Abstract

We develop a model of communication messenger in advertising where advertising can generate product-match signals for consumers. We consider advertising as a problem of Bayesian persuasion with costly information processing, where the type of communication messenger is costless to observe and determines the information structure consumers will face, thereby affecting their attention decisions. We define the type of a messenger (either high-type or low-type) according to its likelihood of generating good signals about product match, and show that the optimal choice of messengers depends on messengers’ signal elasticities and the firm’s decision on whether to induce consumer attention. We then identify a condition for a low-type messenger to be a more effective medium. Especially when the importance of raising the price is high, and a high-type messenger ad overshadows the product value by providing generally good signals, a low-type messenger can better draw consumer attention and persuade them to pay a higher premium. This general insight can still carry over even if a high-type messenger can better grab consumers’ attention by providing additional entertaining value or when some consumers are naive in updating their beliefs.

Keywords: advertising content, messenger, dual-mode of communication, deliberation cost, Bayesian persuasion
1 Introduction

“We’ve said all along that the messenger can be as or more important than the actual message itself.”

– Lisa Sherman, the chief executive, Ad Council

For the past forty years, economists and marketers have studied how advertising can help consumers learn about products or convince them to choose over the competition. The success of an advertising campaign crucially depends on whether its content can clearly communicate the product benefits to the consumers. Also, it is well-known that it is not just what is said (message) that matters, but who says it (messenger) matters too.\(^1\) The messenger of communication plays a vital role in attracting the receiver’s attention and convincing the message, and it can influence the effectiveness of communication. Thus, how to deploy an effective messenger strategy in different contexts and whom to use to deliver the message is of great interest to both academics and practitioners.

The academic literature finds some evidence of the positive effects of using a specific type of messenger, focusing on what messenger traits account for the effectiveness of advertising campaigns. Many studies on the attractiveness model (Eisend and Langner, 2010; Joseph, 1982; Liu et al., 2007) documented that physically attractive messengers positively affect consumers’ attitudes, product evaluation, and purchase intention. Also, a stream of literature (Ohanian, 1991; Schouten et al., 2020; Till and Busler, 1998) suggests that the messenger’s perceived relevance to the advertised product is essential in affecting purchase behavior. Thus, using attractive or more relevant messengers seems appropriate to increase advertising effectiveness.

The industry practices are mixed. Traditionally, many advertising campaigns tend to have endorsements from physically attractive actors or celebrities. However, not all marketers have agreed that the idealized images of such endorsers are the most effective in advertising. In 2004, Unilever developed several Real Beauty campaigns that feature real women, not models.\(^2\) Although the positive results of those campaigns may be due to their provocative nature of novelty (Vézina and Paul, 1997), advertising industries have witnessed the rise of ads featuring real people for their products, notwithstanding the lack of novelty it used to garner in the early days. For example, Schick’s new digital campaign, “Be You. No One Else Can,” features real people, not models.

\(^1\)https://ssir.org/articles/entry/finding_the_right_messenger_for_your_message
Also, Abercrombie & Fitch launched a “Face Your Fierce” campaign in 2020 that showcased a wide spectrum of real people. Those recent ads would not have created the same level of novelty or provocation as the first ones. On the other hand, despite the rise of such realistic ads, almost 20% of all television ads still feature a physically attractive celebrity or famous person. This leaves an important question for businesses: are celebrities or real people, or what types of messenger are more effective for an advertising campaign? A deeper understanding of the mechanism by which messenger leads to persuasion is critical to the success of an advertising campaign. In this study, we seek to understand advertising messenger’s effectiveness by examining the advertising persuasion process. To what extent can the messengers play a role? How do different messengers, such as nonidealized vs. idealized messengers, affect consumers’ attention and purchase decisions? Specifically, we analyze how messengers can lead to persuasion and identify the critical variables associated with their effectiveness.

We take a dual-mode of communication perspective for advertising communication. A messenger serves as a cue communicated through a less costly peripheral route, while the central issue-relevant information requires thoughtful consideration. Consumers can costlessly observe the advertising messenger. However, they can receive a binary signal about their match value (either a good or bad signal) only if they incur a deliberation cost to pay attention to the ad. We first analyze the consumer inference process upon receiving a private signal from an ad featuring different types of messengers. The type of messenger does not change the message the ad delivers but determines the informativeness of consumers’ signal structures. Also, consumers update their beliefs in a Bayesian fashion upon receiving a signal from the ad. For example, an attractive messenger may make a product look more appealing while failing to convince the audience of their product match values, because consumers may be confused about whether the positive appeal is due to the messenger or the product’s true underlying value.

To capture this intricate relationship between the advertising messenger and persuasion, we consider advertising as a problem of Bayesian persuasion, where the type of messenger determines the information structure the consumer will face. More specifically, because the product-match values are ex-ante unknown to all parties and the firm has no additional private information in

---

3https://www.barolin-spencer.com/use-celebrity-real-person-ad/
our model, the firm can commit to a particular signal structure by choosing the type of messenger, making advertising a Bayesian persuasion device. We then investigate the mechanism of how a messenger can attract consumer attention and potentially persuade them to pay a higher price for the product. We characterize the conditions when different signals (either a good or bad signal) from different types of messengers (a low and a high-type messenger based on her ex-ante probability of generating a good signal) can be more valuable. Generally, a high-type (low-type) messenger’s bad (good) signal can be informative because they are less likely to generate that signal. However, a high-type (low-type) messenger’s good (bad) signal can also be useful when its signal is sufficiently sensitive to (and thus indicative of) changes in match values.

We then compare the expected profits to determine the firm’s optimal decision on whether to induce consumer attention or not. When deciding whether or not to induce consumer attention, the firm has to consider both the need and the ability to boost the price, jointly optimizing sales volume versus price premium. We find that as the product cost increases, the need to increase the price becomes higher, and thus, the firm tends to sacrifice demand to increase its price. On the other hand, when the consumer’s deliberation cost increases, the firm’s ability to charge a higher price under no consumer attention increases, enticing the firm to sacrifice profit margin to retain greater demand. Thus, it is optimal to induce consumer attention as the consumer’s deliberation cost becomes smaller or the product cost becomes larger.

After establishing the relationship between the messenger and consumer attention, we show that the optimal choice of messenger depends on each messenger’s signal elasticity (or informativeness) and the firm’s decision on whether to induce consumer attention. When a high-type messenger is sufficiently informative, a high-type messenger ad is more effective in attracting consumer attention and increasing the price. However, because a high-type messenger’s signals are so informative, the firm must set the price very low to dissuade consumer attention. So, the firm chooses a low-type messenger if it opts not to induce consumer attention, which is the case when the product cost is low, or deliberation cost is high. On the other hand, when a high-type messenger becomes less informative, the ad featuring the high-type messenger tends to generate more good signals, which can overshadow the product and thereby providing limited ability to update consumers’ beliefs about the product match. In contrast, as it is rare, the low-type messenger’s good signals
can significantly update consumers’ beliefs to cover its high cost. Therefore, when the low-type messenger’s signals are sufficiently informative and the production cost is sufficiently high, the firm finds it optimal to draw consumer attention and, more importantly, persuade them to pay a higher premium. This general insight can still carry over even if a high-type messenger can better grab consumers’ attention by providing additional entertaining value or when some consumers are naive in updating their beliefs. Even though the range for the low-type to be optimal significantly shrinks once we incorporate such effects, a low-type messenger can still be a more effective medium for advertisers, especially if the importance of raising the price is high, but a high-type messenger is less informative and overshadows the product characteristics.

We know ad content matters, but how it matters is not clearly understood. In this research, we propose one possible mechanism to demonstrate how ad content, especially the messenger, matters in persuading consumers. Therefore, this paper’s main contribution is providing a new framework to consider advertising – we posit that ad content can be thought of as information structure and leverage Bayesian persuasion as a tool to model advertising content and the role of the messenger. The paper is organized as follows. The next section reviews the related literature. Section 3 describe a model that characterizes the role of communication messengers in advertising. Then, we first present the main analyses and results focusing only on the persuasion role of the messenger in Section 4. In Section 5, we demonstrate the robustness of the main insights by analyzing a number of extensions that relax the several key model assumptions. Section 6 concludes.

2 Literature Review

This paper relates to several streams of research: advertising content, persuasion, and information design. First, our research contributes to the burgeoning area of advertising content. The content of advertising provides direct information, such as the existence of the product or its price (Butters 1977; Iyer et al. 2005; Shin 2005). The information can also be indirect, where the mere fact that the firm advertises signals an experience good’s quality (see Nelson 1974 and Milgrom and Roberts 1986). The latter is known as the “money-burning” theory of advertising. The central argument of the money-burning theory suggests that the level of spending, not the advertising content, signals the quality of the product. That is, content is irrelevant for communicating information about
product quality. However, this view has been challenged by a number of recent papers, which explore the role of advertising content in conveying information in a rational equilibrium framework (Anderson and Renault 2006; Mayzlin and Shin 2011). Anderson and Renault (2006) study the optimal amount of information in advertising content, allowing firms to provide both price and match information. Mayzlin and Shin (2011) show how providing product attributes information, along with price, can signal product quality when the bandwidth of advertising messages is limited and consumers can conduct their own search on product information. Our paper contributes to this literature by analyzing the role of the messenger in advertising content.

For the role of advertising messenger, a large body of research finds several key characteristics of messengers that impact the effectiveness of persuasion, such as attractiveness, relevance, credibility, and familiarity (Chaiken and Maheswaran, 1994; Till and Busler, 1998; Zajonc, 1968). For example, Till and Busler (1998) found that when there was a match between the messenger and the endorsed product (e.g., a baseball player endorsing an energy bar), consumers showed greater purchase intent than when there was a perceived mismatch between the messenger and the product. These researches suggest that having the fit between the messengers and the endorsed products tends to make their audience assume they will make valid arguments, and therefore is predisposed to believe what they say and like what they are selling (Chaiken and Maheswaran, 1994). Building on these findings, we assume two types of messengers: a high-type (such as highly relevant or attractive) messenger and a low-type one. We then analyze how these two types of messenger influence the persuasiveness of advertising communication through consumer inference.

We also take the dual mode of communication perspective for persuasion in the psychology literature (Chaiken and Maheswaran, 1994; Kahneman, 2011; Petty and Cacioppo, 1986), which suggests that the decision-makers process information through two different routes (the central and the peripheral route) with different amounts of cognitive loads for processing information. The

---

4 A stream of consumer psychology research has identified different elements of advertising content influence consumer’s attitudes and actions, such as fear (Johnston et al., 2015) and humor (Wöltman Elpers et al., 2004). For celebrity endorsement contexts, Dhar and Stillman (2019) suggest that likability and relevance (or fit) are two critical characteristics of messenger that drives persuasion.

5 A recent paper by Shin and Yu (2021) also examines how advertising affects consumer inference. While we investigate the role of the messenger for consumer inference, they focus on how the mere fact that being targeted by advertisements can affect consumer inference about product match and their subsequent search behaviors.

6 Dewatripont and Tirole (2005) also model the processing of both issue-relevant information and cues. In their paper, the cue is the degree of preference alignment between a sender and a receiver, and the receiver either successfully assimilates the issue-relevant information, or she fails to do so. In our model, the cue, which is the type of
central route involves issue-relevant information that is more cognitively demanding to process and can occur only when one is willing and able to devote substantial mental resources to the message (Petty and Cacioppo, 1986). On the other hand, the peripheral route often entails issue-irrelevant cues and requires little to no cognitive effort from the decision-makers. Several papers study the implications of consumers’ costly information processing or deliberation costs on firms’ decisions, such as pricing (Wathieu and Bertini, 2007), quality signaling (Gardete and Guo, 2021; Guo and Wu, 2016), product line decision (Guo and Zhang, 2012; Kuksov and Villas-Boas, 2010), and optimal information disclosure (Lu and Shin, 2018). For example, in advertising contexts, Dukes et al. (2021) and Lin (2022) investigate the optimal design of media advertising format by focusing on consumers’ incentives to pay attention to the ads in the presence of costly attention (opportunity) costs. We contribute to this line of literature by incorporating the dual-mode of communication in advertising contexts: consumers can observe the advertising messenger costlessly through the peripheral route, while an issue-related product-match signal requires costly deliberation through the central route.

Finally, the problem we study is related to the Bayesian persuasion and information design literature (Bergemann and Morris, 2019; Gentzkow and Kamenica, 2016, 2017; Guo, 2022; Iyer and Zhong, 2022), which seeks to identify the optimal information environment to affect the receiver’s decisions through influencing the posterior beliefs about the state of the world. The pioneering paper by Kamenica and Gentzkow (2011) considers a model with symmetric information where a sender can only affect a receiver’s action by choosing and committing to a particular information structure. In recent papers, Iyer and Zhong (2022) studies a firm’s optimal information notification design to maximize consumer engagement using an information design framework. Guo (2022) also applies the Bayesian persuasion framework in the collaborative customization setting, focusing on the effect of the choice of information structure on customers’ engagement decisions in the customization process. We consider advertising a Bayesian persuasion device: the firm can commit to a specific information structure by choosing the messenger whose characteristics, such as relevance, can affect the information structures that generate the signal about the match value.

---

6 The cognitive process occurring through peripheral and central routes is also referred to as System 1 and System 2 thinking in the literature (Kahneman, 2011).
3 Model

3.1 Strategic players and information environment

We consider a market with a monopolistic firm that sells a single product with a constant production cost $k$ to a unit mass of consumers. Consumers are unaware of the product and can buy it only if they receive an ad. Each consumer has unit demand, and consumer $i$ can obtain a utility of $v_i$ by consuming the product, where $v_i$ is the individual-specific product match value drawn from a distribution $F[0, \bar{v}]$ with density $f(\cdot)$. We assume the information is symmetric: both the probability distribution of product match $F$ and the production cost $k$ are common knowledge.

Upon deliberation, advertising yields a binary product-match signal $s_i \in \{s_g, s_b\}$, where $s_g$ denotes good news, and $s_b$ denotes bad news about the match. In this setting, because the firm has no additional private information about the product match value for an individual consumer when it advertises the product to consumers, we can consider the role of the advertisement as a Bayesian persuasion device (Kamenica and Gentzkow 2011). That is, consumers update their beliefs in a Bayesian fashion upon seeing the signal from the ad instead of updating beliefs based on the firm’s equilibrium advertising strategies. Next, we focus on the characteristics of messengers, which can influence the realization of signals and consumer inference.

