H$. From the perspective of second-period consumers, their posterior belief that the influencer (who sent the positive message about product quality $m = H$) is honest can be computed as follows:

$$\mu(H) = \frac{(1-\beta)D(M)[\lambda\alpha + (1-\lambda)(1-\alpha)]}{(1-\beta)D(M)[\lambda\alpha + (1-\lambda)(1-\alpha)] + \beta[\lambda D(H) + (1-\lambda)D(L)]}, \tag{25}$$

where:

$$M = \lambda H + (1-\lambda)L, \tag{26}$$

[17] It is not necessarily the case that $\hat{\theta}_{1,C}(A) < \tilde{\theta}_{1,N} < \hat{\theta}_{1,N}(A)$, because $\hat{\theta}_{1,C}(A)$ and $\hat{\theta}_{1,N}(A)$ do not depend on λ but $\tilde{\theta}_{1,N}$ does.

$$D(M) = 1 - \tilde{\theta}_{1,N} = 1 - \frac{p}{M}, \tag{27}$$

$$D(L) = 1 - \hat{\theta}_{1,C}(A; L) = 1 - \frac{p - A}{L}. \tag{28}$$

The denominator in (25) is the probability that $m = H$, which is the sum of probabilities of two disjoint compound events (whose probabilities are products of independent probabilities): (i) that the influencer is honest $(1-\beta)$, and he buys the product $(D(M))$, and the signal he receives is H $(\lambda\alpha + (1-\lambda)(1-\alpha))$, whose product is the first term in the denominator; and (ii) that the influencer is dishonest (β), and he buys the product $(\lambda D(H) + (1-\lambda)D(L))$, and sends the positive message (1), whose product is the second term. The numerator is the probability that the message $m = H$ was sent by an honest influencer. Because the honest influencer was not offered a shill contract, he does not know A. Therefore, his expectation about q is $q^e E(q; \lambda)\lambda H + (1-\lambda)L = M$.

The posterior belief that the seller and his product is a high-quality type $(q = H)$ after observing the message $m = H$ can be computed similarly as follows;

$$\begin{aligned} \tilde{\lambda}_2(H) &= \frac{(1-\beta)\lambda\alpha D(M) + \beta\lambda D(H)}{(1-\beta)D(M)[\lambda\alpha + (1-\lambda)(1-\alpha)] + \beta[\lambda D(H) + (1-\lambda)D(L)]} \\ &= \frac{\lambda}{\lambda + (1-\lambda)\frac{(1-\beta)(1-\alpha)D(M)+\beta D(L)}{(1-\beta)\alpha D(M)+\beta D(L)}} > \lambda, \end{aligned} \tag{29}$$

where the inequality follows from $R \equiv \frac{(1-\beta)(1-\alpha)D(M)+\beta D(L)}{(1-\beta)\alpha D(M)+\beta D(L)} < 1$. The denominator of (29) is the probability that $m = H$ is sent, which is the same as $\mu(H)$ as given in (25). The numerator is the probability that $m = H$ when the actual quality is H.

R is the ratio of the probability that a low-quality product is purchased and a high-quality message is sent $(m = H)$ to the probability that a high-quality product is purchased and a high-quality message is sent. The probability that a dishonest influencer purchases and the probability that an honest influencer obtains the signal $s = H$ are both greater when $q = H$. Thus, it is clear that $R < 1$. This means that, unlike the previous model in Section 3, the message $m = H$ is now informative, revealing to second-period consumers that that $q = H$ is more likely.

The least informative case occurs when almost all influencers are honest $(\beta \to 0)$ and the signal is inherently noisy $(\alpha \to \frac{1}{2})$. If the signal is very inaccurate but almost all the influencers are dishonest $(\beta \to 1)$, then a high-quality product is more likely to be purchased by them. Therefore, the fact of receiving any message at all causes the posterior belief

that $q = H$ to adjust upwards. Another difference in the extended model presented in this section is that signal-generating probabilities over types (i.e., a difference in α and $1 - \alpha$) is meaningful for the informativeness of any message sent only when there are some honest influencers ($\beta \neq 1$).

Whenever second-period consumers observe $m = H$, their posterior belief that the product is high-quality increases which, in turn, increases second-period demand. Therefore, one might conjecture that high-quality sellers would be more likely to generate high-quality signals received by influencers ($s = H$) and may have a stronger incentive to manipulate influencers to send high-quality signals ($m = H$) by increasing A, because the influencer's message may have information content and is not entirely babbling (i.e., $\beta < 1$). In the next subsection, we will show that this conjecture is incorrect.

4.3 Shill Contract Decision

Due to the existence of an unoffered influencer who makes no shill contract, demand for the seller's product and the seller's profit are modified (relative to the baseline model in Section 3) accordingly.

Let $\tilde{\lambda}_1(A)$ denote the posterior belief of the first-period influencer who observes offer A in this extended model. In a separating equilibrium, the influencer who is offered a shill payment will have certainty about the seller's product quality based on the shill offer (amount of money offered by the seller to the influencer to shill): $\tilde{\lambda}_1(A_H) = 1$ and $\tilde{\lambda}_1(A_L) = 0$. Denoting the (perceived) product-quality type that the offered influencer infers conditional on A as q^e, the high-type seller's profit is:

$$\Pi(A, H, q^e) = (1 - \beta)p(1 - \frac{p}{M})[1 + \gamma Q(H)] + \beta(1 - \frac{p - A}{q^e})[(p - A) + \gamma p D^{t=2}(H)], \quad (30)$$

where:

$$D^{t=2}(H) = 1 - \frac{p}{E(q|H, \tilde{\lambda}_2(H))}, \quad (31)$$

$$E(q|H, \tilde{\lambda}_2(H)) = \tilde{\lambda}_2(H)H + (1 - \tilde{\lambda}_2(H))L, \quad (32)$$

$$Q(H) = \alpha D^{t=2}(H) + (1 - \alpha)(1 - \frac{p}{L}). \quad (33)$$

The notation in the last expression above (33) makes explicit that $Q(H)$ is distinct from $D^{t=2}(H)$. The latter ($D^{t=2}(H)$) is second-period consumers' conditional demand when

the influencer's message is $m = H$. Dishonest influencers only send the message $m = H$, regardless of the realized value of s he observes. Taking this possibility that some high-quality messages may be false (i.e., manipulated by an unobservable shill contract), second-period consumers make their purchasing decisions based on their expectations about product quality.

On the other hand, $Q(H)$ in equation (33) represents second-period consumers' conditional demand when the seller's type is high-quality ($q = H$), he did not make any shill offer, and an honest (unoffered) influencer sent a message of either type ($m = H$ or $m = L$). If $q = H$, the influencer's message can be either H or L, depending on the realization of the noisy experience-good signal s that the honest influencer receives after purchasing. Because the unoffered influencer does not manipulate the message, he sends the message $m = H$ with probability α and $m = L$ with probability $1-\alpha$. Therefore, second-period consumer demand in (33) is a weighted sum of second-period demand when $m = H$ (with probability α) and when $m = L$ (with probability $1 - \alpha$). Note that second-period demand conditional on $s = H$ is $D^{t=2}(H)$—not $1 - \frac{p}{H}$—even though the influencer is honest, because second-period consumers do not know whether the influencer is honest or not.

Similarly, low-type seller's profit is:

$$\Pi(A, L, q^e) = (1 - \beta)p(1 - \frac{p}{M})[1 + \gamma Q(L)] + \beta(1 - \frac{p - A}{q^e})[(p - A) + \gamma p D^{t=2}(H)], \quad (34)$$

where

$$Q(L) = (1 - \alpha)D^{t=2}(H) + \alpha(1 - \frac{p}{L}). \quad (35)$$

A dishonest influencer always sends message $m = H$ even if he buys the product and receives a low-quality signal $s = L$. Therefore, second-period conditional demand is $D^{t=2}(H)$, not $D^{t=2}(L)$.

It is clear that $\Pi(A, q, H) > \Pi(A, q, L)$ for any $A \geq 0$. It is also easy to see that $\Pi(A, H, q^e) > \Pi(A, L, q^e)$ for any $A \geq 0$, since $Q(H) > Q(L)$ due to $\alpha > \frac{1}{2}$.

In this model, a low-type seller may want to pretend to be a high type, because $\Pi(A, L, H) > \Pi(A, L, L)$. By doing so, he can increase first-period demand from the dishonest influencer (in expectation, by increasing the likelihood that the dishonest influencer will accept the seller's offer and purchase the product). It does not increase the demand from the unoffered honest influencer, however, because the unoffered influencer cannot observe A.

The low-quality seller's incentive to pretend by posing as a high-quality type can be prevented, because his profit depends on actual type (the second argument in the total profit function, in addition to perceived type which is the third argument): $\Pi(A, H, q^e) \neq \Pi(A, L, q^e)$, which implies that signaling costs differ across actual types.

Intuitively, if a seller increases A, he lowers the effective price $p - A$ for influencers (also known as first-period consumers), thereby increasing first-period demand (in expectation or, equivalently, the probability that an influencer will use the product). When a seller increases their shill offer A, there is no effect on the effective price that second-period consumers face, because the seller does not pay advertisement fees (A) to second-period consumers. However, the seller's shill offer A can increase second-period consumer demand (only in the extended model in Section 4) by generating more positive messages from more purchases by dishonest shill influencers.

The amount by which the seller's shill offer A increases demand consists of both quality and price effects. Demand increases if the seller's perceived quality increases or if the product's effective price decreases, both of which can be accomplished by the seller's decision to increase A.

Increasing first-period demand (which means increasing the likelihood that influencers will use the product and transmit messages about it) is a price effect, because increasing A lowers the effective price, thereby increasing first-period demand among influencers. This price effect is larger for low-quality sellers (i.e., the sensitivity to first-period demand is more price-sensitive when the seller's true product quality is low). We refer to this difference in price sensitivity by seller's quality type as the *elasticity effect*, which means that the seller of a low-quality product chooses a lower (effective) price.

On the other hand, an increase in A increases the second-period demand more when product quality is high, because high-quality products are more likely to generate high-quality signals ($s = H$) from honest non-shill influencers. Therefore, sellers with products of different quality types will make different trade-offs based on the first-period price effect and the quality effect that affects both the probability of positive message transmission from honest influencers in the first period and its knock-on effect of increasing second-period consumer demand. The relative importance of price and quality effects on first-period uptake by influencers and second-period consumer demand is scaled by the parameter γ, which measures the mass of second-period consumers relative to mass of influencers that has been

normalized to 1.

If γ is so small that the seller regards first-period consumers as more important, then increasing A benefits a low-type seller more. For sufficiently small γ, a low-type seller will prefer a higher A than a high-type seller. In this low-*gamma* case, small shill offers (low A) signals high quality.

On the other hand, if γ is sufficiently large so that second-period consumers are more important to the seller, then increasing A benefits a seller of high-quality products more. In this case, high-A offers signal high-type product quality, because a high-type seller chooses higher A to separate himself by preventing low-type sellers from imitating him.

In both low- and high-γ cases mentioned above, separation will be possible due to the differences, by the seller's product-quality type, in the benefit the seller receives from shill-marketing expenditure A. However, we will show in the remaining subsections below that the latter (high-γ) case in which A serves to signal high-quality product disappears in equilibrium(i.e., is degenerated to a zero-measure subset of the exogenous parameters in the model).

Thus, in equilibrium, high quality can only be signaled by sellers who make low-A shill offers. The prediction is stark (regardless of the relative size of the second-period consumer market (γ)): sellers who use hidden contracts to pay large amounts to influencers are unlikely to supply high-quality products. For consumers who follow highly-paid influencers to decide which products to purchase, our model would suggest a warning is in order. Shilling pays disproportionately well for sellers of low-quality products.

4.4 Equilibrium Analysis

As a benchmark to begin from, we consider the special case of $\gamma = 0$, which removes any consideration by sellers about trading off first- against second-period profits because there are no second-period consumers. The sellers' total profit functions (Equations (21) and (25)) then simplify to:

$$\Pi(A, H, H; \gamma = 0) = (1 - \beta)p(1 - \frac{p}{M}) + \beta(p - A)(1 - \frac{p - A}{H}), \tag{36}$$

$$\Pi(A, L, L; \gamma = 0) = (1 - \beta)p(1 - \frac{p}{M}) + \beta(p - A)(1 - \frac{p - A}{L}). \tag{37}$$

Note that both profit equations above are quadratic in A. Therefore, this special case yields unique global maxima, one for the seller of each type. If we denote the first-best solution for H-and L-type sellers by $A^*(H)$ and $A^*(L)$, respectively, one easily sees that $A^*(H) = p - \frac{H}{2}$ and $A^*(L) = p - \frac{L}{2}$, from which it follows that low-type sellers spend more on shilling than high-type sellers do: $A^*(L) > A^*(H)$, because $H > L$.

Lemma 2. *There exists* $\underline{\gamma}(> 0)$ *such that for any* $\gamma \leq \underline{\gamma}$, *(i)* $\Pi(A, H, H; \gamma)$ *and* $\Pi(A, L, L; \gamma)$ *have unique global maxima with respect to* A, *and (ii)* $A^*(H; \gamma) < A^*(L; \gamma)$.

For now, we restrict attention to the case that $\gamma \leq \underline{\gamma}$. To characterize the whole set of separating equilibria, we will assume the most pessimistic off-the-equilibrium belief $\tilde{\lambda}_1(A) = 0$ for any off-the-equilibrium payment A, which gives the most severe punishment when the seller deviates from an equilibrium payment. The following lemma turns out to be useful for characterizing the set of all possible separating equilibria.

Lemma 3. *In any separating equilibrium,* $A_L(\gamma) = A^*(L; \gamma)$.