The role of communication messengers

Who delivers the message is important in attracting attention and convincing the receiver about the product match. Specifically, some characteristics of the messenger, such as attractiveness, credibility, and expertise, can influence the effectiveness of an ad (Dhar and Stillman, 2019; Till and Busler, 2000). We postulate that a messenger’s type (either a high-type $m_H$ or a low-type $m_L$) can probabilistically affect the realization of the signal that consumers obtain from the ad. For example, many beauty industry firms tend to feature good-looking celebrities to affect consumers’ attitudes toward the brand by borrowing the celebrities’ personal images or attractiveness. Formally, given the true product match value $v$ and the messenger $m$, consumers obtain private signals upon

---

*In practice, consumers perceive the product either as a good match (a good signal) or not a good fit for their needs (a bad signal) after deliberation. Those impressions are signals that consumers receive.*
paying attention to the ad, with the following probability:

\[ \sigma_j(v) \equiv \Pr(s_i = s_g|v, m_j), \text{ where } j \in \{L, H\}. \]  \hspace{1cm} (1)

The firm knows that the consumer can obtain this signal \( s_i \) with the above probabilities but does not observe what signal a consumer ultimately receives. Still, the signal structure consumers face is affected by the messenger’s type as follows:

**Assumption.** The probability of receiving a good signal \( \sigma_j(v) \) is increasing in the true match value \( v \) and the messenger’s type \( m_j \):

\[ \frac{\partial \sigma_j}{\partial v} \geq 0, \text{ and } \sigma_H(v) > \sigma_L(v) \text{ for all } v. \]  \hspace{1cm} (2)

Equation (2) implies that a higher (true) match value will more likely generate a good signal. Therefore, a positive news \( s_g \) is really “good news” regarding the true match value (Milgrom, 1981). Also, a high-type messenger is more likely to generate a good signal than a low-type messenger for any level of true value \( v \). Let us define the average probability of receiving a good signal from a particular type of messenger as

\[ \bar{\sigma}_H \equiv \Pr(s_i = s_g|m_H) = \int_{0}^{\bar{v}} \sigma_H(v) dF, \hspace{0.5cm} \bar{\sigma}_L \equiv \Pr(s_i = s_g|m_L) = \int_{0}^{\bar{v}} \sigma_L(v) dF. \]  \hspace{1cm} (3)

Then, \( \bar{\sigma}_H > \bar{\sigma}_L \). We use these two terms \( \bar{\sigma}_H, \bar{\sigma}_L \) in our subsequent analysis.

We also note that a messenger’s type may determine how easily the messenger can command attention and motivate people to think about the information in the advertising message. In particular, high-type messengers can lower the deliberation cost consumers incur to pay attention to the ad. For example, the firm can feature a celebrity to provide consumers additional entertaining value to offset the cost.\(^9\) Formally, the deliberation cost can be a function of a messenger’s type, \( c = c(m) \), where \( 0 \leq c(m_H) < c(m_L) \). However, we first focus on the relationship between a messenger’s type and the signal structure in the main model by fixing the deliberation cost constant for both types of messengers \( c(m_H) = c(m_L) = c \) to convey the insights cleanly. We later relax this

---

\(^9\)Other ways to manipulate the deliberation cost include adjusting the ad’s length, changing the content’s readability (format), and so on. We abstract away those features in ad content to focus on the role of messenger.
assumption by incorporating the effect of a messenger’s type on the deliberation cost to capture the messenger’s attention-grabbing role.

**Dual model of communication**

Consumers need different cognitive loads to process different types of information in advertising. We take a dual-mode of communication perspective, where the information is conveyed through two different routes (Petty and Cacioppo, 1986). First, there is a less costly peripheral route delivering issue-irrelevant cues such as the expertise or attractiveness of the messenger (Petty and Cacioppo, 1986). Because messenger’s attractiveness might be the first thing consumers notice even before they start thinking about the product information, such ostensible information about the messenger can be communicated through mindless processing (Kahneman, 2011; Petty and Cacioppo, 1986). On the other hand, the more costly central route of persuasion requires thoughtful consideration of advertising contents, such as the personal relevance to a product or the product’s match value (Guo, 2016; Guo and Zhang, 2012; Wathieu and Bertini, 2007). Therefore, the detailed information about the actual product would only be examined when consumers pay close attention to the ad.

Here, we assume that consumers can costlessly observe the advertising messenger. In contrast, they can receive an issue-relevant signal through the central route of communication only if they decide to pay attention to the ad by incurring costly deliberation costs, \( c > 0 \).\(^{10}\) For example, when an ad appears on TV, consumers can mindlessly see an ad delivered to their eyes and recognize the appearance of a messenger in it. Then, she can decide whether to pay attention by incurring deliberation costs to process the central information in the ad, thereby receiving a private signal about the product match.

**3.2 Sequence of the game.**

The firm’s action space consists of its choice of the product price, \( p \), and the type of advertising messenger, \( m \in \{m_L, m_H\} \). Consumers’ strategies are deciding (1) whether to pay attention to the advertising message after observing the messenger, and (2) whether to buy a product.

\(^{10}\)We impose the assumption that \( c < \tau \equiv \min_{j \in \{L, H\}} \int_0^v \sigma_j - \sigma_j'(v) \, dF \), so the firm can set prices to induce consumers’ attention for any given type of messenger. Otherwise, the deliberation cost is too high and it might never be optimal for consumers to pay attention to the ad regardless of the price. Thus, the problem becomes trivial.
Nature draws product match values \( v_i \)'s.

Stage 0: Nature draws each consumer \( i \)'s product match value \( v_i \).

Stage 1: Firm \( j \) decides (1) the price \( p_j \) and (2) the messenger \( m_j \in \{m_L, m_H\} \).

Stage 2: Consumers decide whether to pay attention to the ad upon seeing the messenger \( m_j \) and the price \( p_j \). If so, a consumer incurs the deliberation cost \( c \) and receives a signal \( s_i \in \{s_g, s_b\} \).

Stage 3: Consumers decide whether to purchase the product based on the price and their updated beliefs about the product match value.

The timeline of the game goes as follows. In stage 0, Nature draws each consumer \( i \)'s product match value \( v_i \). In stage 1, the firm \( j \) chooses the type of messenger \( m_j \in \{m_L, m_H\} \) and a price \( p_j \) to maximize its expected profit. In stage 2, consumers decide whether to pay attention to the ad upon seeing the messenger \( m_j \) and the price \( p_j \) chosen by the firm.\(^{11}\) If a consumer decides to pay attention, she incurs a deliberation cost \( c \) and receives a noisy private signal \( s_i \in \{s_g, s_b\} \) about the product match. The firm knows that the consumer will receive a private good signal about the product match with probabilities \( \sigma_j(v) \) but does not observe whether the consumer actually pays attention to the advertising by incurring a deliberation cost, let alone what signal the consumer ultimately receives if she chooses to do so. Finally, consumers decide whether to purchase the product based on the price and their updated beliefs about the product match value in stage 3. We summarize the game sequence in Figure 1.

4 Analysis

When the firm evaluates the effectiveness of a messenger, it compares the expected profits from different types of messengers, which are affected by (1) the signal structure generated by a messenger, (2) the messenger’s ability to attract consumer attention, and (3) the price premium it can command based on updated beliefs of consumers after deliberating the ad. We start by analyzing consumers’ inference process upon receiving a private signal. Then, we study the mechanism of how a messenger can attract consumer attention and potentially raise price premiums. In our main model, we fix the messenger’s attraction ability, which directly affects the consumer’s deliberation cost. Nevertheless, as we show in 4.2, the messenger’s type can still affect the consumers’ attention.

\(^{11}\)In a model extension in Section 5.3, we consider a case where price is ex-ante unobservable and can be only known after they pay attention. We show that the main insights carry over.
decision rationally through influencing the consumers’ inference. In our extension, we consider the
direct attention-grabbing effect by incorporating the messenger’s differential attraction ability. Af-
after analyzing consumers’ inference processes and attention decisions, we turn to the firm’s decisions
about pricing and optimal choice of messenger in equilibrium. Our solution concept is Subgame
Perfect Equilibrium, and all proofs can be found in the appendix.

4.1 Consumer inference

Consumers update their beliefs in a Bayesian fashion upon receiving a signal from the ad (conditional
on paying attention to it). Formally, upon receiving a private signal $s_i$, a consumer’s posterior beliefs
about the underlying match value distribution follow the next two densities\textsuperscript{12}:

$$f(v|m_j, s_g) = \frac{f(v) \cdot \sigma_j(v)}{Pr(s_i = s_g|m_j)} = \frac{f(v) \cdot \sigma_j(v)}{\sigma_j}, \quad (4)$$

$$f(v|m_j, s_b) = \frac{f(v) \cdot [1 - \sigma_j(v)]}{Pr(s_i = s_b|m_j)} = \frac{f(v) \cdot [1 - \sigma_j(v)]}{1 - \sigma_j}. \quad (5)$$

To compare the informativeness of the signal from different messengers, we first establish the
following proposition, which connects consumers’ posterior beliefs to the different signal structures
generated by different messengers.

**Proposition 1.** Upon receiving a good signal $s_g$, the posterior belief generated from the high-type
messenger $m_H$ satisfies the MLRP with respect to the low-type messenger $m_L$ (that is, $\frac{f(v|m_H, s_g)}{f(v|m_L, s_g)}$ is
increasing in $v$) if and only if, for all $v$,

$$\varepsilon_H(v) \equiv \frac{d\sigma_H(v)}{d\sigma_H(v)/v} \geq \frac{d\sigma_L(v)}{d\sigma_L(v)/v} \equiv \varepsilon_L(v). \quad (6)$$

Otherwise (i.e., $\varepsilon_H(v) < \varepsilon_L(v)$ for all $v$), the posterior belief from the low-type messenger $m_L$
satisfies the MLRP with respect to the high-type messenger $m_H$: $\frac{f(v|m_H, s_g)}{f(v|m_L, s_g)}$ is decreasing in $v$.

The elasticities $\varepsilon_H(v)$ and $\varepsilon_L(v)$ measure how sensitive a messenger’s good signal is to changes
in match values, and their relationship determines the informational environments that consumers

\textsuperscript{12}Note that the posterior beliefs are Bayes plausible as defined in (Kamenica and Gentzkow, 2011). That is, the
expected posterior belief equals the prior belief: $\tilde{\sigma}_j f(v|m_j, s_g) + (1 - \tilde{\sigma}_j) f(v|m_j, s_b) = f(v) \sigma_j(v) + f(v)[1 - \sigma_j(v)] = f(v)$. 11
Figure 2: Two Different Signal Structures

Face. Figure 2 demonstrates two possible cases suggested by the lemma. It is important to note that, as the figures show, both cases are consistent with our assumption in equation (2): (1) the probability of receiving a good signal always increases in \( v \): \( \frac{\partial \sigma}{\partial v} \geq 0 \), and (2) \( \sigma_H(v) > \sigma_L(v) \) for all \( v \). Therefore, receiving a good signal \( s_g \) always improves the posterior belief about the product match value \( f(v|m_j, s_g) \) regardless of messenger’s type. Moreover, one may intuit that a good signal from a low-type messenger may update consumers’ posterior beliefs about the match value more simply because it is less likely to generate a good signal (i.e., \( \sigma_H(v) > \sigma_L(v) \) for all \( v \)). However, this lay intuition fails to incorporate a more nuanced consumer inference process. In fact, the relative effects of messengers’ signals on the posterior beliefs depend on the relationship between those two elasticities of signal structures.

First, when \( \varepsilon_H(v) \geq \varepsilon_L(v) \) (see Figure 2-(a)), the likelihood ratio of generating a good signal between high and low type messengers increases in \( v \), and different messengers’ signals are fairly divergent only when the product match is sufficiently large. In particular, for the low-type messenger ads, the likelihood of generating a good signal is quite similar across all the match values \( v \). Therefore, even though a good signal positively updates the consumer belief about the match value, it is not so informative about the match value of \( v \). In contrast, receiving a good signal \( s_g \) from a high-type messenger is clear “good” news about the potential match value since the probability to generate a good signal \( \sigma_H(v) \) is especially sensitive to the match value \( v \). It would more likely generate a good signal only under high match values. Figure 3-(a), where \( \sigma_L(v) = 0 \) for all \( v \) shows the extreme case of such an information environment, where a signal from the low-type messenger is not informative about the true match value.
On the other hand, when $\varepsilon_H(v) < \varepsilon_L(v)$ (see Figure 2-(b)), the pattern is reversed, and the likelihood ratio of generating a good signal between high and low-type messengers decreases in $v$, so different messengers’ signals are fairly divergent only when the product match is sufficiently small. Under this information environment, the high-type messenger’s signals generally tend to be positive regardless of the match value, while the low-type messenger’s signal is quite sensitive to the true match value $\varepsilon_H(v) < \varepsilon_L(v)$. If a consumer receives a good signal from a low-type messenger, it is highly likely to come from high match values. Thus, receiving a good signal $s_g$ from a low-type messenger is more discriminating and informative than a high-type messenger’s signal to resolve their uncertainties in this information environment. Again, Figure 3-(b), where $\sigma_H(v) = 1$ for all $v$ shows the extreme case of such an information environment, where a signal from the high-type messenger is always positive $(s_g)$ but not informative about the true match value.

The signal structures in practice are somewhere between these two extreme cases in Figure 3. The proposition characterizes those general signal structures. When a particular messenger type’s signals are more sensitive to changes in match values, those elastic signals are more discriminating and thus more helpful for the consumers to resolve their uncertainties about the true match value $v$. Thus, depending on the signal structure generated from the different messengers, the values of information from the same signal outcome (i.e., either a good or bad signal $s_i \in \{s_g, s_b\}$) differ.

### 4.2 Consumer attention

When a consumer receives an ad, she decides whether to pay attention to the ad by costly deliberating about the content. If the additional information from the advertisement won’t change her
decision, there is no value in incurring the deliberation cost. Thus, upon deliberation, a consumer purchases the product only if she receives a good signal $s_g$ and does not purchase a product if she receives a bad signal $s_b$. The expected utility of paying attention is

$$EU(\text{attention}|m_j) = \tilde{\sigma}_j \cdot [E(v|m_j, s_g) - p] - c \quad (7)$$

On the other hand, even when she decides not to pay attention to the advertisement, she would still purchase a product if $E(v) - p \geq 0$. Thus, the expected utility of not paying attention is

$$EU(\text{no attention}) = \max \{0, E(v) - p\} \quad (8)$$

We characterize consumers’ attention decisions in the following lemma.\textsuperscript{13}

**Lemma 1.** 1. When $p \geq E(v)$, a private good signal $s_g$ is useful for consumers’ purchase decisions. Consumers pay attention to the ad if and only if $p \leq E(v|m_j, s_g) - \frac{c}{\sigma_j}$.

2. When $p < E(v)$, a private bad signal $s_b$ is useful for consumers’ purchase decisions. Consumers decide to pay attention to the ad if and only if $p > E(v|m_j, s_b) + \frac{c}{1-\sigma_j}$.

The decision rule is intuitive: a consumer compares the cost and benefit of paying attention (represented by equations (7) and (8)) after costlessly observing the messenger (which determines the signal structure that she will face), where the benefit of paying attention depends on her default decision in the case of no advertisement. The lemma shows how the price can change the information value for different signals ($s_i \in \{s_g, s_b\}$). First, when $E(v) > p$, her default action without paying attention to the ad is to purchase the product. In this case, the marginal benefit of getting the private signal is to prevent unnecessary purchasing if receiving a bad signal $s_b$, which updates her belief sufficiently negatively to avoid a purchase. Thus, consumers decide to pay attention to the ad when the cost of paying attention is smaller than the marginal benefit from receiving a bad signal.