This lemma implies that a low type's equilibrium payment will not be distorted from its first-best payment $A^*(L; \gamma)$. The proof is omitted because it is clear. If $A_L(\gamma) \neq A^*(L; \gamma)$, then a low-type seller would prefer deviating to $A^*(L; \gamma)$ under the most pessimistic belief.

The high type's equilibrium payment A_H must satisfy the following two incentive-compatibility conditions:

$$\Pi(A_L, L, L) \geq \Pi(A_H, L, H), \qquad [ICL1]$$

$$\Pi(A_H, H, H) \geq \Pi(A, H, L), \forall A \neq A_H. \qquad [ICH1]$$

The first inequality, labeled as [ICL1], is the L-type's incentive-compatibility condition. It requires that A_H is too costly for a low-type seller to imitate (i.e., ensuring that the low-type seller does not find it profitable to imitate the high type's payment A_H.) On the other hand, [ICH1], which is the H type's incentive compatibility condition, requires that A_H is not so costly for a high-type seller that he would not prefer deviating from A_H to be perceived as a low type.

Figure 3 shows the region of A_H that satisfies both incentive compatibility conditions. In Figure 3, $\underline{A}_L$ and $\overline{A}_L$ are values of A that satisfy [ICL1] with equality, and $\underline{A}_H$ and $\overline{A}_H$ are values of A that satisfy [ICH1] with equality. Figure 3 illustrates a continuum of separating equilibria, denoted by the segment E, in which high quality is signaled by a low payment. We

can single out a unique separating equilibrium by invoking the stronger equilibrium concept of Cho and Kreps' (1987) Intuitive Criterion (abbreviated as "C-K") as follows.

We say that a weak perfect Bayesian equilibrium (A_H, A_L) fails to pass the C-K Intuitive Criterion if there exists a payment $A(\neq A_H, A_L)$ such that:

$$(CK-i) \qquad \Pi(A_L, L, L) \geq \Pi(A, L, q^e), \forall q^e = L, H, \tag{38}$$

$$(CK-ii) \qquad \Pi(A_H, H, H) < \Pi(A, H, H). \tag{39}$$

Inequality (38) implies that an off-the-equilibrium A is equilibrium-dominated for type L. Inequality (39) implies that whenever the first-period consumer believes that the payment A was chosen by type H (for whom A is not equilibrium-dominated), then type H would have an incentive to deviate to A from A_H. If there exists a payment (A) that satisfies these two conditions, then (A_H, A_L) cannot pass the C-K Intuitive Criterion, because type H would have an incentive to deviate. Applying C-K Intuitive Criterion, we obtain the following proposition.

Proposition 4. *Assume that $\frac{H}{L} > \frac{4}{3}$ and $H - L < p < L$. Then, for any $\gamma \leq \underline{\gamma}$, there exists a unique separating equilibrium $(A_L, A_H) = (A^*(L), \overline{A}_L)$ that satisfies the C-K Intuitive Criterion. In this equilibrium, $A_L > A_H > 0$.*

This proposition implies that a seller's high quality is signaled only by a low payment for a shill contract. Also, it says that the unique separating equilibrium exists if the difference between H and L is large enough and $H - L < p < L$. This is just a sufficient condition to ensure positive amounts of equilibrium advertising fees.[18]

Proposition 4 says that the high-type seller does not need to use any signal that costs more than $\underline{A}_L$, which is the well-known "Riley outcome." The Riley outcome $\underline{A}_L$ is the most efficient separating equilibrium because it incurs the least signaling cost. To see why applying the Intuitive Criterion can eliminate all separating equilibria except for the most efficient one $(\underline{A}_L)$, for any equilibrium payment of the high type, $A_H \in [\underline{A}_H, \underline{A}_L)$, consider the following off-the-equilibrium message $A' \in (A_H, \underline{A}_L)$ as depicted in Figure 3.

[18]Let $\underline{A}_H$ be A satisfying $\Pi(A, H, H; \gamma = 0) = \Pi(A^*(L), L, L; \gamma = 0)$. For $\underline{A}_H > 0$, it is required that $\Pi(0, H, H; \gamma = 0) < \Pi(A^*(L), L, L; \gamma = 0)$, which is reduced to $p(1 - \frac{p}{H}) < (p - A^*(L))(1 - \frac{p - A^*(L)}{L})$. Since $A^*(L) = p - \frac{L}{2}$, this condition boils down to $p(1 - \frac{p}{H}) < \frac{L}{4}$, or equivalently, $\varphi(p) \equiv p^2 - Hp + \frac{LH}{4} < 0$. Since $\max \varphi(p) = \varphi(\frac{H}{2}) = \frac{H}{4}(L - H) < 0$ and $\varphi(L) = \varphi(H - L) = L(L - \frac{3}{4}H) < 0$ if $\frac{H}{L} > \frac{4}{3}$, it is proved that $A_L > A_H > \underline{A}_H > 0$ for $p \in (H - L, L)$ and for small γ.

In Figure 3, it is easy to see that $\Pi(A^*(L), L, L) > \Pi(A', L, L), \Pi(A', L, H)$, which implies that A' is equilibrium -dominated for low-quality types. Once the influencer observes A' and therefore eliminates the possibility that the seller is a low type, it follows that $\Pi(A_H, H, H) < \Pi(A', H, H)$, since $A_H < A' < \underline{A}_L < A^*(H)$. Therefore, C-K Intuitive Criterion requires that $\tilde{\lambda}_1(A') = 1$. This result implies that the equilibrium payment for a high type, $A_H \in [\underline{A}_H, \underline{A}_L)$, fails to pass the Intuitive Criterion.

Due to the elasticity effect, a given amount offered (as shill payment) increases the low-type seller's first-period demand more than it does the high type's. A high type knows that his marginal benefit with respect to an increase in A is lower. He therefore offers less to the influencer. The low type may want to imitate the high type because of the quality effect (meaning that an increase in perceived quality increases product demand). But he cannot profitably imitate the low payment offered by the high type due to the credibility effect, which means that a message that product-quality is high-type becomes more credible as fewer dishonest influencers purchase the product. A high-type seller prefers the dishonest influencer being less likely to purchase the product because it increases $\tilde{\lambda}_2(H)$, the credibility of the honest influencer's message H. His gain from the difference between $D^{t=2}(H)$ and $D^{t=2}(L)$ is higher as the credibility of his WoM communication is higher. Therefore, the high-type seller will prefer a high effective price (i.e., low shill offer A). On the other hand, the gain is lower to a low type. So, credibility is not so much important as to a high type. That is, if a seller loses credibility of his message by lowering the advertising fee, it is more costly to a high type.

One may conjecture that if γ is very large, a high-quality seller may have a stronger incentive to offer more (higher A) to induce more first-period influencers to buy the product (at the expense of first-period profits), because higher A could generate more high-type experience signals among influencers, thereby increasing second-period consumer demand by an amount that more than offsets first-period reductions in profit. However, it turns out that this conjecture is false.

This mechanism would only be possible if shills (i.e., first-period influencers that buy the product after accepting the seller's shill offer) could transmit informative signals of high product quality. The messages that shills send (in both baseline and extended models) are never informative, however. The fact that shills have accepted payment A and are therefore shills implies that the messages they transmit can never rationally increase second-period

consumer demand. This intuition is formalizes in the second-period conditional demands for high- and low-quality products, which are precisely the same: $D^{t=2}(H)$.

Messages sent by honest influencers do convey information that influence second-period conditional demand: $Q(H) > Q(L)$. Second-period consumers never observe A directly and are never contacted with any messages directly from the seller. Therefore, second-period consumers' purchasing decisions are never affected by small changes in the magnitude of A beyond the seller having chosen A to land in the proper regions satisfying the threshold conditions to support the separating equilibrium. Moreover, second-period consumers cannot tell which influencers are honest or not. Thus, increasing A is no more beneficial for a high- than for a low-type seller. Therefore, even if γ is large, Proposition 4 will not be threatened insofar as the sellers' profit functions remain concave in A so that unique global optima continue to exist.

What is the main effect of increasing A under incomplete information? Because second-period consumers do not know whether the influencer is honest or dishonest, increasing A only lowers the credibility of messages, regardless of whether $q = L$ or $q = H$ and regardless of whether the influencer is honest or dishonest. Conversely, reducing A increases the credibility of the honest influencer's message. If there is a greater proportion of honest influencers (i.e., β is low), then the beneficial marginal effect of increasing credibility by reducing A is larger. Also, as the signal s becomes more accurate (i.e., α is larger), this beneficial credibility-enhancing effect becomes larger. In business environments with these characteristics, a high-type seller does not need to engage in costly signaling (by employing shill influencers) much beyond the first-best shill-marketing intensity, which implies that $A_H(\gamma)$ will be closer to the first-best solution $A^*(H;\gamma)$.

In terms of social welfare, the general insight that we obtained in Section 3 remains unaffected. Since the profit of a seller when shills are illegal is the same as (21) and it is reduced to (23) if $\alpha = \frac{1}{2}$, the disadvantage of allowing shills that WoM communication is less credible disappears when $\alpha = \frac{1}{2}$, and the clear advantage is that the seller sells more either in a separating equilibrium or in a pooling equilibrium. Again, allowing shills can be socially better if the signal that consumption experience generates is not so precise.

4.5 Discussion on Endogenizing the Price

To highlight the role of secret advertising as a quality signal, we have, so far, assumed that price is exogenously given. In this section, we relax the assumption by assuming that both price and advertising are choice variables, and examine whether our result is robust against this modification.

Strategic interaction between the seller and consumers remains the same except that the seller chooses p and A both of which are observable to the first-period consumers, whereas A is not observable to the second-period consumers.

If the price can signal product quality, the seller has no reason to use costly shill contracts because any second-period consumer who observes the price can infer the true quality of the product regardless of the influencer's message. So, we focus only on separating equilibria in which the seller uses a separating advertising payment and a pooling price.

The seller has a pair of signaling devices, p and A, which jointly determine an effective price (denoted $\tilde{p}$) to first-period consumers given by the formula: $\tilde{p} = p - A$.

By abusing notation, we denote each type's advertising choices by A_H and A_L, and the corresponding effective prices by $\tilde{p}_H$ and $\tilde{p}_L$. Let $\pi(\tilde{p}, p, q, q^e)$ represent the profit of the q-type seller given that the first-period consumers' belief about the seller's type is q^e. Again, we assume that influencer' belief formation is pessimistic in the sense that: $\hat{\lambda}(A) = 0$ for any $A \neq A_H$.

In the benchmark case that $\gamma = 0$, by using $\tilde{p}$, we can rewrite (36) and (37) as

$$\Pi(\tilde{p}, p, H, H; \gamma = 0) = (1-\beta)p(1 - \frac{p}{M}) + \beta\tilde{p}(1 - \frac{\tilde{p}}{H}), \tag{40}$$

$$\Pi(\tilde{p}, p, L, L; \gamma = 0) = (1-\beta)p(1 - \frac{p}{M}) + \beta\tilde{p}(1 - \frac{\tilde{p}}{L}). \tag{41}$$

Let p^* be the equilibrium price. The high type's equilibrium effective price $\tilde{p}_H$ must satisfy the following two incentive-compatibility conditions:

$$\Pi(\tilde{p}_H, p^*, H, H) \geq \Pi(\tilde{p}, p, H, L), \forall \tilde{p} \neq \tilde{p}_H, \forall p \neq p^*, \qquad [ICH2]$$

$$\Pi(\tilde{p}_L, p^*, L, L) \geq \Pi(\tilde{p}_H, p^*, L, H). \qquad [ICL2]$$

Note that the iso-profit curves of each type sellers are ellipses if $\gamma = 0$. So, if γ is very small, the iso-profit curves are slightly deformed from the ellipses. In Figure 4, [ICH2]

corresponds with the inside of the larger right (almost) ellipse illustrated in blue, while [ICL2] corresponds with the outside of the left ellipse illustrated in green. The set of $(\tilde{H}, p^*)$ satisfying both incentive compatibility conditions is the intersection of the two areas which is illustrated in a crescent shape.

However, not all points satisfy the Intuitive Criterion. For example, consider the point $P = (\tilde{p}_H, p^*)$. If a high-type seller deviates to $Q = (\tilde{p}', p')$, we have

$$\Pi(\tilde{p}', p', H, H) > \Pi(\tilde{p}_H, p^*, H, H) \geq \max_{\tilde{p}, p} \Pi(\tilde{p}, p, H, L).$$

Thus, Intuitive Criterion requires the influencer to believe that the message came from a high type. This implies that the original equilibrium P is overturned, because a high type will have an incentive to deviate from P to Q. This means that for $(\tilde{p}_H, p^*)$ to pass the Intuitive Criterion, there must be no way to move inside the ellipse with satisfying [ICL2], implying that the iso-profit curves of the low type and the high type must be tangent, as depicted by red point R. Also, note that $\tilde{p}_H < p$, i.e., $A_H > 0$ if $\frac{H}{L} > \frac{4}{3}$ and $p^* \in (H - L, L)$, i.e., $\frac{2H-3L}{H-L} < \lambda < \frac{L}{H-L}$.[19] It shows that (possibly positive) lower advertising signals higher quality.

Milgrom and Roberts (1986) argue that any separating equilibrium that involves dissipative advertising ($A_H > 0$) fails to pass the Intuitive Criterion if the profit functions of different types satisfy the single crossing property. Why do we obtain a contrasted result in a model with the single crossing property? It is mainly due to their assumption that advertising is just dissipative and has no effect on the credibility of WoM communication, contrary to our assumption that it can affect the credibility by altering the first-period sales. In our model, if a seller pays an influencer a higher advertising fee, it lowers the effective price that the influencer bears. So, an increase in the advertising fee has the direct effect of reducing the seller's profit through increasing the seller's cost but also has the indirect effect of increasing the profit through increasing the sales. That is, a seller's advertising payment is not dissipative in our model. Therefore, a seller can signal its high type by nonzero advertising payment.