\textsuperscript{13}Mayzlin and Shin (2011) also characterize the consumers’ search decisions and highlight the relationship between the uncertainty and price range that can encourage consumer search. In their model, a consumer can engage in a costly search to acquire an additional signal about the product quality, which is either low, medium, or high ($q \in \{L, M, H\}$). The quality is the firm’s private information, and the additional information comes from an external source, such as word of mouth. Thus, advertising serves as a signaling device. In contrast, we focus on the role of advertising messenger in the same ad, and the firm has no additional private information. Therefore, the advertising serves as a Bayesian persuasion (not signaling) device, where the firm commits to a certain signal structure by choosing the type of messenger.
That is \( c < (1 - \bar{\sigma}_j)[p - E(v|m_j, s_b)] \), which is equivalent to the condition \( p > E(v|m_j, s_b) + \frac{c}{1 - \bar{\sigma}_j} \).

On the other hand, when \( E(v) \leq p \), her default action is not to purchase the product. In this case, the marginal benefit of getting the private signal is to help her buy the product if receiving a good signal \( s_g \), which updates her beliefs sufficiently positively to make a purchase. Thus, consumers pay attention to the advertisement when \( c \leq \bar{\sigma}_j[E(v|m_j, s_g) - p] \), which is again equivalent to \( p \leq E(v|m_j, s_g) - \frac{c}{\bar{\sigma}_j} \).

Overall, consumers pay attention to the ad if \( E(v|m_j, s_b) + \frac{c}{1 - \bar{\sigma}_j} < p < E(v) \) or if \( E(v) \leq p \leq E(v|m_j, s_g) - \frac{c}{\bar{\sigma}_j} \). Given our assumption \( c \leq \tau \equiv \min_{j \in \{L, H\}} \int_0^\infty v[\bar{\sigma}_j - \sigma_j(v)]dF \), we can find that consumers pay attention to the ad if and only if the price is in the moderate range such that

\[
P_j \equiv E(v|m_j, s_b) + \frac{c}{1 - \bar{\sigma}_j} < p \leq \bar{P}_j \equiv E(v|m_j, s_g) - \frac{c}{\bar{\sigma}_j}.
\]

We next study how the messenger’s type \( m_j \in \{m_L, m_H\} \) can influence consumers’ expected utility and their decision processes. Regardless of the realized signal \( s_i \in \{s_g, s_b\} \), a high-type messenger can generate an ex-ante higher expected utility than a low-type messenger if \( EU(attention|m_H) \geq EU(attention|m_L) \iff \bar{\sigma}_H \cdot [E(v|m_H, s_g) - p] - c \geq \bar{\sigma}_L \cdot [E(v|m_L, s_g) - p] - c \).

This is equivalent to

\[
\int_0^\infty (v - p) \cdot [\sigma_H(v) - \sigma_L(v)]dF \geq 0.
\]

The following comparative statics identify the condition for a high-type messenger to generate a higher expected utility of paying attention.

**Lemma 2.** A high-type messenger ad generates a higher expected utility of paying attention than a low-type messenger if the price becomes sufficiently low such that \( p \leq \bar{p} \equiv \frac{\int_0^\infty v[\sigma_H(v) - \sigma_L(v)]dF}{\sigma_H - \sigma_L} \).

The difference in marginal benefit of paying attention between a high-type and a low-type messenger ad comes from two components: the average probability of getting good signals \( \bar{\sigma}_j = \Pr(s_i = s_g|m_j) \) and the net expected utility upon receiving good signals \( E(v|m_j, s_g) - p \). A high-type messenger has a higher chance of generating good signals. So, if \( E(v|m_H, s_g) - p \) is not too small compared to \( E(v|m_L, s_g) - p \), the effect of \( \bar{\sigma}_H \) would dominate such that a high-type messenger
generates a higher expected utility of paying attention. Lemma 2 shows that this holds when the price is lower than some threshold \( \tilde{p} \) that satisfies
\[
\int_{\bar{v}}^{\tilde{v}} (v - \tilde{p}) \cdot [\sigma_H(v) - \sigma_L(v)] \, dF = 0.
\]

The following proposition connects the signal structures from different messengers to the value of information that can influence consumers’ attention decisions.

**Proposition 2.** 1. If \( \varepsilon_H(v) \geq \lambda \varepsilon_L(v) \) for all \( v \) where \( \lambda \equiv \frac{\sigma_L(v)}{\sigma_H(v)} \), we have \( \tilde{p} \geq E(v) \). Then, for a low-type messenger, only the good signal \( s_g \) is useful for consumers’ purchase decisions. On the other hand, for a high-type messenger, both good and bad signals can be useful.

2. If \( \varepsilon_H(v) < \lambda \varepsilon_L(v) \) for all \( v \), we have \( \tilde{p} < E(v) \). Then, for a low-type messenger, both good and bad signals can be useful for consumers’ purchase decisions. On the other hand, for a high-type messenger, only the bad signal \( s_b \) is useful.

The proposition suggests that the values of good/bad signals are determined depending on the signal structures generated from messengers. Figure 4 illustrates those findings in Proposition 2 by showing three different cases. In these figures, the solid line is \( p = \tilde{p} \). The areas below this line \( p \leq \tilde{p} \) represent the parameter space where a high-type messenger ad generates a higher expected utility of paying attention than a low-type one (in Lemma 2). Also, the 45-degree line represents \( p = E(v) \), the threshold where the private good/bad signal becomes useful for consumers’ purchase decisions (in Lemma 1).

We begin by noting that when \( \varepsilon_H(v) = \lambda \varepsilon_L(v) \) for all \( v \) (see Figure 4-(a)), we have \( E(v) = \tilde{p} \), and a high-type messenger is less likely to generate bad signals. Thus, its bad signals are more useful when \( p \leq E(v) \), preventing unwanted purchases. Also, a low-type messenger is less likely to generate good signals, which makes its good signals more useful when \( p > E(v) \), picking up potentially high-match products.

Next, when \( \varepsilon_H(v) > \lambda \varepsilon_L(v) \), we have \( E(v) \leq \tilde{p} \). As Figure 4-(b) shows, there are regions \( (E(v) \leq p \leq \tilde{p}) \) where a high-type messenger’s good signals are the most useful for consumers. Even though a high-type messenger would generally generate good signals, its signals are more sensitive to changes in match values, so they can update the posterior beliefs more than a low-type’s signals to convince consumers to purchase a product. On the other hand, when \( \varepsilon_H(v) < \lambda \varepsilon_L(v) \), we have

\[^{14}\text{The threshold price } \tilde{p} \text{ satisfies } \int_{\bar{v}}^{\tilde{v}} (v - \tilde{p}) \cdot [\sigma_H(v) - \sigma_L(v)] \, dF = 0. \text{ When } \varepsilon_H(v) = \lambda \varepsilon_L(v), \sigma_H(v) - \sigma_L(v) \text{ is constant, so the condition } \int_{\bar{v}}^{\tilde{v}} (v - \tilde{p})[\sigma_H(v) - \sigma_L(v)] \, dF = 0 \text{ degenerates to } \int_{\bar{v}}^{\tilde{v}} (v - \tilde{p}) \, dF = E(v) - \tilde{p} = 0.\]
$E(v) > \tilde{p}$. In this case, as Figure 4-(c) shows, there are regions ($\tilde{p} \leq p \leq E(v)$) where a low-type messenger’s bad signals can be useful for consumers, because they are more sensitive and can update the posterior beliefs more than a high-type to change the consumers’ purchase decisions.\footnote{Note that $|\varepsilon_b_j(v)| = \frac{\varepsilon_j}{1-\sigma_j(v)}$, where $\varepsilon_j = \frac{d(1-\sigma_j(v))}{dv}$ is the elasticity of bad signal from messenger $m_j$. Therefore, the more elastic a type of messengers’ good signals are, the more elastic their bad signals are. So, $\varepsilon_H(v) \leq \lambda \varepsilon_L(v)$ for all $v$ also implies that low-type messengers’ bad signals are sufficiently sensitive to changes in match values compared to that of high-type messengers.}

### 4.3 Price premiums

Given messenger type $m \in \{m_L, m_H\}$, consumers decide to pay attention when $p_j < p \leq \bar{p}_j$. We next compare the highest prices using different messengers with and without consumer attention.

First, we consider the case where the firm induces consumer attention. The highest prices associated with different messengers the firm can achieve are: $p_H$ and $p_L$.

**Proposition 3.**

1. When $\varepsilon_H(v) \geq \varepsilon_L(v)$, we always have $p_H \geq p_L$.
2. When $\lambda \varepsilon_L(v) \leq \varepsilon_H(v) < \varepsilon_L(v)$, we have $p_L > p_H$ if and only if $c < c^*$.\footnote{We can see $p_H \geq p_L \Leftrightarrow E(v|m_H, s_g) - \frac{\tilde{\sigma}_H}{\sigma_H} \geq E(v|m_L, s_g) - \frac{\tilde{\sigma}_L}{\sigma_L}$, which is $c \geq c^* \equiv \frac{\sigma_H \int_0^{v_H} \varepsilon_L(v) dF - \sigma_L \int_0^{\varepsilon_L(v)} \varepsilon_H(v) dF}{\sigma_H - \sigma_L}$.}
3. When $\varepsilon_H(v) < \lambda \varepsilon_L(v)$, we always have $p_L > p_H$.

The proposition suggests that depending on the signal environments, different messengers can induce a higher price premium under consumer attention. Figure 5 illustrates those findings. We note that $p_j = E(v|m_j, s_g) - \tilde{\sigma}_j$ has two parts: (1) the consumers’ updated beliefs conditional on receiving a good signal $E(v|m_j, s_g)$, and (2) the net deliberation cost per unit of good signal $\tilde{\sigma}_j$.\footnote{We can see $p_H \geq p_L \Leftrightarrow E(v|m_H, s_g) - \frac{\tilde{\sigma}_H}{\sigma_H} \geq E(v|m_L, s_g) - \frac{\tilde{\sigma}_L}{\sigma_L}$, which is $c \geq c^* \equiv \frac{\sigma_H \int_0^{v_H} \varepsilon_L(v) dF - \sigma_L \int_0^{\varepsilon_L(v)} \varepsilon_H(v) dF}{\sigma_H - \sigma_L}$.}
The magnitude of consumers’ updated beliefs given a good signal is positively correlated to the informativeness of a messenger’s good signal, while a higher deliberation cost always has a more negative effect on the low-type messenger’s ability to raise the price because $c < c^*$. 

First, when $\varepsilon_H(v) \geq \lambda \varepsilon_L(v)$, a high-type messenger’s good signals is always more informative according to Proposition 1. So, it is always the case that $p_H \geq p_L$ (Figure 5-(a)). When $\varepsilon_H(v) < \varepsilon_L(v)$, a low-type messenger’s good signals become more informative. In particular, when $\kappa \varepsilon_L(v) \leq \varepsilon_H(v) < \lambda \varepsilon_L(v)$ (Figure 5-(b)), a low-type messenger’s good signals are moderately informative, and $p_L$ can be larger than $p_H$ when the deliberation cost is low such that the low-type messenger’s ability to raise the price is not undermined. Finally, when $\varepsilon_H(v) < \kappa \varepsilon_L(v)$ for all $v$, a low-type messenger’s good signals are sufficiently informative to the point that they can be useful for consumers purchase decisions (Proposition 2), and it is always the case that $p_L > p_H$ (Figure 5-(c)).

Next, we compare the highest price premiums from different types of messengers that the firm can achieve under no consumer attention: $p_H$ and $p_L$.

**Proposition 4.**

1. When $\varepsilon_H(v) \geq \lambda \varepsilon_L(v)$, we have $p_L > p_H$.

2. When $\kappa \varepsilon_L(v) \leq \varepsilon_H(v) < \lambda \varepsilon_L(v)$ ($\kappa \equiv \frac{\sigma_L(v)(1-\sigma_H(v))}{\sigma_H(v)(1-\sigma_L(v))}$), we have $p_L > p_H$ if and only if $c < c^{**}$.\footnote{We have $\frac{p_H}{c} \geq \frac{p_L}{c} \iff E(v|m_H,s_b) + \frac{\varepsilon_H}{\varepsilon_L} \geq E(v|m_L,s_b) + \frac{\varepsilon_L}{\varepsilon_H}$, which is equivalent to $c \geq c^{**} \equiv \frac{\int_{v_0}^{v_H} [1-\sigma_H(v)] dF - \int_{v_L}^{v_0} [1-\sigma_L(v)] dF}{\sigma_H - \sigma_L}$.}

3. When $\varepsilon_H(v) < \kappa \varepsilon_L(v)$, we have $p_H \geq p_L$.

Again, $p_j \equiv E(v|m_j,s_b) + \frac{\varepsilon_j}{1-\sigma_j}$ also has two parts: (1) consumers’ updated beliefs conditional on receiving a bad signal $E(v|m_j,s_b)$, and (2) the net deliberation cost per unit of bad signal
Similar to the case with consumer attention, the magnitude of consumers’ updated beliefs is related to the informativeness of a messenger’s signal, but negatively because the bad signal makes consumers more pessimistic about the match value. Also, a higher deliberation cost has a positive effect on the price, and it always has a bigger effect on the high-type messenger’s ability to maintain the price under no attention because \( \frac{c}{1-\sigma_H} > \frac{c}{1-\sigma_L} \). Thus, more elastic (or informative) signals from the messenger put downward pressure on the price in the no consumer attention case, while a higher deliberation cost always has a more positive effect for the high-type messenger.

When \( \varepsilon_H(v) \geq \lambda \varepsilon_L(v) \), a high-type messenger’s bad signals become sufficiently informative to the extent that they can be useful for consumer purchase decisions (Proposition 2), leading to a lower price under no attention: \( p_L > p_H \). When \( \varepsilon_H(v) < \lambda \varepsilon_L(v) \), a high-type messenger’s bad signals become less informative, which enhances its ability to maintain a high price under no attention. In particular, when \( \kappa \varepsilon_L(v) \leq \varepsilon_H(v) < \lambda \varepsilon_L(v) \), a high-type messenger’s bad signals are moderately uninformative, so \( p_L \) can be higher than \( p_L \) as the deliberation cost becomes sufficiently high such that the high-type has a better advantage of maintaining a high price under no attention. Finally, when \( \varepsilon_H(v) < \kappa \varepsilon_L(v) \), a high-type messenger’s bad signals are particularly uninformative such that it is always the case that \( p_H \geq p_L \). Figure 6 demonstrates those findings.

4.4 Optimal choice of messenger

We now turn to the firm’s problem. The firm chooses a type of messenger and the price, affecting consumer attention and purchasing decisions. The firm’s objective is to maximize its expected
profit, which is

$$\mathbb{E}\Pi = D(p, m_j) \cdot (p - k),$$  \hspace{1cm} (11)$$

where \(D(p, m_j)\) denotes the demand for the product, and \(p - k\) is the profit margin. We first assume that consumers can costlessly observe the price before they make attention decisions, and later we analyze an alternative scenario where price is unobservable prior to deliberation in Section 5.3. Propositions 3 and 4 state which types of the messenger can generate a higher profit margin. However, the demand function \(D(p, m_j)\) is also a function of the price \(p\) and the type of messenger \(m_j\). Thus, we need to consider the overall effects of the messenger on the firm’s profit through both channels (profit margin and demand) together. Given the consumer attention decision in Lemma 1, the firm considers two demand regimes: full market coverage and partial market coverage.