[19] (40) and (41) both achieve their maxima at $p = \frac{M}{2}$. If a pooling equilibrium price is $p^* = \frac{M}{2} \in (H-L, L)$, we get $A_H > 0$ for the price and for any $\gamma \leq \underline{\gamma}$. Since $M = \lambda H + (1 - \lambda)L$, this condition is reduced to $\lambda \in (\frac{2H-3L}{H-L}, \frac{L}{H-L})$.

5 Conclusion and Caveats

In this paper, we showed that a seller of an experience good can signal the quality of its product by offering a lower (rather than higher-than-average) payment for shill advertising. More generally, expenditures on shill influencers can help sellers, regardless of their quality type, sell more in the first period by lowering the product's effective price, although whether it can help increase second-period consumer demand may depend on the quality of the product, the proportions of paid versus unpaid influencers that consumers follow, search costs that sellers incur when seeking to establish shill contracts with paid influencers, and the subjectivity of product quality (i.e., accuracy of the signal that influencers and consumers receive when they consume the product).

Appendix

Proof of Lemma 1: The payoffs of the influencer who buys the product when he makes no shill contract and when he makes a shill contract are:

$$V_1^N = \theta E(q|A) - p, \tag{42}$$

$$V_1^C = \theta E(q|A) - p + A. \tag{43}$$

Comparing (42) and (43) clearly implies that $V_1^C > V_1^N$ for any $A > 0$. Since the choices of the second-period consumers do not affect V_1^C and V_1^N, the proof is immediate. ■

Proof of Proposition 1: It remains to prove the contracting decision. Note that the contracting is a binding decision, meaning that if an influence makes the shill contract, he must buy the product. Therefore, to see the contracting decision, it suffices to compare the payoffs of an influencer who made a contract when he buys and when he does not buy. From (43), we have:

$$V_1^C = \theta E(q|A) - p + A \geq 0,$$

if and only if $\theta \geq \hat{\theta}_{1,C}(A)$. This implies that $\hat{\theta}_1(A) = \hat{\theta}_{1,C}(A)$. It is clear from Lemma 1 that any influencer with $\theta \in (\hat{\theta}_{1,C}(A), \hat{\theta}_{1,N}(A))$ prefers to make a shill contract rather than make no contract. ■

Proof of Proposition 2: Assuming that $\hat{\Pi}(A, q^e)$ has the global maximum for any $q^e = H, L$, we let $A^*(q^e) = \arg\max_A \hat{\Pi}(A; q^e)$. Since a low-type seller wants to imitate a high-type seller for any A under incomplete information, we will assume the worst off-the-equilibrium belief $\hat{\lambda}_1(A) = 0$ for any $A \neq A_H, A_L$ in any separating equilibrium. Then, the equilibrium payments of both types, A_H and A_L, must satisfy the following two incentive compatibility conditions:

$$\Pi(A_L, L, L) \geq \Pi(A_H, L, H), \tag{44}$$

$$\Pi(A_H, H, H) \geq \Pi(A, H, L), \forall A. \tag{45}$$

Inequality (44) is the incentive-compatibility condition of type L (ICL), and inequality (45) is the incentive-compatibility condition of type H (ICH). Note that $A_L = A^*(L)$. Otherwise, i.e., if $A_L \neq A^*(L)$, a low type would profitably deviate to $A^*(L)$, because $\Pi(A^*(L), L, L) > \Pi(A_L, L, L)$. Since $\hat{\Pi}(A, L) = \Pi(A, H, L) = \Pi(A, L, L)$, $\hat{\Pi}(A, H) = \Pi(A, H, H) = \Pi(A, L, H)$ and (45) must be also satisfied for $A = A_L$, satisfying (44) and (45) is reduced to satisfying:

$$\hat{\Pi}(A_H, H) = \hat{\Pi}(A_L, L). \tag{46}$$

Figure 1 shows the two points for the high-type seller's equilibrium offer A_H that satisfy (46). ■

Proof of Proposition 3: What remains is to analyze the purchase decision of an unoffered influencer. If an influencer is not offered a shill contract, he has no option to make the contract. Without a contract, the honest influencer buys the product if and only if:

$$V_1 = E(U(q; p); \lambda) = \theta E(q; \lambda) - p = \theta[\lambda H + (1 - \lambda)L] - p \geq 0,$$

i.e., $\theta > \tilde{\theta}_{1,N} \equiv \frac{p}{M}$. ■

Proof of Lemma 2: Let $\hat{\Pi}(A, q, \gamma) = f(A, q) + \gamma g(A, q)$, where:

$$f(A, q) = (1 - \beta)p(1 - \frac{p}{M}) + \beta(1 - \frac{p - A}{q})(p - A),$$

$$g(A, q) = (1 - \beta)p(1 - \frac{p}{M})Q(q) + \beta(1 - \frac{p - A}{q})pD^{t=2}(H).$$

We have $\hat{\Pi}'(A^*(q, 0); \gamma = 0) = 0$ and $\hat{\Pi}''(A; \gamma = 0) < 0$.

(ii) Since $\hat{\Pi}(A, q; \gamma)$ is continuous in A and γ for any $q = L, H$, $A^*(q, \gamma)$ is continuous in γ by Berge's maximum theorem. It is easy to see that $A^*(H; 0) < A^*(L; 0)$ because $f(A; q)$ is a quadratic function and $\frac{p-A}{H} < \frac{p-A}{L}$ if $p - A > 0$. This implies that $A^*(H; \gamma) < A^*(L; \gamma)$ for any $\gamma < \gamma_1$ for some γ_1.

(i) Since $\hat{\Pi}''$ is continuous in γ, there exists $\gamma_2 (> 0)$ such that for any $\gamma < \gamma_2$, $\hat{\Pi}''(A, q, \gamma) < 0$ for any A, q, because $\hat{\Pi}''(A, q, \gamma = 0) < 0$ for any A, q.

Taking $\underline{\gamma} = \min\{\gamma_1, \gamma_2\}$ completes the proof. ■

Proof of Proposition 4: The two incentive compatibility conditions given by ICL1 and ICH1 imply that the set of A_H that are possible in any separating equilibrium is $[\underline{A}_L, \underline{A}_H] \cup [\bar{A}_L, \bar{A}_H]$.

By Lemma 1, it remains to show that $A_H = \bar{A}_L$ is the unique A_H that satisfies the C-K Intuitive Criterion.