First, the firm can always set the price sufficiently low such that all consumers purchase the product without incurring deliberation costs, making \(D(p, m_j) = 1\), which is the full market coverage without consumer attention case. Given a messenger \(m_j\) for any \(j \in \{L, H\}\), the price has to be set lower than \(p_j\) so that consumers purchase the product without paying attention. Thus, under full coverage, the firm’s optimization problem is

$$\max_{p, m_j \in \{m_L, m_H\}} \mathbb{E}\Pi_{\text{full}} = 1 \cdot (p - k),$$  \hspace{1cm} (12)$$

subject to \(k \leq p \leq p_j\).

The (IC) constraint \(p \leq p_j\) ensures all the consumers purchase without paying attention. Also, the firm has to set the price higher than the production cost, \(k \leq p\), which is the (IR) constraint for the firm. Suppose that \(k \leq \min\{p_L, p_H\}\) so that the firm’s (IR) constraint can be satisfied when the firm sets \(p = p_j\). Then, under full coverage, the firm’s optimal price must make the (IC) constraint bind such that \(p = p_j\) for some \(j \in \{L, H\}\). Solving (12) yields the following result. The optimal messenger choice is from Proposition 4, and the price is from (IC) constraint in (12).

**Result 1.** Under full coverage with no consumer attention, when (1) \(\varepsilon_H(v) \geq \lambda \varepsilon_L(v)\) for all \(v\), or (2) \(\kappa \varepsilon_L(v) \leq \varepsilon_H(v) < \lambda \varepsilon_L(v)\) and \(c < c^{**}\), a low-type messenger ad is optimal: \(m^*_{\text{full}} = m_L\), \(p^*_{\text{full}} = p_L\), and \(\mathbb{E}\Pi_{\text{full}}(m_L) = \frac{c}{1 - \sigma_L} - k\). Otherwise, a high-type messenger ad is optimal:
Next, for a given messenger $m_j$, the firm can set the price $p_j < p \leq \bar{p}_j$ so that consumers pay attention to the ad and purchase the product only if they receive a good signal, making $D(p, m_j) = \tilde{\sigma}_j = \Pr(s_g = g|m_j)$ (partial market coverage). Thus, the firm’s optimization problem under partial coverage with consumer attention is

$$\max_{p, m_j \in \{m_L, m_H\}} \Pi_{\text{part}} = \tilde{\sigma}_j \cdot (p - k),$$

subject to $p_j < p \leq \bar{p}_j$ and $k \leq p$.

Similar to the full coverage case, the optimal price must make the constraint $p \leq \bar{p}_j$ binds and for some $j \in \{L, H\}$, and the price must be higher than the production cost $k$.

Under partial coverage with consumer attention, the firm faces a trade-off of sales $\tilde{\sigma}_j$ (where $\tilde{\sigma}_L < \tilde{\sigma}_H$) versus margin $p_j - k$. When $\bar{p}_H \geq \bar{p}_L$, it is optimal for the firm to choose a high-type messenger. However, when $\bar{p}_H < \bar{p}_L$, a low-type messenger ad is not necessarily optimal despite its higher margin due to lower sales. Thus, only when the profit margin $\bar{p}_L - k$ is sufficiently larger than $p_H - k$, the low-type messenger can be optimal, which is the case when $\Pi_{\text{part}}(m_H) = \tilde{\sigma}_H \cdot (\bar{p}_H - k) < \Pi_{\text{part}}(m_L) = \tilde{\sigma}_L \cdot (\bar{p}_L - k) \iff k > \tilde{p}$. We assume that the production cost is less than an average consumer’s consumption utility: $k < E(v)$. So, $k > \tilde{p}$ is not possible when $\varepsilon_H(v) \geq \lambda \varepsilon_L(v)$ for all $v$ (from Proposition 2), which leads to the following result.

**Result 2.** Under partial coverage with consumer attention, when $\varepsilon_H(v) < \lambda \varepsilon_L(v)$ for all $v$, and $k > \tilde{p}$, a low-type messenger ad is optimal: $m^*_{\text{part}} = m_L$, $p^*_{\text{part}} = \bar{p}_L$, and $\Pi_{\text{part}}(m_L) = \tilde{\sigma}_L (\bar{p}_L - k) = \tilde{\sigma}_L (E(v|m_L, s_g) - k) - c$. Otherwise, a high-type messenger ad is optimal: $m^*_{\text{part}} = m_H$, $p^*_{\text{part}} = \bar{p}_H$, and $\Pi_{\text{full}}(m_H) = \tilde{\sigma}_H (\bar{p}_H - k) = \tilde{\sigma}_H (E(v|m_H, s_g) - k) - c$.

Table 1 summarizes our findings from results 1 and 2 about the firm’s optimal messenger and price under each regime.

Next, we compare the expected profits under partial and full coverage to determine the firm’s coverage choice with optimal messenger and price, which is the main result of this research.

**Proposition 5.** 1. When $\varepsilon_H(v) \geq \lambda \varepsilon_L(v)$, it is optimal to induce partial coverage with $(m_H, \bar{p}_H)$ if $k \geq k_1(c)$. Otherwise, full coverage with $(m_L, \bar{p}_L)$ is optimal.
Coverage $\varepsilon_H(v) \geq \lambda \varepsilon_L(v)$ $\kappa \varepsilon_L(v) \leq \varepsilon_H(v) < \lambda \varepsilon_L(v)$ $\varepsilon_H(v) < \kappa \varepsilon_L(v)$

<table>
<thead>
<tr>
<th>Coverage</th>
<th>$\varepsilon_H(v) \geq \lambda \varepsilon_L(v)$</th>
<th>$\kappa \varepsilon_L(v) \leq \varepsilon_H(v) &lt; \lambda \varepsilon_L(v)$</th>
<th>$\varepsilon_H(v) &lt; \kappa \varepsilon_L(v)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full (Without consumer attention)</td>
<td>$(m_L, p_L)$</td>
<td>$\begin{cases} (m_L, p_L), \ (m_H, p_H), \text{ otherwise} \end{cases}$</td>
<td>$(m_H, p_H)$</td>
</tr>
<tr>
<td>Partial (With consumer attention)</td>
<td>$(m_H, \bar{p}_H)$</td>
<td>$\begin{cases} (m_L, \bar{p}_L), \ (m_H, \bar{p}_H), \text{ otherwise} \end{cases}$</td>
<td>$(m_H, \bar{p}_H)$</td>
</tr>
</tbody>
</table>

Table 1: Optimal choice of messenger and price $(m^*_j, p^*_j)$ under full and partial coverage

2. When $\kappa \varepsilon_L(v) \leq \varepsilon_H(v) < \lambda \varepsilon_L(v)$, it is optimal to induce partial coverage if $k \geq k_2(c)$. Otherwise, it is optimal to induce full coverage, where $(m_H, p_H)$ is optimal if and only if $c \geq c^{**}$.

3. When $\varepsilon_H(v) < \kappa \varepsilon_L(v)$, it is optimal to induce partial coverage if $k \geq k_3(c)$. Otherwise, it is optimal to induce full coverage with $(m_H, p_H)$.

Moreover, all thresholds are increasing in $c$: $\frac{\partial k_j(c)}{\partial c} \geq 0$, for all $j \in \{1, 2, 3\}$.

Figure 7 illustrates the findings from proposition 5, where the darker areas in each sub-figure represent the firm’s partial coverage choice with consumer attention. When deciding whether or not to induce consumer attention, the firm has to consider both sales volume and price premium, optimizing its profit. As the production cost $k$ increases, the need to increase the price becomes higher. Thus, the firm tends to sacrifice demand to increase its price. On the other hand, when the consumer’s deliberation cost increases, the firm’s ability to charge a higher price under no consumer attention increases, tempting the firm to sacrifice profit margin to retain greater demand. Thus, full coverage without consumer attention is more likely as the consumer’s deliberation cost $c$ becomes larger or the product cost $k$ becomes smaller.

Also, all the threshold levels of $k$ for inducing consumer attention (i.e., partial coverage) are increasing in the deliberation cost: $\frac{\partial k_j(c)}{\partial c} \geq 0$, for all $j \in \{1, 2, 3\}$. When the deliberation cost is small, the firm can raise the price sufficiently by inducing consumer attention (recall that $p_j$ decreases in $c$). Thus, even when the need to boost the price is not too strong ($k$ is small), the firm still prefers to induce consumer attention. Conversely, when the deliberation cost becomes larger, the firm has less incentive to raise the price through consumer attention. So, the firm chooses full coverage without consumer attention unless the need to boost the price is strong ($k$ is large).

Moreover, the optimal choice of messengers depends on each messenger’s elasticity and the
firm’s coverage choice (which is again a function of the consumer’s deliberation cost and production cost). First, when a high-type messenger is sufficiently informative such that $\varepsilon_H(v) \geq \lambda \varepsilon_L(v)$ (see Figure 7-(a)), a high-type messenger ad is more effective for raising the price, making it optimal for attracting consumer attention and boosting the price when $k$ is large relative to $c$ ($k \geq k_1(c)$).

At the same time, because a high-type messenger’s signals are so informative, it also means that the firm has to set the price very low in order to dissuade consumer attention, making a low-type messenger optimal for full coverage without consumer attention.

When a high-type messenger becomes less informative such that $\varepsilon_H(v) < \lambda \varepsilon_L(v)$ (both Figure 7-(b) and 7-(c)), a high-type messenger ad generally provides only good signals, which can overshadow the product, thus limiting its ability to update consumers’ beliefs about the product match (for example, see Figure 3-b). In contrast, as it is rare, the low-type messenger’s good signal can significantly update consumers’ beliefs. Therefore, for partial coverage ($k \geq k_2(c)$ or $k \geq k_3(c)$), when (1) the low-type messenger’s signals are sufficiently elastic (or informative), and (2) the production cost is sufficiently high ($k > \tilde{p}$), a low-type messenger can indeed be optimal for inducing consumer attention and charging a high price to cover its product cost. On the other hand, under full coverage, the firms should use the type of messenger that can command a higher price without consumer attention. For example, a high-type messenger is optimal under full coverage without consumer attention when $\kappa \varepsilon_L(v) \leq \varepsilon_H(v) < \lambda \varepsilon_L(v)$ and $c \leq c^{**}$ (Figure 7-(b)).

**Proposition 6.** Suppose $\varepsilon_H(v) < \lambda \varepsilon_L(v)$. For partial coverage ($k \geq k_2(c)$ or $k \geq k_3(c)$), it is optimal for the firm to choose a low-type messenger with $(m_L, \tilde{p}_L)$ when $k > \tilde{p}$. 

Figure 7: Optimal choice of market coverage and messenger
5 Extensions

5.1 Attention-grabbing role

We assumed the deliberation cost is constant and fixed for both types of messengers \((c(m_H) = c(m_L) = c)\) to get a cleaner intuition. We now relax this assumption to capture the attention-grabbing role of the advertising messenger. In practice, a high-type messenger can directly lower the consumer’s deliberation cost by providing extra entertaining value that offsets the deliberation cost. To capture this idea, we assume that a high-type messenger ad can reduce the deliberation cost to zero, thus making all consumers pay attention to the ad: \(c(m_H) = 0 < c(m_L) = c\).

The analysis is similar to the main model, except that the demand from a high-type messenger ad changes. Because the deliberation cost of a high-type messenger ad is zero \(c(m_H) = 0\), consumers always obtain noisy signals from it. Therefore, the consumer’s posterior belief is either \(E(v|m_H, s_g)\) when receiving a good signal or \(E(v|m_H, s_b)\) when receiving a bad signal and the demand function associated with a high-type messenger ad becomes

\[
D(p, m_H) = \begin{cases} 
1, & p \leq E(v|m_H, s_b) \\
\tilde{\sigma}_H, & E(v|m_H, s_b) < p \leq E(v|m_H, s_g) \\
0, & p > E(v|m_H, s_g) 
\end{cases}. \tag{14}
\]

Again, the firm has to consider both the demand and the profit margin when maximizing the profit. Similar to the analyses in the main model, we first determine the optimal price and messenger under the full and partial coverage cases. According to equation (14), the optimal price for a high-type messenger ad is \(E(v|m_H, s_b) \equiv p_H^*\) under full coverage and \(E(v|m_H, s_g) \equiv p_H^*\) under partial coverage. Under full coverage without consumer attention, a low-type messenger ad is optimal if (1) \(\varepsilon_H(v) \geq \kappa \varepsilon_L(v)\) or (2) \(\varepsilon_H(v) < \kappa \varepsilon_L(v)\) and \(c > (1 - \tilde{\sigma}_L)(E(v|m_H, s_b) - E(v|m_L, s_b)) \equiv c^{***}\). Also, under partial coverage with consumer attention, a low-type messenger ad is optimal if

\[
\tilde{\sigma}_L \left( E(v|m_L, s_g) - \frac{c}{\tilde{\sigma}_L} - k \right) > \tilde{\sigma}_H \left( E(v|m_L, s_g) - k \right) \iff k > \bar{p} + \frac{c}{\tilde{\sigma}_H - \tilde{\sigma}_L}, \tag{15}
\]

where \(\bar{p} = \int_{\varepsilon_L(v)}^{\varepsilon_H(v)} \frac{v \sigma_H(v) - \sigma_L(v)}{\tilde{\sigma}_H - \tilde{\sigma}_L} dv\) as before. When \(\varepsilon_H(v) \geq \lambda \varepsilon_L(v)\), \(\bar{p} \geq E(v)\) and a high-type messenger
ad is always optimal under partial coverage with consumer attention. We summarize the results from the full and the partial coverage cases in the Table 2.

<table>
<thead>
<tr>
<th>Coverage</th>
<th>$\epsilon_H(v) \geq \lambda \epsilon_L(v)$</th>
<th>$\kappa \epsilon_L(v) \leq \epsilon_H(v) &lt; \lambda \epsilon_L(v)$</th>
<th>$\epsilon_H(v) &lt; \kappa \epsilon_L(v)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full</td>
<td>$(m_L, p_L)$</td>
<td>$(m_L, \bar{p}_L)$, $c &gt; c^{***}$</td>
<td>$(m_H, \bar{p}_H)$, otherwise</td>
</tr>
<tr>
<td>Partial</td>
<td>$(m_H, \bar{p}_H)$</td>
<td>$(m_L, \bar{p}_L)$, $k &gt; \bar{p} + \frac{c}{\sigma_H - \sigma_L}$</td>
<td>$(m_H, \bar{p}_H)$, otherwise</td>
</tr>
</tbody>
</table>

Table 2: Optimal choice of messenger and price under full and partial coverage

We next compare the full and the partial cases to determine the firm’s optimal coverage with optimal messenger and price in the following proposition.

**Proposition 7.** 1. When $\epsilon_H(v) \geq \lambda \epsilon_L(v)$, it is optimal to induce partial coverage with $(m_H, \bar{p}_H)$ if $k \geq k_4(c)$. Otherwise, full coverage with $(m_L, p_L)$ is optimal.

2. When $\kappa \epsilon_L(v) \leq \epsilon_H(v) < \lambda \epsilon_L(v)$, it is optimal to induce partial coverage if $k \geq k_5(c)$. Otherwise, full coverage with $(m_L, p_L)$ is optimal. Moreover, under partial coverage, the firm chooses a low-type messenger with $(m_L, \bar{p}_L)$ if and only if $k > \bar{p} + \frac{c}{\sigma_H - \sigma_L}$.

3. When $\epsilon_L(v) < \kappa \epsilon_L(v)$, it is optimal to induce partial coverage if $k \geq k_6(c)$. Otherwise, full coverage is optimal. Moreover, under partial coverage, the firm chooses a low-type messenger with $(m_L, \bar{p}_L)$ if and only if $k \geq \bar{p} + \frac{c}{\sigma_H - \sigma_L}$. Also, under full coverage, the firm chooses a low-type messenger with $(m_L, p_L)$ if and only if $c > c^{***}$.

Moreover, all thresholds are increasing in $c$: $\frac{\partial k_4(c)}{\partial c} \geq 0$, $\frac{\partial k_5(c)}{\partial c} \geq 0$, and $\frac{\partial k_6(c)}{\partial c} \geq 0$.

Figure 8 illustrates the results from Proposition 7. Similar to the main analysis, it is optimal to induce partial coverage with a higher price when the production cost is sufficiently large relative to the deliberation cost. An interesting departure from the main analysis arises: the additional attention-grabbing effect of the high-type messenger effectively diminishes the range for the low-type to be optimal under partial coverage with consumer attention. As the deliberation cost increases, the firm has to lower the price for a low-type messenger ad to attract consumer attention, while there is no such need for a high-type messenger ad. Therefore, we see the parameter region for adopting the low-type messenger shrinks and ultimately disappears as the deliberation cost $c$
becomes higher. Thus, even under high production cost $k$, a high-type messenger becomes more effective for advertising communication (see Figures 8-(b) and 8-(c)). Also, in the main model, a higher deliberation cost enables the firm to charge a higher price $p_j$ while dissuading consumer attention under full coverage. However, such an advantage disappears for a high-type messenger. Thus, under full coverage, a low-type messenger dominates in most cases. Only when (1) $\varepsilon_H(v)$ is sufficiently small that $p^*_H$ remains high, and (2) the deliberation cost is also small such that the firm has to charge a low price with a low-type messenger, a high-type messenger is optimal.

Nevertheless, we still find that a low-type messenger can be a more effective medium for advertisers, even with the attention-grabbing effect from a high-type messenger. That is, the general insight that a low-type messenger can still draw consumer attention and, more importantly, persuade them to pay a higher premium carries over if the importance of boosting the price is high, but a high-type messenger is less informative ($\varepsilon_H(v) < \lambda \varepsilon_L(v)$).

### 5.2 Naïve consumers

In our main model, we assume all consumers are rational. An alternative scenario is when some consumers are naïve, in the sense that they do not rationally make inference and attention decisions. Instead, they always pay attention to the ad and take private signals at their face values regardless of who generated those signals. In this section, we explore such a realistic situation where a $\alpha$ portion of consumers are naïve consumers. Specifically, those naïve consumers’ posteriors upon observing the signals are $E^*(v|m_j, s_g) = \bar{v}$ and $E^*(v|m_j, s_b) = 0$ for any $j \in \{L, H\}$. That is, they believe that
the signals are completely informative about the product match. Therefore, a high-type messenger has the advantage of generating more good signals and being able to convert more naïve consumers.

Because naïve consumers always pay attention to the ad and purchase the product if and only if they receive a good signal, the demand function is

\[ D(m_j, p) = \begin{cases} 
\alpha \tilde{\sigma}_j + (1 - \alpha), & p \leq \bar{p}_j \\
\tilde{\sigma}_j, & \bar{p}_j < p \leq \bar{p}_j \\
\alpha \tilde{\sigma}_j, & \bar{p}_j < p \leq \bar{v} 
\end{cases} \]

(16)

Note that as \( \alpha \to 1 \), \( D(m_j, p) \to \tilde{\sigma}_j \) for all prices \( p \leq \bar{v} \), making \( \bar{v} \) as the optimal price. Then, a high-type messenger is optimal because \( \tilde{\sigma}_H > \tilde{\sigma}_L \). Naïve consumers don’t take into account the difference in signal generating processes from different messengers when making inference.

For general \( \alpha \), we can discuss three cases in terms of the coverage of “rational” consumers. First, in the full coverage case without (rational) consumer attention \( (p \leq \bar{p}_j) \), the optimal price given the messenger \( m_j \) is \( \bar{p}_j \), and a high-type messenger is optimal if and only if \([\alpha \tilde{\sigma}_H + (1 - \alpha)](\bar{p}_L - k) \geq [\alpha \tilde{\sigma}_L + (1 - \alpha)](\bar{p}_L - k) \Leftrightarrow k \leq \frac{\bar{p}_H - \bar{p}_L}{\alpha(\tilde{\sigma}_H - \tilde{\sigma}_L)} + \tilde{\sigma} \equiv \tilde{k}(c, \alpha) \). Next, in the partial coverage case with (rational) consumer attention \( (\bar{p}_j < p \leq \bar{p}_j) \), the optimal price given the messenger \( m_j \) is \( \bar{p}_j \), and a high-type messenger is optimal if and only if \( \tilde{\sigma}_H(\bar{p}_j - k) \geq \tilde{\sigma}_L(\bar{p}_j - k) \Leftrightarrow k \leq \tilde{p} \), which always holds when \( \varepsilon_H(v) \geq \lambda \varepsilon_L(v) \) for all \( v \). Moreover, there is a new case where no (rational) consumers purchase – no coverage case \( (\bar{p}_j < p \leq \bar{v}) \), where only naïve consumers purchase when they receive private good signals. In this case, a high-type messenger is always optimal and the optimal price is \( \bar{v} \). Table 3 summarizes the results about the firm’s optimal messenger and price under each regime.

<table>
<thead>
<tr>
<th>Coverage</th>
<th>( \varepsilon_H(v) \geq \lambda \varepsilon_L(v) )</th>
<th>( \varepsilon_H(v) &lt; \lambda \varepsilon_L(v) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full (Without rational consumer attention)</td>
<td>( {m_L, \bar{p}_L}, k &gt; \tilde{k}(c, \alpha) )</td>
<td>( {m_H, \bar{p}_H}, \text{otherwise} )</td>
</tr>
<tr>
<td>Partial (With rational consumer attention)</td>
<td>( (m_H, \bar{p}_H) )</td>
<td>( {m_L, \bar{p}_L}, k &gt; \bar{p} )</td>
</tr>
<tr>
<td>No (Only Naive Consumers)</td>
<td>( (m_H, \bar{v}) )</td>
<td>( (m_H, \bar{v}) )</td>
</tr>
</tbody>
</table>

Table 3: Optimal choice of messenger and price \( (m^*_j, p^*) \) when some consumers are naïve
Figure 9: Optimal choice of messenger when some consumers are naïve (\(\alpha\) close to 0)

If \(\alpha\) becomes sufficiently large, the case reverts to a trivial case. Intuitively, no coverage case where the firm targets only naïve consumers becomes an optimal choice, making the high-type messenger a more promising option. We assume that \(\alpha\) is small enough such that
\[
\alpha \leq \min \left\{ \frac{1 - \tilde{\sigma}_H}{1 - \tilde{\sigma}_L}, \frac{\tilde{\sigma}_L}{\tilde{\sigma}_H} \right\},
\]
where the general demand structure in the main model holds.

Proposition 8. Suppose \(\alpha \leq \min \left\{ \frac{1 - \tilde{\sigma}_H}{1 - \tilde{\sigma}_L}, \frac{\tilde{\sigma}_L}{\tilde{\sigma}_H} \right\}\). Then, the optimal choice of coverage and messenger is characterized as follows.

1. When \(\varepsilon_H(v) \geq \lambda \varepsilon_L(v)\), it is optimal to induce no coverage with \((m_H, \bar{v})\) if \(k \geq k_7(c, \alpha)\), and it is optimal to induce full coverage if \(k \leq k_8(c, \alpha)\). Otherwise, partial coverage with \((m_H, \bar{p}_H)\) is optimal. Moreover, under full coverage, \((m_L, \bar{p}_L)\) is optimal if and only if \(k > \tilde{k}(c, \alpha)\).

2. When \(\varepsilon_H(v) < \lambda \varepsilon_L(v)\), it is optimal to induce no coverage with \((m_H, \bar{v})\) if \(k \geq k_9(c, \alpha)\), and it is optimal to induce full coverage if \(k \leq k_{10}(c, \alpha)\). Otherwise, partial coverage is optimal. Moreover, under full coverage, \((m_L, \bar{p}_L)\) is optimal if and only if \(k > \tilde{k}(c, \alpha)\). Also, under partial coverage, \((m_L, \bar{p}_L)\) is optimal if and only if \(k > \tilde{p}\).

Moreover, all thresholds are decreasing in \(\alpha\):
\[
\left(\frac{\partial k_7(c, \alpha)}{\partial \alpha}\right) \leq 0, \left(\frac{\partial k_8(c, \alpha)}{\partial \alpha}\right) \leq 0, \left(\frac{\partial k_9(c, \alpha)}{\partial \alpha}\right) \leq 0 \text{ and } \left(\frac{\partial k_{10}(c, \alpha)}{\partial \alpha}\right) \leq 0.
\]

Figure 9 and 10 illustrate Proposition 8. In particular, a low-type messenger can optimally

\[\text{constant in the partial coverage case: } \frac{\partial \sigma_L}{\partial \alpha} = 0, \text{ and decreasing in the full coverage case: } \frac{\partial \sigma_H}{\partial \alpha} = -\left(1 - \sigma_L\right)\left(\bar{p} - k\right) < 0. \]

Thus, as \(\alpha\) becomes larger, the profit under no coverage case dominates the other cases. Without naïve consumers, the consumer demand in the full coverage without inducing consumer attention is always greater than the partial coverage without consumer attention. However, introducing naïve consumers can disrupt such hierarchy. Thus, the condition ensures the demand hierarchy among different coverage cases stays constant. The partial coverage demand is smaller than the full coverage, and the demand is smaller in the no coverage than in the partial coverage: \(\alpha \sigma_H + (1 - \alpha) \geq \sigma_H \Leftrightarrow \alpha \leq \frac{\tilde{\sigma}_H}{1 - \tilde{\sigma}_L} \text{ and } \sigma_L \geq \alpha \sigma_H \Leftrightarrow \sigma_L \geq \frac{\tilde{\sigma}_L}{\tilde{\sigma}_H} \).
Figure 10: Optimal choice of messenger when some consumers are naïve ($\alpha$ close to $\min \{1 - \tilde{\sigma}_H, \tilde{\sigma}_L\}$) attract rational consumers’ attention when $\varepsilon_H(v) < \lambda \varepsilon_L(v)$ for all $v$ and $\tilde{p} < k < k_9(c, \alpha)$ (Figure 9-(b)). However, as the proportion of naïve consumers increases and $\alpha \rightarrow \min \{1 - \tilde{\sigma}_H, \tilde{\sigma}_L\}$, (Figure 10-(b)), such region would shrink and can eventually disappear. This is because the no-coverage case would dominate the low-type messenger’s advantage of boosting the price under partial coverage.

As more consumers are naïve and willing to pay the highest price upon receiving good signals, the high-type messenger can boost the price even more ($p = \bar{v}$).

5.3 Unobservable Pricing

In this section, we relax the assumption that consumers can costlessly observe the price of the product. Consumers can only costlessly observe the type of the messenger, and they must make attention decisions based on that information and the expected price of the product, $p^e$. However, this would lead to a classic hold-up problem. When consumers receive a good signal, the firm would charge a higher price than expected price to extract consumer surplus since consumers’ deliberation cost is sunk. Anticipating such opportunistic behaviors of the firm, consumers are unwilling to pay attention to the ad ex-ante. To address such problem, we introduce consumer heterogeneity by assuming that a portion $\beta$ of consumers have no deliberation cost so that they always pay attention to the ad, while $(1 - \beta)$ of “regular” consumers have a deliberation cost $c > 0$.

The consumers without deliberation costs always obtain a private signal and observe the price freely, so they would purchase the product if the actual price is lower than their updated beliefs about product match: $p \leq E(v|m_j, s_i)$. Next, we assume that when regular consumers commit not
Table 4: Optimal choice of messenger and price \((m_j^*, p^*)\) when pricing is unobservable

to pay attention to the ad (i.e., when \(p^e \leq p_j\) or \(p^e > P_j\)), they can costlessly observe the price if and only if they expect they would purchase the product (i.e., \(p^e \leq p_j\)). Therefore, the consumer demand \(D(m_j, p^e, p)\) depends on the type of messenger, and the expected and the actual price:

\[
D(m_j, p^e, p) = \begin{cases} 
\beta \{ \tilde{\sigma}_j I(p \leq E(v|m_j, s_g)) + (1 - \tilde{\sigma}_j)I(p \leq E(v|m_j, s_b)) \} + (1 - \beta)I(p \leq E(v)), & p^e \leq p_j \\
\tilde{\sigma}_j I(p \leq E(v|m_j, s_g)) + (1 - \tilde{\sigma}_j)I(p \leq E(v|m_j, s_b)), & p_j < p^e \leq P_j \\
\beta \{ \tilde{\sigma}_j I(p \leq E(v|m_j, s_g)) + (1 - \tilde{\sigma}_j)I(p \leq E(v|m_j, s_b)) \}, & p^e > P_j 
\end{cases}
\]

Given the consumer demand, the profit-maximizing price is either \(E(v|m_j, s_b), E(v),\) or \(E(v|m_j, s_g)\). However, \(p = E(v)\) can never be consistent with the expected price \(p^e\) because \(p = E(v)\) can be optimal only when \(p^e \leq p_j < E(v)\). Thus, only \(E(v|m_j, s_b)\) and \(E(v|m_j, s_g)\) can possibly be consistent with the expected price in the equilibrium. We discuss these two cases in terms of the coverage of the regular consumers: the full coverage without regular consumer attention case with \(p = E(v|m_j, s_b)\) and the partial coverage with regular consumer attention case with \(p = E(v|m_j, s_g)\).\(^{20}\)

In the full coverage case, the high-type messenger is optimal if and only if \(E(v|m_H, s_b) \geq E(v|m_L, s_b)\), which holds when \(\varepsilon_H(v) < \kappa \varepsilon_L(v)\) for all \(v\). In the partial coverage case, the high-type messenger is optimal if and only if \(\beta \tilde{\sigma}_H (E(v|m_H, s_g) - k) \geq \beta \tilde{\sigma}_L (E(v|m_L, s_g) - k) \Leftrightarrow k \leq \tilde{p}\), which is always the case when \(\varepsilon_H(v) \geq \lambda \varepsilon_L(v)\). Table 4 summarizes the results about the firm’s optimal messenger and price under each regime.

**Proposition 9.** 1. When \(\varepsilon_H(v) \geq \lambda \varepsilon_L(v)\), it is optimal to induce partial coverage with \((m_H, E(v|m_H, s_g))\) if \(k \geq k_{11}\). Otherwise, full coverage with \((m_L, E(v|m_L, s_b))\) is optimal.