Claim 1. *(i)* $\underline{A}_H < \underline{A}_L$*, i.e.,* $[\underline{A}_H, \underline{A}_L] \neq \emptyset$*. (ii)* $\bar{A}_H < \bar{A}_L$*, i.e.,* $[\bar{A}_L, \bar{A}_H] = \emptyset$*.*

Proof. Let us begin with the case that $\gamma = 0$. Then, we have:

$$\Pi(A, q, q^e) = (1 - \beta)p(1 - \frac{p}{M}) + \beta(p - A)(1 - \frac{p - A}{q^e}),$$

which does not depend on q. Thus, $\Pi(A, H, H) = \Pi(A, L, H)$ and $\Pi(A, H, L) = \Pi(A, L, L)$ for all A. These results imply that $\underline{A}_H = \underline{A}_L$ and $\bar{A}_H = \bar{A}_L$, as shown in Figure 1.

Now, if $\gamma > 0$, $\Pi(A, H, H; \gamma) > \Pi(A, L, H; \gamma)$ and $\Pi(A, H, L; \gamma) > \Pi(A, L, L; \gamma)$. Let $\Delta\Pi(H; \gamma) \equiv \Pi(A, H, H; \gamma) - \Pi(A, L, H; \gamma)$ and $\Delta\Pi(L; \gamma) \equiv \Pi(A, H, L; \gamma) - \Pi(A, L, L; \gamma)$. We have:

$$\Delta(H; \gamma) = (1 - \beta)p(1 - \frac{p}{M})\gamma\phi, \tag{47}$$

$$\Delta(L; \gamma) = (1 - \beta)p(1 - \frac{p}{M})\gamma\phi, \tag{48}$$

where $\phi \equiv Q(H) - Q(L) = (2\alpha - 1)[D^{t=2}(H; A) - (1 - \frac{p}{L})] > 0$, since $\alpha > \frac{1}{2}$ and $D^{t=2}(H) > 1 - \frac{p}{L}$. Note that ϕ is strictly decreasing in A, because an increase in A decreases $\mu(A), \hat{\lambda}_2(A)$, so $D^{t=2}(H; A)$.

(i) We have $\Delta\Pi(A, H; \gamma) = \Delta\Pi(A, L; \gamma)$. Since $\underline{A}_H < A^*(H)$, it is clear that $\Delta\Pi(\underline{A}_H, H; \gamma) > \Delta\Pi(A^*(H), L; \gamma)$, because $\Delta\Pi(A, H; \gamma)$ and $\Delta\Pi(A, L; \gamma)$ are both strictly increasing in A. This implies that $\underline{A}_H < \underline{A}_L$.

(ii) It suffices to show that $\Delta\Pi(A^*(H), H; \gamma) > \Delta\Pi(\bar{A}_L, L; \gamma)$. Straightforward calculations reduce the inequality to:

$$(1 - \frac{p}{H})D^{t=2}(H; A^*(L)) > (1 - \frac{p}{L})D^{t=2}(H; \bar{A}_L). \tag{49}$$

We can easily see that $D^{t=2}(H; A^*(L)) > D^{t=2}(H; \bar{A}_L)$, because $A^*(L) < \bar{A}_L$. It is clear that $1 - \frac{p}{H} > 1 - \frac{p}{L}$. Therefore inequality (49) follows. ■

Claim 2. *Every $A_H \in [\underline{A}_H, \underline{A}_L)$ fails to satisfy the C-K Intuitive Criterion.*

Proof. For any $A_H \in [\underline{A}_H, \underline{A}_L)$, take $A' = A_H + \epsilon \in (A_H, \underline{A}_L)$, and observe the following two inequalities:

$$\pi(A^*(L), L, L) \geq \pi(A', L, q^e), \forall q^e = L, H, \tag{50}$$

$$\pi(A_H, H, H) < \pi(A', H, H), \tag{51}$$

since $A_H < A' < A^*(H)$. Therefore, no $A_H \in [\underline{A}_H, \underline{A}_L)$ satisfies the C-K Intuitive Criterion. ■

Claim 3. *$A_H = \underline{A}_L$ satisfies the C-K Intuitive Criterion.*

Proof. If we take any $A' < A_H$, then A' satisfies (CK-i) given by (38) but does not satisfy (CK-ii) given by (39), because $A^*(H) > A_H > A'$. If we take any $A' \in (\underline{A}_L, \bar{A}_L)$, then A' does not satisfy (CK-i) by the definitions of $\underline{A}_L$ and $\bar{A}_L$. Finally, if we take any $A' \geq \bar{A}_L$, then A' satisfies (CK-i) but does not satisfy (CK-ii), because $\Delta\Pi(A, H)$ is strictly decreasing in A. ■

References

[1] Bagwell, K.and M. Riordan, 1991, High and Declining Prices Signal Product Quality, American Economic Review 81, 224-239.

[2] Balough, C.D., 2012, A Survey of False Advertising in Cyberspace The Business Lawyer 68, 297-304.

[3] Berg, N., J.-Y. Kim and I. Seon, 2021, A Performance-Based Payment: Signaling the Quality of a Credence Good, Managerial and Decision Economics, 42, 1117-1131.

[4] Chang, J.-W., 2022, KFTC and KCA Investigate Shill Advertisements, Dailian Newspaper, February 2. https://www.dailian.co.kr/news/view/1079170.

[5] Chen, Y., N. Kartik and J. Sobel, 2008, Selecting Cheap-Talk Equilibria, Econometrica 76, 117-136.

[6] Chen, Y., R. Zhang and B. Liu , 2019, Pricing Decisions on Reward-Based Crowdfunding with Bayesian Review System Facing Strategic Consumers, Discrete Dynamics in Nature and Society 2019, 1-14.

[7] Chevalier, J. A., and D. Mayzlin, 2006, The Effect of Word of Mouth on Sales: Online Book Reviews, Journal of Marketing Research, 43, 345-354.

[8] Chintagunta, P.K., Gopinath, S. and Venkataraman, S., 2010, The Effects of Online User Reviews on Movie Box Office Performance: Accounting for Sequential Rollout and Aggregation Across Local Markets. Marketing Science 29, 944-957.

[9] Cho, I. and D. Kreps, 1987, Signaling Games and Stable Equilibria, Quarterly Journal of Economics 102, 179-221.

[10] Crawford, V. and J. Sobel, 1982, Strategic Information Transmission, Econometrica 50, 1431-1451.

[11] Darby, M. and E. Karni, 1973, Free Competition and the Optimal Amount of Fraud, Journal of Law and Economics, 16, 67–88.

[12] Daughety, A. and J. Reinganum, 1995, Product Safety: Liability, R&D, and Signaling, American Economic Review 85, 1187-1206.

[13] Ellison, G. and D. Fudenberg, 1995, Word-of-Mouth Communication and Social Learning, Quarterly Journal of Economics 110, 93-125.

[14] Evans, P., 2016, How celebrities like the Kardashians are bending the advertising rules in the social media age. Canadian Broadcasting Corporation, Nov 11, https://www.cbc.ca/news/business/marketplace-celebrity-endorsements-1.3841922.

[15] Fradkin, A., E. Grewal, and D. Holtz, 2021, Reciprocity and Unveiling in Two-sided Reputation Systems: Evidence from an Experiment on Airbnb, Mimeo.

[16] Grether, D., D. Porter and M. Shum, 2015, Cyber-Shilling in Automobile Auctions: Evidence from a Field Experiment. American Economic Journal: Microeconomics 7, 85-103.

[17] Grossman, S., 1980, The Role of Warranties and Private Disclosure about Product Quality, Journal of Law and Economics 24, 461-483.

[18] He, S., B. Hollenbeck and D. Proserpio, 2022, The Market for Fake Reviews, Marketing Science 41, 871-1027.

[19] Hertzendorf, M., 1993, I'm Not A High-Quality Firm – But I Play One on TV, RAND Journal of Economics 24, 236-247.

[20] Hung, K. and S. Li, 2007, The Influence of eWOM on Virtual Consumer Communities: Social Capital, Consumer Learning, and Behavioral Outcomes, Journal of Advertising Research 47, 485-495.

[21] Jin, C., L. Yang, K. Hosanagar, 2022, To Brush or Not to Brush: Product Rankings, Consumer Search, and Fake Orders. Information Systems Research. 1-21.

[22] Judd, K. and M. Riordan, 1994, Price and Quality in a New Product Monopoly, Review of Economic Studies 61, 773-789.

[23] Kartik, N., 2009, Strategic Communication with Lying Costs, Review of Economic Studies 76, 1359-1395.

[24] Kihlstrom, R. E. and M. H. Riordan, 1984, Advertising as a Signal, Journal of Political Economy 92, 427-450.

[25] Kim, J.-Y. and W. Xu, 2022, Endogenous Word-of-Mouth Communication: A Signaling Theory of the Consumer Review Policy, Mimeo.

[26] Korea Fair Trade Committee, 2023, The Result of Monitoring SNS Shill Advertising, Press Release, February 6.

[27] Laskowski, A. (2022) Kim Kardashian, Sponcon, and the Rules of Being an Influencer. BU Today https://www.bu.edu/articles/2022/kim-kardashian-sponsored-content-influencer-rules/.

[28] Li, L., S. Tadelis, and X. Zhou, 2020, Buying Reputation as a Signal of Quality: Evidence from an Online Marketplace, Rand Journal of Economics 51, 965-988.

[29] Liu, X., 2017, The Signaling Study of Advance Selling Considering the Product Diffusion and Online Reviews Effect, Journal of Service Science and Management 10, 72-86.

[30] Liu, Y., 2006, Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue, Journal of Marketing 70, 74-89.

[31] Luca, M., G. Zervas, 2016, Fake it Till You Make it: Reputation, Competition, and Yelp Review Fraud. Management Science, 62(12), 3412-3427.

[32] Milgrom, P. and J. Roberts, 1986, Price and Advertising Signals of Product Quality, Journal of Political Economy 94, 796-821.

[33] Nelson, P., 1970, Information and Consumer Behavior, Journal of Political Economy 78, 311-329.

[34] Nelson, P., 1974, Advertising as Information, Journal of Political Economy 81, 729-754.

[35] New York Times, 2004, Amazon Glitch Unmasks War of Reviewers, February 14, Al.

[36] Office of the New York State Attorney General (2009) Attorney General Cuomo Secures Settlement With Plastic Surgery Franchise That Flooded Internet With False Positive Reviews (Press release). https://ag.ny.gov/press-release/2009/attorney-general-cuomo-secures-settlement-plastic-surgery-franchise-flooded.

[37] Roettgers, J., 2016, Kardashians in Trouble Over Paid Product Endorsements on Instagram. Variety, August 22, https://variety.com/2016/digital/news/kardashians-instagram-paid-ads-product-placements-1201842072/.

[38] Satterthwaite, M., 1979, Consumer Information, Equilibrium Industry Price, and the Number of Sellers, Bell Journal of Economics 10, 483-502.

[39] Schmalensee, R., 1978, A Model of Advertising and Product Quality, Journal of Political Economy 86, 485-503.

[40] Schuckert, M., X. Liu, R. Law, 2016, Insights into suspicious online ratings: direct evidence from TripAdvisor. Asia Pacific Journal of Tourism Research, 21, 259-272.

[41] Shifferaw, A. (2019, May 04) Legal Documents Reveal Kim Kardashian Makes $300K$ to $500K$ Per Instagram Post. Complex Networks, https://www.complex.com/pop-culture/2019/05/legal-documents-reveal-kim-kardashian-makes-300k-to-500k-per-instagram-post-ad.

[42] Thompson, D., P. Malaviya, 2013, Consumer-Generated Ads: Does Awareness of Advertising Co-Creation Help or Hurt Persuasion?, Journal of Marketing 77, 33-47.

[43] Tirole, J., 1988, The Theory of Industrial Organization, MIT Press.

[44] Whitworth, D., 2010, Man fined over fake eBay auctions, July 5. BBC https://www.bbc.co.uk/news/newsbeat-10508913.

[45] Wilkinson, J., 2013, Court orders shill bidding car dealer to pay back deposit. New Zealand Herald, Dec 1, https://www.nzherald.co.nz/nz/court-orders-shill-bidding-car-dealer-to-pay-back-deposit/OOMIBMI4YFB2FFU3HCVFKTA4I4/.

[46] Wigmore, I. (2013) Internet shill. https://www.techtarget.com/whatis/definition/Internet-shill.

[47] Wolinsky, A., 1983, Price as Signals of Product Quality, Review of Economic Studies 50, 647-658.

[48] Xu, H., D. Liu, H. Wang, and A. Stavrou, 2015, E-commerce Reputation Manipulation: The Emergence of Reputation-Escalation-as-a-Service, Proceedings of 24th World Wide Web Conference (WWW 2015), 1296–1306.