2. When \(\kappa \varepsilon_L(v) \leq \varepsilon_H(v) < \lambda \varepsilon_L(v)\), it is optimal to induce partial coverage if \(k \geq k_{12}\). Otherwise,

\(^{20}\)We assume that \(\beta\) is large enough so that either \(p = E(v|m_j, s_b)\) or \(p = E(v|m_j, s_g)\) is optimal. Specifically, \(\beta \geq \max_{E(v|m_j, s_b) \leq E(v|m_j, s_g)} \frac{2(\varepsilon_L(v|m_j, s_g) - E(v|m_j, s_g))}{2E(v|m_j, s_g) - 2\varepsilon_L(v|m_j, s_g) + 2\lambda \varepsilon_L(v|m_j, s_g) + 2E(v|m_j, s_b) - 2\varepsilon_L(v|m_j, s_b) + 2\lambda \varepsilon_L(v|m_j, s_b)}\). When \(\beta < \frac{\max_{E(v|m_j, s_b) \leq E(v|m_j, s_g)}}{\min_{E(v|m_j, s_b) \leq E(v|m_j, s_g)}}\), there are regions of parameters where a pure-strategy equilibrium does not exist.
full coverage with \((m_L, E(v|m_L, s_b))\) is optimal. Moreover, under partial coverage, the firm chooses a low-type messenger with \((m_L, E(v|m_L, s_g))\) if and only if \(k > \tilde{p}\).

3. When \(\varepsilon_H(v) < \kappa \varepsilon_L(v)\), it is optimal to induce partial coverage if \(k \geq k_{13}\). Otherwise, it is optimal to induce full coverage with \((m_H, E(v|m_H, s_b))\). Moreover, under partial coverage, the firm chooses a low-type messenger with \((m_L, E(v|m_L, s_g))\) if and only if \(k > \tilde{p}\).

The results are similar to the main results, except that the regions where a messenger-price pair is optimal are independent of the deliberation cost. This is because no price can credibly attract attention from regular consumers. Therefore, the firm either gives up those consumers by setting a very high price \(E(v|m_j, s_g)\) or sets a very low price \(E(v|m_j, s_b)\) to make all consumers purchase without paying attention to the ad.

5.4 Other extensions

We also carry out other extensions to assess the robustness of our main insight. We overview those extensions and the details relegated to the Online Appendix. Overall, we find that the qualitative insights from the main model carry over to these extension models.

1. **Social utility and fandom:** Sometimes, high-type messengers such as celebrities can directly affect consumers’ preference for a product because consumers might project their feelings about celebrities toward the product. Thus, we assume that other than the consumption utility \(v_i\), consumers can obtain additional social utilities \(\eta > 0\) if the product is endorsed by a high-type messenger. The results are qualitatively the same as the main results. However, as \(\eta\) becomes larger, we find that the high-type messenger has the advantage of boosting the price, and thus, the regions where a high-type messenger is optimal become larger.

2. **Differential fixed costs:** In the main model, we assumed the cost of featuring the messenger as fixed and the same across different messengers, normalizing it to zero. In practice, featuring a celebrity would generally be more expensive than featuring ordinary people. We analyze such a case where a different messenger ad requires different fixed costs.

3. **Exogenous pricing case:** In many situations, the prices of different products are similar within a product category. For example, most mobile apps are priced at $0.99, and consumers know the price level through past interactions with the product category. So, they can make
attention decisions based on this knowledge of the price of the product category. We explore this case by assuming that the price is exogenously given.

6 Conclusion

A nuanced understanding of the mechanism by which advertising messenger leads to persuasion is critical to the success of an advertising campaign. This paper aims to provide a framework for understanding how the advertising messenger can affect the campaign’s effectiveness. We take a dual-mode of communication perspective where the type of messenger is costless to observe while paying attention to other product-related information is costly for consumers. Because the type of messenger is costless to observe and can foretell the signal structure consumers will face upon attention, firms can affect consumers’ attention decisions by committing to a type of messenger. We use a Bayesian persuasion framework to study the role of the messenger where a firm persuades consumers using product match signals, contrary to most of the literature where firms usually have private information and advertising is used to signal firms’ types.

We identify a set of variables pertinent to the choice of messenger - different messengers’ signal structures, the deliberation cost, and the production cost. The messengers’ signal structures determine which type of messenger’s signals are more sensitive to consumers’ product matches, whereas the deliberation and production costs determine a firm’s ability and the need to boost the price. We find that a more informative type of messenger is optimal for the firm when the importance of boosting the price is high because it is more effective for attracting consumer attention and thereby raising the price. In particular, a low-type messenger can be optimal when a high-type messenger would overshadow the product-related information, and the benefits of raising the price outweigh the loss in consumer demand due to consumer deliberation. On the other hand, a more uninformative messenger can be useful when the importance of retaining higher sales is high because it is more capable of dissuading consumer attention even at a higher price. Our results provide managerial implications for firms as to what factors might be important when considering which type of messengers to adopt.

In the main model, we assume that different types of messengers require the same amount of
effort for consumers to process the product-related information in an ad. In reality, a high-type messenger might have several advantages over a low-type one, such as better attracting consumer attention by providing additional entertaining value and converting more consumers when they naively update their beliefs. We discuss these possibilities in the extensions and found that the main insight carries over, except that the regions where a low-type messenger is more effective for boosting the price become smaller. Moreover, we also relax the assumption that consumers can costlessly observe the price. Our main results still hold, except that the deliberation cost does not affect the messenger choice because the firm would either only capture the demand from consumers with zero deliberation costs or set a sufficiently low price to dissuade attention from all consumers.

This paper has proposed some general principles to guide the selection of different types of advertising messengers. However, these guidelines alone do not necessarily account for all the factors determining whether a particular type of messenger would be effective for a specific campaign. Other factors might also affect a firm’s choice of messenger. For example, a firm might choose a less conventional messenger to provoke consumer attention, or a firm can choose a specific type of messenger to appeal to consumers’ values. In addition, advertising messenger can also serve as a signal device for product quality. It would be interesting to explore how the signaling role of advertising can interact with the Bayesian persuasion motives we propose in the current paper.

Finally, we are interested in and motivated by how the messenger affects the persuasion, focusing on the “who is saying it” issue. Nevertheless, the current framework of advertising as a Bayesian persuasion device can be generalized to broader contexts of ad format (“how it is being said”), for example, sophisticated vs. lousy computer graphics, amusing script vs. dry text, etc. However, it is not immediate how those different ad formats can affect the information structures and how credible the commitment power of the firm is. We hope future research can generalize the current framework for different ad contents to understand advertising. We leave those important issues for future research.
Appendix

Proof of Proposition 1.

Note that \( \frac{f(v|m_H,s_g)}{f(v|m_L,s_g)} = \frac{f(v|\sigma_H(v))}{f(v|\sigma_L(v))} \cdot \frac{\tilde{\sigma}_L}{\tilde{\sigma}_H} \). Because \( \frac{\tilde{\sigma}_L}{\tilde{\sigma}_H} \) does not depend on \( v \), \( \frac{f(v|m_H,s_g)}{f(v|m_L,s_g)} \) is increasing in \( v \) if and only if \( \frac{d}{dv} \left( \frac{f(v|\sigma_H(v))}{f(v|\sigma_L(v))} \right) \geq 0 \iff \frac{d}{dv} \left( \frac{\sigma_H(v)}{\sigma_L(v)} \right) \geq 0 \iff \frac{d\sigma_H(v)/dv}{\sigma_H(v)/v} \geq \frac{d\sigma_L(v)/dv}{\sigma_L(v)/v} \) for all \( v \). Conversely, \( \frac{f(v|m_H,s_g)}{f(v|m_L,s_g)} \) is decreasing in \( v \) if and only if \( \frac{d\sigma_H(v)/dv}{\sigma_H(v)/v} < \frac{d\sigma_L(v)/dv}{\sigma_L(v)/v} \) for all \( v \). \( \blacksquare \)

Proof of Lemma 1.

First, when \( p \geq E(v) \), a consumer pays attention to the ad if and only if \( EU(attention|m_j) = \tilde{\sigma}_j \cdot [E(v|m_j,s_g) - p] - c \geq EU(no\ attention) = \max\{0,E(v) - p\} = 0 \iff p \leq \tilde{\sigma}_j E(v|m_j,s_g) - c = E(v|m_j,s_g) - \tilde{\sigma}_j \cdot p \). On the other hand, when \( p < E(v) \), consumers pay attention to the ad if and only if \( \tilde{\sigma}_j \cdot [E(v|m_j,s_g) - p] - c > \max\{0,E(v) - p\} = E(v) - p \iff p > \frac{E(v) - \tilde{\sigma}_j E(v|m_j,s_g) + c}{1 - \tilde{\sigma}_j} \). Because \( E(v) = \tilde{\sigma}_j E(v|m_j,s_g) + (1 - \tilde{\sigma}_j) E(v|m_j,s_b) \), this is equivalent to \( p > E(v|m_j,s_b) + \frac{c}{1 - \tilde{\sigma}_j(v)} \). \( \blacksquare \)

Proof of Lemma 2.

Inequality (10) can be rewritten as \( \int_0^\theta (v - p)[\sigma_H(v) - \sigma_L(v)] \, dF \geq 0 \iff p \int_0^\theta [\sigma_H(v) - \sigma_L(v)] \, dF \leq \int_0^\theta v[\sigma_H(v) - \sigma_L(v)] \, dF \iff p \leq \frac{\int_0^\theta v[\sigma_H(v) - \sigma_L(v)] \, dF}{\int_0^\theta [\sigma_H(v) - \sigma_L(v)] \, dF} \equiv \tilde{p} \). \( \blacksquare \)

Proof of Proposition 2.

When \( \varepsilon_H(v) \geq \lambda \varepsilon_L(v) \) for all \( v \), we have that \( \varepsilon_H(v) \sigma_H(v) \geq \varepsilon_L(v) \sigma_L(v) \iff \frac{d\sigma_H(v)/dv}{\sigma_H(v)/v} \geq \frac{d\sigma_L(v)/dv}{\sigma_L(v)/v} \iff \frac{d\sigma_H(v)/dv}{\sigma_H(v)/v} \geq \frac{d\sigma_L(v)/dv}{\sigma_L(v)/v} \geq 0 \). Also, note that \( \tilde{p} = \int_0^\theta v[\sigma_H(v) - \sigma_L(v)] \, dF \geq \int_0^\theta E(v)[\sigma_H(v) - \sigma_L(v)] \, dF \iff \int_0^\theta [v - E(v)][\sigma_H(v) - \sigma_L(v)] \, dF \geq 0 \).

Thus, when \( \varepsilon_H(v) \geq \lambda \varepsilon_L(v) \) for all \( v \), we can get that \( \int_0^\theta [v - E(v)][\sigma_H(v) - \sigma_L(v)] \, dF \geq \int_0^\theta [v - E(v)][\sigma_H(E(v)) - \sigma_L(E(v))] \, dF = [\sigma_H(E(v)) - \sigma_L(E(v))] \int_0^\theta [v - E(v)] \, dF = 0 \), where the first inequality holds because the positive part of the integral decreases while the negative part of the integral increases. Thus, \( \tilde{p} \geq E(v) \). Conversely, when \( \varepsilon_H(v) < \lambda \varepsilon_L(v) \) for all \( v \), \( \frac{d[\sigma_H(v) - \sigma_L(v)]}{dv} < 0 \), which implies that \( \int_0^\theta [v - E(v)][\sigma_H(v) - \sigma_L(v)] \, dF < 0 \). Therefore, \( \tilde{p} < E(v) \). \( \blacksquare \)

Proof of Proposition 3.

Note that \( c^v = \frac{E(v|m_H,s_g) - E(v|m_L,s_g)}{\tilde{\sigma}_L - \tilde{\sigma}_H} \). First, when \( \varepsilon_H(v) \geq \varepsilon_L(v) \) for all \( v \), \( f(v|m_H,s_g) \) satisfies the MLRP with respect to \( f(v|m_L,s_g) \) from Proposition 1, which implies that \( E(v|m_L,s_g) - \)
\(E(v|m_H,s_g) \leq 0\). Thus, \(c^* \leq 0\), and \(\tilde{p}_H \geq \tilde{p}_L\). Second, when \(\lambda \varepsilon_L(v) \leq \varepsilon_H(v) < \varepsilon_L(v)\) for all \(v, f(v|m_L,s_g)\) satisfies the MLRP with respect to \(f(v|m_H,s_g)\) from Proposition 1, which implies that \(E(v|m_L,s_g) - E(v|m_H,s_g) > 0\) and \(c^* > 0\). Moreover, we have \(c^* \leq \bar{c} = \min_{j \in \{L,H\}} \int_0^\sigma \bar{v}[\sigma_j(v) - \bar{\sigma}_j]dF\) (that is, \(c^*\) is within our range of consideration of \(c\)) if and only if \(\frac{d}{\sigma_H - \bar{\sigma}_L} \frac{\int_0^\sigma \bar{v}[\sigma_j(v) - \bar{\sigma}_j]dF}{\int_0^\sigma \bar{v}[\sigma_H(v) - \sigma_L(v)]dF} \geq \frac{\tilde{p}_L}{\sigma_H - \bar{\sigma}_L} \geq \tilde{p} = \frac{\int_0^\sigma \bar{v}[\sigma_H(v) - \sigma_L(v)]dF}{\sigma_H - \bar{\sigma}_L} \geq \int_0^\sigma \bar{v}dF = E(v)\) which holds when \(\varepsilon_H(v) \geq \lambda \varepsilon_L(v)\) from Proposition 2. So, \(c^* \in (0, \bar{c}]\), and \(\tilde{p}_L > \tilde{p}_H\) if and only if \(c < c^*\). Finally, when \(\varepsilon_H(v) < \lambda \varepsilon_L(v)\), we have \(c^* \geq \bar{c}\) if and only if \(\int_0^\sigma \bar{v}[\sigma_H(v) - \sigma_L(v)]dF < \int_0^\sigma \bar{v}[\sigma_H - \sigma_L]dF\), which holds from Proposition 2. Therefore, \(\tilde{p}_L > \tilde{p}_H\).

**Proof of Proposition 4.**

Note that \(c^{**} = \frac{E(v|m_L,s_b) - E(v|m_H,s_b)}{1 - \bar{\sigma}_L}\). First, when \(\varepsilon_H(v) \geq \lambda \varepsilon_L(v)\), we have \(c^{**} \geq \bar{c}\) (that is, \(c^{**}\) is outside the range of consideration) if and only if \(\frac{(1 - \bar{\sigma}_L) \int_0^\sigma \bar{v}[\sigma_L(v) - \sigma_L]dF - (1 - \bar{\sigma}_L) \int_0^\sigma \bar{v}[\sigma_H(v)]dF}{\bar{\sigma}_H - \bar{\sigma}_L} \geq \min_{j \in \{L,H\}} \int_0^\sigma \bar{v}[\sigma_j(v) - \bar{\sigma}_j]dF \iff \int_0^\sigma \bar{v}[\sigma_H(v) - \sigma_L(v)]dF \geq \int_0^\sigma \bar{v}[\sigma_H - \sigma_L]dF\), which holds according to Proposition 2. Thus, \(\tilde{p}_L > \tilde{p}_H\). Second, when \(\kappa \varepsilon_L(v) \leq \varepsilon_H(v) < \lambda \varepsilon_L(v)\) for all \(v\), we have \(\varepsilon_H(v) \geq \kappa \varepsilon_L(v) \iff \frac{d\varepsilon_H(v)}{dv}[1 - \sigma_L(v)] \geq \frac{d\varepsilon_L(v)}{dv}[1 - \sigma_H(v)] \iff \frac{d}{dv} \left(\frac{1 - \sigma_H(v)}{1 - \sigma_L(v)}\right) \geq 0 \iff \frac{d}{dv} \left(\frac{f(v|m_L,s_b)}{f(v|m_H,s_b)}\right) \geq 0\). That is, \(f(v|m_L,s_b)\) satisfies MLRP with respect to \(f(v|m_H,s_b)\), which implies that \(E(v|m_H,s_b) \geq E(v|m_L,s_b)\) and thus \(c^{**} \geq 0\). And, \(\varepsilon_H(v) < \lambda \varepsilon_L(v)\) implies that \(c^{**} < \bar{c}\). Thus, \(c^{**} \in [0, \bar{c}]\) and we have \(\tilde{p}_L > \tilde{p}_H\) if and only if \(c < c^{**}\). Finally, when \(\varepsilon_H(v) < \kappa \varepsilon_L(v)\) for all \(v\), \(f(v|m_H,s_b)\) satisfies MLRP with respect to \(f(v|m_L,s_b)\), which implies that \(E(v|m_H,s_b) > E(v|m_L,s_b)\). Therefore, \(c^{**} < 0\), and \(\tilde{p}_H \geq \tilde{p}_L\).

**Proof of Proposition 5.**

First, when \(\varepsilon_H(v) \geq \lambda \varepsilon_L(v)\) for all \(v\), \((m_H,\tilde{p}_H)\) is optimal under partial coverage and \((m_L,\tilde{p}_L)\) is optimal under full coverage. Thus, it is optimal to induce partial coverage with \((m_H,\tilde{p}_H)\) if \(\sigma_H[\tilde{p}_H - k] = \tilde{\sigma}_H[E(v|m_H,s_g) - k] - c \geq \tilde{p}_L - k \iff k \geq \frac{1}{\bar{\sigma}_H}E(v|m_H,s_g) + \frac{c}{\bar{\sigma}_L} - k \iff k \geq \frac{1}{\bar{\sigma}_H} \left(\frac{2 - \bar{\sigma}_H}{1 - \bar{\sigma}_L} c + E(v|m_H,s_g) - \sigma_H E(v|m_H,s_g)\right) \equiv k_1(c)\) Otherwise, it is optimal to induce full coverage with \((m_L,\tilde{p}_L)\). Second, when \(\kappa \varepsilon_L(v) \leq \varepsilon_H(v) < \lambda \varepsilon_L(v)\) for all \(v\), it is optimal to induce partial coverage if \(\max_{j \in \{L,H\}} \tilde{\sigma}_j[\tilde{p}_j - k] \geq \max_{j' \in \{L,H\}} \tilde{p}_j - k \iff k \geq \min_{j \in \{L,H\}} \max_{j' \in \{L,H\}} \frac{1}{\bar{\sigma}_j} \left[\frac{2 - \bar{\sigma}_j}{1 - \bar{\sigma}_L} c + E(v|m_{j'},s_b) - \sigma_j E(v|m_{j'},s_g)\right] \equiv k_2(c)\). Otherwise, it is optimal to induce full coverage. Finally, when \(\varepsilon_H(v) < \kappa \varepsilon_L(v)\) for all \(v\), only \((m_H,\tilde{p}_H)\) is optimal under full coverage. Therefore, it is optimal to induce partial coverage if
Proof of Proposition 7.

First, when \( \varepsilon_H(v) \geq \lambda \varepsilon_L(v) \) for all \( v \), \((m_H, \bar{p}_H)\) is optimal under partial coverage and \((m_L, \bar{p}_L)\) is optimal under full coverage. Thus, it is optimal to induce partial coverage with \((m_H, \bar{p}_H)\) if \( \bar{\sigma}_H[\bar{p}_H - k] = \bar{\sigma}_H[E(v|m_H, s_g) - k] \geq \bar{p}_L - k = E(v|m_L, s_b) + \frac{c}{1 - \bar{\sigma}_L} - k \Leftrightarrow k \geq \max \{ \frac{1}{1 - \bar{\sigma}_H} [\frac{2 - \bar{\sigma}_H}{1 - \bar{\sigma}_H} c + E(v|m_H, s_b) - \bar{\sigma}_H E(v|m_H, s_g)] \} \equiv k_4(c). \) Otherwise, it is optimal to induce full coverage with \((m_H, \bar{p}_H)\). The rest of the results follows from resutls 1 and 2.

Proof of Proposition 8.

First, when \( \varepsilon_H(v) \geq \lambda \varepsilon_L(v) \) for all \( v \), only \((m_H, \bar{v})\) is optimal under the no coverage case and only \((m_H, \bar{p}_H)\) is optimal under the partial coverage case. Therefore, no coverage is optimal if \( \alpha \bar{\sigma}_H(\bar{v} - k) \geq \max \{ \bar{\sigma}_H(\bar{p}_H - k), \max_{j \in \{L, H\}} \alpha \bar{\sigma}_j + (1 - \alpha)(\bar{p}_j - k) \} \Leftrightarrow k \geq \max \left\{ \frac{\bar{\sigma}_H E(v|m_H, s_b) - \alpha \bar{\sigma}_H \bar{v} - c}{\bar{\sigma}_H (1 - \alpha)}, \max_{j \in \{L, H\}} \frac{[\alpha \bar{\sigma}_j + (1 - \alpha)] E(v|m_j, s_b) - \alpha \bar{\sigma}_H \bar{v} - c}{\alpha \bar{\sigma}_j + (1 - \alpha) - \alpha \bar{\sigma}_H} \right\} \equiv k_7(c, \alpha). \) And, full coverage is optimal if

\[
\text{max}_{j \in \{L, H\}} \left[ \frac{\alpha \bar{\sigma}_j + (1 - \alpha)] E(v|m_j, s_b) - \alpha \bar{\sigma}_H \bar{v} - c}{\alpha \bar{\sigma}_j + (1 - \alpha) - \alpha \bar{\sigma}_H} \right] \geq \max \{ \bar{\sigma}_H(\bar{p}_H - k), \alpha \bar{\sigma}_H(\bar{v} - k) \} \Leftrightarrow k \leq \max_{j \in \{L, H\}} \left\{ \frac{\alpha \bar{\sigma}_j + (1 - \alpha)] E(v|m_j, s_b) - \alpha \bar{\sigma}_H \bar{v} - c}{\alpha \bar{\sigma}_j + (1 - \alpha) - \alpha \bar{\sigma}_H} \right\} \equiv k_8(c, \alpha).
\]

Next, when \( \varepsilon_H(v) < \lambda \varepsilon_L(v) \) for all \( v \), only \((m_H, \bar{v})\) is optimal under no coverage, so no coverage is optimal if \( \alpha \bar{\sigma}_H(\bar{v} - k) \geq \max \{ \max_{j \in \{L, H\}} \bar{\sigma}_j(\bar{p}_j - k), \max_{j \in \{L, H\}} \alpha \bar{\sigma}_j + (1 - \alpha)(\bar{p}_j - k) \} \Leftrightarrow k \geq \max \left\{ \frac{\bar{\sigma}_j E(v|m_j, s_b) - \alpha \bar{\sigma}_H \bar{v} - c}{\bar{\sigma}_j - \alpha \bar{\sigma}_H}, \max_{j \in \{L, H\}} \frac{[\alpha \bar{\sigma}_j + (1 - \alpha)] E(v|m_j, s_b) - \alpha \bar{\sigma}_H \bar{v} - c}{\alpha \bar{\sigma}_j + (1 - \alpha) - \alpha \bar{\sigma}_H} \right\} \equiv k_9(c, \alpha).

And, full coverage is optimal if \( \text{max}_{j \in \{L, H\}} \left[ \frac{\alpha \bar{\sigma}_j + (1 - \alpha)] E(v|m_j, s_b) - \alpha \bar{\sigma}_H \bar{v} - c}{\alpha \bar{\sigma}_j + (1 - \alpha) - \alpha \bar{\sigma}_H} \right] \geq \max \{ \max_{j \in \{L, H\}} \bar{\sigma}_j(\bar{p}_j - k) \} \geq \max \left\{ \frac{\bar{\sigma}_j E(v|m_j, s_b) - \alpha \bar{\sigma}_H \bar{v} - c}{\bar{\sigma}_j - \alpha \bar{\sigma}_H}, \max_{j \in \{L, H\}} \frac{[\alpha \bar{\sigma}_j + (1 - \alpha)] E(v|m_j, s_b) - \alpha \bar{\sigma}_H \bar{v} - c}{\alpha \bar{\sigma}_j + (1 - \alpha) - \alpha \bar{\sigma}_H} \right\} \equiv k_9(c, \alpha). \)
\[ k, \alpha \hat{\sigma}_H(\bar{v} - k) \implies k \leq \max_{j \in \{L,H\}} \min \left\{ \min_{j' \in \{L,H\}} \left[ \frac{[a\hat{\sigma}_j + (1 - \alpha)]E(v|m_j, s_b) - \hat{\sigma}_j E(v|m_j, s_b) + \frac{1 + (1 - \alpha)(1 - \hat{\sigma}_j)}{1 - \hat{\sigma}_j}c}{(1 - \alpha) + \alpha(\hat{\sigma}_j - \hat{\sigma}_j)} \right] \right\} \equiv k_{10}(c, \alpha). \]

Because as \( \alpha \) increases, the profit increases in the no coverage case \( \left( \frac{\partial \alpha \hat{\sigma}_H(\bar{v} - k)}{\partial \alpha} = \hat{\sigma}_H(\bar{v} - k) > 0 \right) \) and decreases in the full coverage case \( \left( \frac{\partial \alpha \hat{\sigma}_H(\bar{v} - k)}{\partial \alpha} = -(1 - \hat{\sigma}_j)(\bar{p}_j - k) < 0 \text{ when } k < \bar{p}_j \right) \), while the profit is constant in the partial coverage case \( \left( \frac{\partial \hat{\sigma}_j(\bar{p}_j - k)}{\partial \alpha} = 0 \right) \), the regions where no coverage is optimal must be weakly increasing in \( \alpha \) and the regions where full coverage is optimal must be weakly decreasing in \( \alpha \). Therefore, \( \frac{\partial k_{10}(c, \alpha)}{\partial \alpha} \leq 0, \frac{\partial k_{10}(c, \alpha)}{\partial \alpha} \leq 0, \frac{\partial k_{10}(c, \alpha)}{\partial \alpha} \leq 0 \) and \( \frac{\partial k_{10}(c, \alpha)}{\partial \alpha} \leq 0 \). The rest of the results follow from Table 3.

**Proof of Proposition 9.**

First, when \( \varepsilon_H(v) \geq \lambda \varepsilon_L(v) \) for all \( v \), only \( (m_L, E(v|m_L, s_b)) \) can be optimal under full coverage and only \( (m_H, E(v|m_H, s_g)) \) can be optimal under partial coverage. Therefore, partial coverage with \( (m_H, E(v|m_H, s_g)) \) is optimal if and only if \( \beta \hat{\sigma}_H(E(v|m_H, s_g) - k) \geq E(v|m_L, s_b) - k \iff k \geq \frac{E(v|m_L, s_b) - \beta \hat{\sigma}_H E(v|m_H, s_g)}{1 - \beta \hat{\sigma}_H} \equiv k_{11} \). Next, when \( \kappa \varepsilon_L(v) \leq \varepsilon_H(v) < \lambda \varepsilon_L(v) \) for all \( v \), only \( (m_L, E(v|m_L, s_b)) \) can be optimal under full coverage. Therefore, partial coverage is optimal if and only if \( \hat{\sigma}_j(E(v|m_j, s_g) - k) \geq E(v|m_L, s_b) - k \iff k \geq \min_{j \in \{L,H\}} \frac{E(v|m_j, s_b) - \hat{\sigma}_j E(v|m_j, s_g)}{1 - \hat{\sigma}_j} \equiv k_{12} \). Finally, when \( \varepsilon_H(v) < \kappa \varepsilon_L(v) \) for all \( v \), only \( (m_H, E(v|m_H, s_b)) \) can be optimal under full coverage. Therefore, partial coverage is optimal if and only if \( \max_{j \in \{L,H\}} \beta \hat{\sigma}_j(E(v|m_j, s_g) - k) \geq E(v|m_H, s_b) - k \iff k \geq \min_{j \in \{L,H\}} \frac{E(v|m_H, s_b) - \beta \hat{\sigma}_j E(v|m_j, s_g)}{1 - \beta \hat{\sigma}_j} \equiv k_{13} \). The rest of the results follow from Table 4.
References


Online Appendix for “The Role of Messenger in Advertising Content: Bayesian Persuasion Perspective”

A Social Utility and Fandom

In this section, we assume that other than the consumption utility $v_i$, consumers can obtain additional social utilities $\eta > 0$ if the product is endorsed by a high-type messenger. Therefore, when facing a high-type messenger ad, a consumer’s utility of paying attention to the ad is $\bar{\sigma}_H(E(v|m_H, s_\eta) - p + \eta) - c$, and the utility of not paying attention is $\max\{0, E(v) - p - \eta\}$. By comparing the two utilities, consumers pay attention to the ad if and only if $p_H + \eta < p < \bar{p}_H + \eta$. Therefore, given the messenger is high-type, the optimal price under full coverage is $p_H + \eta$ and the optimal price under partial coverage is $\bar{p}_H + \eta$.

Under full coverage, a high-type messenger ad is optimal if and only if $p_H + \eta - k \geq p_L - k \Leftrightarrow c \geq c^* - \frac{\eta}{\bar{\sigma}_H - \bar{\sigma}_L}$, which always holds when $\varepsilon_H(v) < \kappa \varepsilon_L(v)$ for all $v$. Under partial coverage, a high-type messenger ad is optimal if and only if $\bar{\sigma}_H(p_H + \eta - k) \geq \bar{\sigma}_L(p_L - k) \Leftrightarrow k \leq \bar{p} + \frac{\bar{\sigma}_H - \bar{\sigma}_L \eta}{\bar{\sigma}_H - \bar{\sigma}_L}$, which always holds when $\varepsilon_H(v) \geq \lambda \varepsilon_L(v)$ for all $v$. Table 1 summarizes the results about the firm’s optimal messenger and price under each regime.