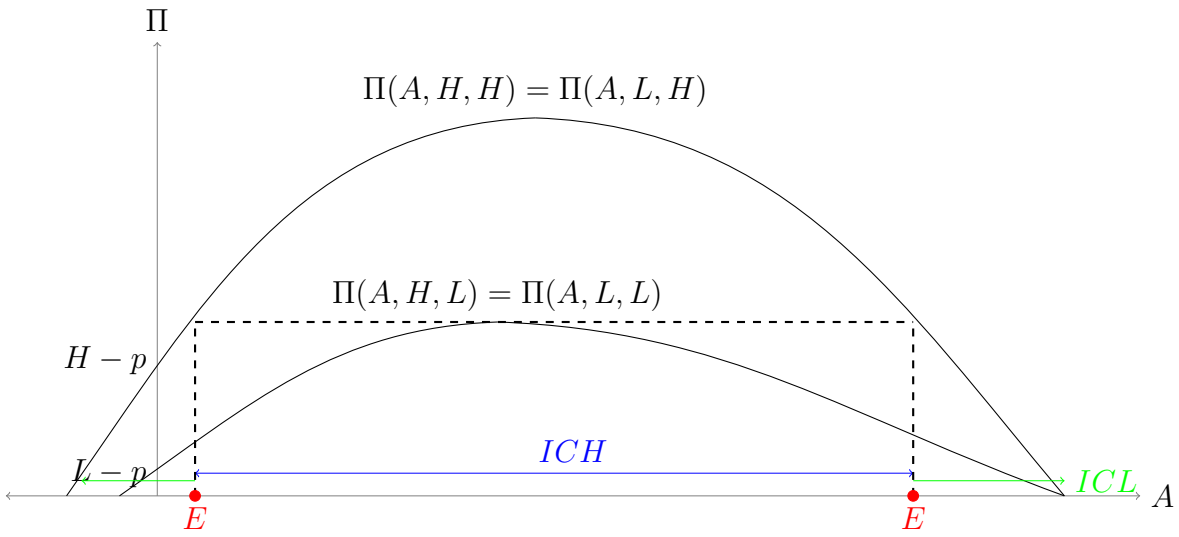

Figure 1. Two Separating Equilibria

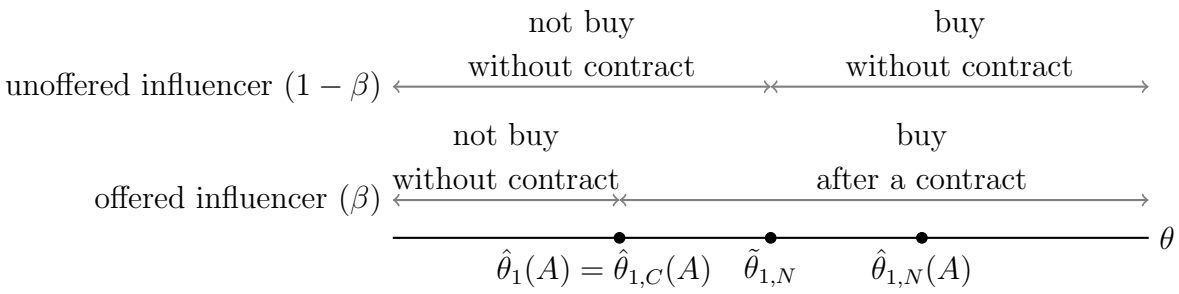

Figure 2. Purchasing Decision of an Influencer

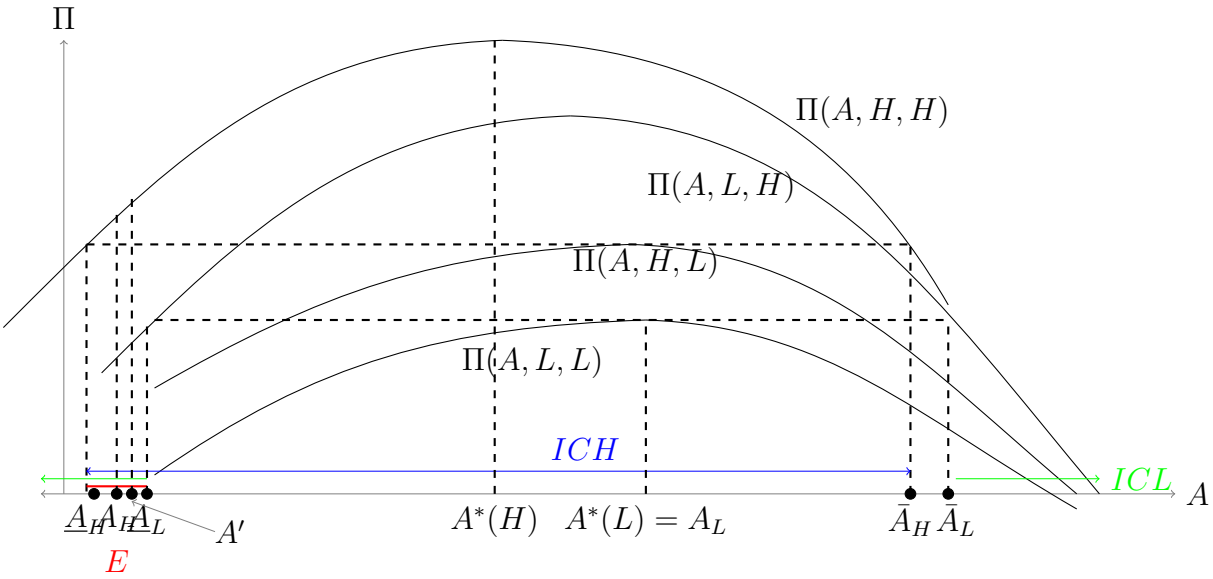

Figure 3. Unique Separating Equilibrium that passes the Intuitive Criterion

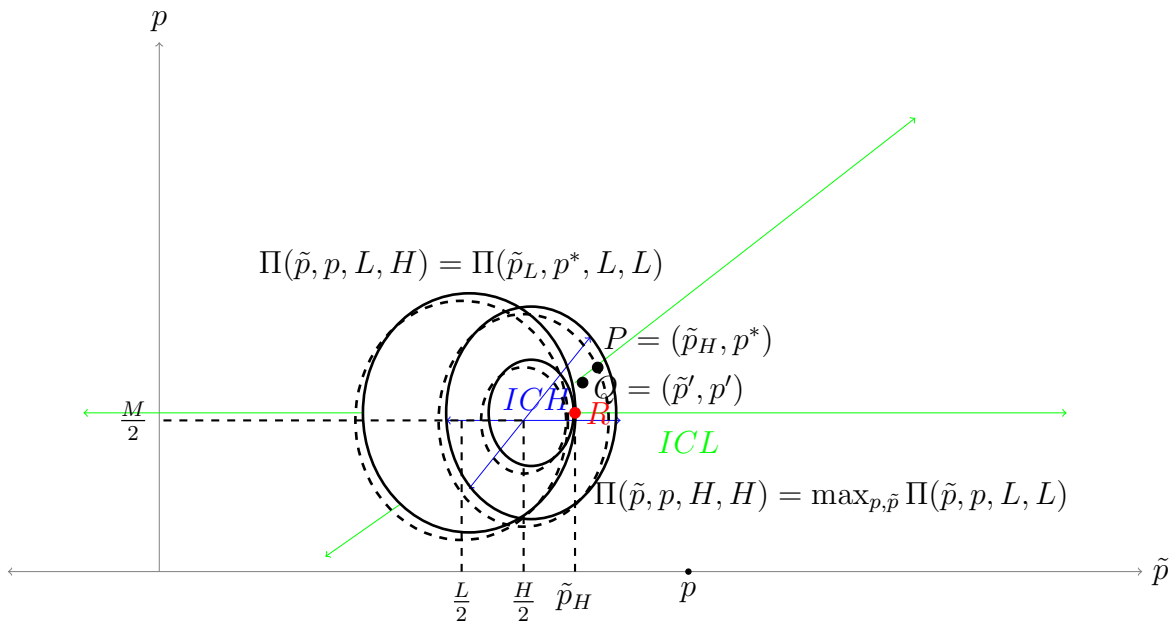

Figure 4. Price and Advertising as Signals