<table>
<thead>
<tr>
<th>Coverage</th>
<th>$\varepsilon_H(v) \geq \lambda \varepsilon_L(v)$</th>
<th>$\kappa \varepsilon_L(v) \leq \varepsilon_H(v) &lt; \lambda \varepsilon_L(v)$</th>
<th>$\varepsilon_H(v) &lt; \kappa \varepsilon_L(v)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full</td>
<td>$(m_L, \bar{p}_L)$, $c &lt; c^* - \frac{\eta}{\bar{\sigma}_H - \bar{\sigma}_L}$</td>
<td>$(m_H, \bar{p}_H + \eta)$, otherwise $c &lt; c^* - \frac{\eta}{\bar{\sigma}_H - \bar{\sigma}_L}$</td>
<td>$(m_H, \bar{p}_H + \eta)$, otherwise $c &lt; c^* - \frac{\eta}{\bar{\sigma}_H - \bar{\sigma}_L}$</td>
</tr>
<tr>
<td>Partial</td>
<td>$(m_H, \bar{p}_H + \eta)$</td>
<td>$(m_L, \bar{p}_L)$, $k &gt; \bar{p} + \frac{\bar{\sigma}_H - \bar{\sigma}_L \eta}{\bar{\sigma}_H - \bar{\sigma}_L}$</td>
<td>$(m_H, \bar{p}_H + \eta)$, otherwise $c &lt; c^* - \frac{\eta}{\bar{\sigma}_H - \bar{\sigma}_L}$</td>
</tr>
</tbody>
</table>

Table 1: Optimal choice of messenger and price $(m^*_p, p^*_H)$ under full and partial coverage when the product endorsed by a high-type messenger provides additional utilities

**Proposition 1.** 1. When $\varepsilon_H(v) \geq \lambda \varepsilon_L(v)$, it is optimal to induce partial coverage with $(m_H, \bar{p}_H + \eta)$ if $k \geq k_{14}(c)$. Otherwise, full coverage is optimal. Moreover, under full coverage, the firm chooses a low-type messenger with $(m_L, \bar{p}_L)$ if and only if $c < c^* - \frac{\eta}{\bar{\sigma}_H - \bar{\sigma}_L}$.

2. When $\kappa \varepsilon_L(v) \leq \varepsilon_H(v) < \lambda \varepsilon_L(v)$, it is optimal to induce partial coverage if $k \geq k_{15}(c)$. Otherwise, full coverage is optimal. Moreover, under partial coverage, the firm chooses a low-type messenger with $(m_L, \bar{p}_L)$ if and only if $k < \bar{p} + \frac{\bar{\sigma}_H - \bar{\sigma}_L \eta}{\bar{\sigma}_H - \bar{\sigma}_L}$. And, under full coverage, the firm chooses a low-type messenger with $(m_L, \bar{p}_L)$ if and only if $c < c^* - \frac{\eta}{\bar{\sigma}_H - \bar{\sigma}_L}$.

3. When $\varepsilon_H(v) < \kappa \varepsilon_L(v)$, it is optimal to induce partial coverage if $k > k_{16}(c)$. Otherwise, it is optimal to induce full coverage with $(m_H, \bar{p}_H + \eta)$. Moreover, under partial coverage, the firm chooses a low-type messenger with $(m_L, \bar{p}_L)$ if and only if $k > \bar{p} + \frac{\bar{\sigma}_H - \bar{\sigma}_L \eta}{\bar{\sigma}_H - \bar{\sigma}_L}$.

Moreover, all thresholds are increasing in $c$: $\frac{\partial k_{14}(c)}{\partial c} \geq 0$, $\frac{\partial k_{15}(c)}{\partial c} \geq 0$, and $\frac{\partial k_{16}(c)}{\partial c} \geq 0$.

**Proof.** First, when $\varepsilon_H(v) \geq \lambda \varepsilon_L(v)$ for all $v$, only $(m_H, \bar{p}_H + \eta)$ can be optimal under partial coverage. Therefore, partial coverage with $(m_H, \bar{p}_H + \eta)$ is optimal if and only if $\bar{\sigma}_H(p_H + \eta - k) \geq \max\{\bar{p}_H + \eta -
Figure 1: Optimal choice of messenger when the product endorsed by a high-type messenger provides additional utilities

\[ k_p - k \] \iff \ k \geq \max \left\{ \frac{2 - \frac{\sigma}{\eta} c + E(v|m_H,s)}{1 - \sigma_H}, \frac{2 - \frac{\sigma}{\eta} c + E(v|m_L,s)}{1 - \sigma_L} \right\} \equiv k_{14}(c). \]  

Next, when \( \kappa \varepsilon_L(v) \leq \varepsilon_H(v) < \kappa \varepsilon_L(v) \) for all \( v \), partial coverage is optimal if and only if \( \max\{\bar{\sigma}_H(\bar{p}_H + \eta - k), \bar{\sigma}_L(\bar{p}_L - k)\} \geq \max\{p_H + \eta - k, p_L - k\} \iff k \geq \min \left\{ \frac{2 - \frac{\sigma}{\eta} c + E(v|m_H,s)}{1 - \sigma_H}, \frac{2 - \frac{\sigma}{\eta} c + E(v|m_L,s)}{1 - \sigma_L} \right\} \equiv k_{15}(c). \]

Finally, when \( \varepsilon_H(v) < \kappa \varepsilon_L(v) \) for all \( v \), only \((m_H, p_H + \eta)\) can be optimal under full coverage. Therefore, partial coverage is optimal if and only if \( \max\{\bar{\sigma}_H(\bar{p}_H + \eta - k), \bar{\sigma}_L(\bar{p}_L - k)\} \geq p_H + \eta - k \iff k \geq \min \left\{ \frac{2 - \frac{\sigma}{\eta} c + E(v|m_H,s)}{1 - \sigma_H}, \frac{2 - \frac{\sigma}{\eta} c + E(v|m_L,s)}{1 - \sigma_L} \right\} \equiv k_{16}(c). \]  
The rest of the results follow from Table 1.

Figure 1 illustrates Proposition 1. In particular, the results are the same as the main results when \( \eta = 0 \). As \( \eta \) becomes greater, the high-type messenger has the advantage of boosting the price by \( \eta \) both under consumer attention and under no consumer attention. Therefore, the regions where a high-type messenger is optimal becomes larger.

**B Differential Fixed Costs**

So far, we have only discussed the profit generated by different messenger ads. In practice, the cost of featuring a messenger can also affect firms’ advertising decisions. We assume that different messenger ads require different amount of fixed costs \( \phi(m_j) \). In particular, we discuss the case where \( \phi \equiv \phi(m_H) \geq \phi(m_L) = 0 \). For example, featuring a celebrity or an expert would generally be more expensive than featuring normal people.

Under full coverage, a high-type messenger is optimal if and only if \( p_H - k - \phi \geq p_L - k \iff c \geq c^{**} + \frac{\phi}{1 - \sigma_H - 1 - \sigma_L} \), which would never hold when \( \varepsilon_H(v) \geq \lambda \varepsilon_L(v) \) for all \( v \). On the other hand, under partial coverage, a high-type messenger is optimal if and only if \( \bar{\sigma}_H(\bar{p}_H - k) - \phi \geq \bar{\sigma}_L(\bar{p}_L - k) - c \iff
only if \( k \leq \tilde{p} - \frac{\phi}{\sigma_H - \sigma_L} \). Table 2 summarizes the results about the firm’s optimal messenger and price under each regime.

<table>
<thead>
<tr>
<th>Coverage</th>
<th>( \varepsilon_H(v) \geq \lambda \varepsilon_L(v) )</th>
<th>( \varepsilon_H(v) &lt; \lambda \varepsilon_L(v) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full ( (m_L, \overline{p}_L) )</td>
<td>( (m_L, \overline{p}_L), \quad c &lt; c^{**} + \frac{\phi}{1 - \sigma_H - \sigma_L} )</td>
<td>( (m_H, \overline{p}_H), \quad ) otherwise |</td>
</tr>
<tr>
<td>Partial ( (m_L, \overline{p}_L) )</td>
<td>( (m_L, \overline{p}_L), \quad k &gt; \tilde{p} - \frac{\phi}{\sigma_H - \sigma_L} )</td>
<td>( (m_H, \overline{p}_H), \quad ) otherwise |</td>
</tr>
</tbody>
</table>

Table 2: Optimal choice of messenger and price \((m_j^*, p^*)\) under full and partial coverage when different messengers have different fixed costs

**Proposition 2.**

1. When \( \varepsilon_H(v) \geq \lambda \varepsilon_L(v) \), it is optimal to induce partial coverage if \( k \geq k_{17}(c) \).
   
   Otherwise, full coverage with \((m_L, \overline{p}_L)\) is optimal. Moreover, under partial coverage, \((m_L, \overline{p}_L)\) is optimal if and only if \( k > \tilde{p} - \frac{\phi}{\sigma_H - \sigma_L} \).

2. When \( \varepsilon_H(v) < \lambda \varepsilon_L(v) \), it is optimal to induce partial coverage if \( k \geq k_{18}(c) \). Otherwise, full coverage is optimal. Moreover, under partial coverage, \((m_L, \overline{p}_L)\) if and only if \( k > \tilde{p} - \frac{\phi}{\sigma_H - \sigma_L} \).

Also, under full coverage, \((m_L, \overline{p}_L)\) is optimal if and only if \( c < c^{**} + \frac{\phi}{1 - \sigma_H - \sigma_L} \).

Moreover, all thresholds are increasing in \( c \): \( \frac{\partial k_{17}(c)}{\partial c} \geq 0 \) and \( \frac{\partial k_{18}(c)}{\partial c} \geq 0 \).

**Proof.** First, when \( \varepsilon_H(v) \geq \lambda \varepsilon_L(v) \) for all \( v \), only \((m_L, \overline{p}_L)\) can be optimal under full coverage. Therefore, partial coverage is optimal if \( \max \{\tilde{\sigma}_H(E[v|m_H, s_h] - k) - c - \phi, \tilde{\sigma}_L(E[v|m_L, s_g] - k) - c\} \geq E[v|m_L, s_h] - k \Leftrightarrow k \geq \min \left\{ \frac{c + \phi + E[v|m_L, s_h] - \tilde{\sigma}_H E[v|m_H, s_h] - k}{1 - \sigma_H}, \frac{c + E[v|m_L, s_g] - \tilde{\sigma}_L E[v|m_L, s_g] - k}{1 - \sigma_L} \right\} \equiv k_{17}(c) \).

And, when \( \varepsilon_H(v) < \lambda \varepsilon_L(v) \) for all \( v \), partial coverage is optimal if \( \max \{\tilde{\sigma}_L(E[v|m_H, s_h] - k) - c - \phi\} \geq \max \{E[v|m_L, s_h] - k, E[v|m_H, s_h] - k - \phi\} \), which holds if and only if \( k \geq \min \left\{ \frac{c + \phi + E[v|m_L, s_h] - \tilde{\sigma}_L E[v|m_L, s_h]}{1 - \sigma_L}, \frac{E[v|m_L, s_h] - \tilde{\sigma}_L E[v|m_L, s_h]}{1 - \sigma_L} \right\}, \max \left\{ \frac{c + E[v|m_H, s_h] - \tilde{\sigma}_L E[v|m_H, s_h]}{1 - \sigma_L}, \frac{E[v|m_H, s_h] - \tilde{\sigma}_L E[v|m_H, s_h]}{1 - \sigma_L} \right\} \equiv k_{18}(c) \). The rest of the results follow from Table 2.

Figure 2 illustrates Proposition 2. Note that when \( \phi = 0 \), the results are exactly the same as the main results. As \( \phi \) becomes larger, a high-type messenger becomes less favorable for the firm to generate profits, so the regions where a high-type messenger is optimal become smaller.

**C Exogenous pricing**

In many situations, the prices of different products are similar within a product category. For example, most mobile apps are priced at \$0.99, and consumers know the price level through past interactions with the product category. So, they can make attention decisions based on this knowledge of the price of the product category. We explore this case by assuming that the price is exogenously given. This assumption allows us to set aside the price signaling issue and focus on the role of the messenger cleanly as a main medium of information.
When the price is exogenously given, the profit margin $p - k$ is fixed, so the firm's goal is to maximize the demand for the product $D(p, m_j)$. When $p > E(v)$, consumers' default action is to not purchase the product, so attracting consumer attention is beneficial for capturing part of the demand from those who receive private good signals. However, when $p \leq E(v)$, consumers' default action is to purchase the product, and attracting consumer attention can be harmful for the firm, because some consumers would opt to not purchasing the product upon deliberation, thereby lowering the demand. We characterize the conditions under which a low-type messenger ad can be optimal in the proposition below.

**Proposition 3.**

1. When $E(v) < p$, it is optimal for the firm to attract consumer attention. In this case, a low-type messenger ad generates a higher profit if $p_H < p \leq p_L$, which is possible only if (1) $\varepsilon_H(v) < \lambda \varepsilon_L(v)$ for all $v$ or (2) $\lambda \varepsilon_L(v) \leq \varepsilon_H(v) < \varepsilon_L(v)$ for all $v$ and $c < c^*$.  

2. When $E(v) \geq p$, it is optimal for the firm to dissuade consumer attention. In this case, a low-type messenger ad generates a higher profit if $p_H \leq p < p_L$, which is possible only if (1) $\varepsilon_H(v) \geq \lambda \varepsilon_L(v)$ for all $v$ or (2) $\lambda \varepsilon_L(v) \leq \varepsilon_H(v) < \lambda \varepsilon_L(v)$ for all $v$ and $c < c^{**}$.  

**Proof.** First, when $E(v) < p$, attracting consumer attention is optimal. Thus, a low-type messenger can be optimal when $p_H < p \leq p_L$, in which case the price is too high for consumers to pay attention to a high-type messenger, but still low enough for them to pay attention to a low-type one. From Proposition 3, this is possible only when (1) $\varepsilon_H(v) < \lambda \varepsilon_L(v)$ for all $v$ or (2) $\lambda \varepsilon_L(v) \leq \varepsilon_H(v) < \varepsilon_L(v)$ for all $v$ and $c < c^*$.  

Next, when $E(v) \geq p$, dissuading consumer attention is optimal. Thus, a low-type messenger can be optimal when $p_H \leq p < p_L$, in which case the price is high enough for consumers to pay attention to the high-type messenger, but not high enough for them to pay attention to a low-type one. From Proposition 4, this is possible only when (1) $\varepsilon_H(v) \geq \lambda \varepsilon_L(v)$ for all $v$ or (2) $\lambda \varepsilon_L(v) \leq \varepsilon_H(v) < \varepsilon_L(v)$ for all $v$ and $c < c^{**}$.  

Finally, when $\max_{j \in \{L,H\}} p_j < p \leq \min_{j \in \{L,H\}} p_j$, both types of messenger can induce consumer attention. However, a high-type messenger ad is always optimal, since more consumers will receive private good signals upon paying attention to a high-type messenger ad and end up purchasing the product: $\tilde{\sigma}_H > \tilde{\sigma}_L$.  

Figure 3 illustrates the proposition 3. The dark grey regions are where a high-type messenger
ad generates a higher profit, and the light grey regions are where a low-type messenger ad generates a higher profit. In the white regions, because consumers either never pay attention or always pay attention regardless of the type of the messenger, the firm is indifferent between a high-type and a low-type messenger ad.

To get a clearer intuition, we discuss the two extreme cases (1) $\varepsilon_H(v) \geq \varepsilon_L(v)$ for all $v$ and (2) $\varepsilon_H(v) < \kappa \varepsilon_L(v)$ for all $v$. First, when a high-type messenger’s signals are sufficiently informative such that $\varepsilon_H(v) \geq \varepsilon_L(v)$ for all $v$, a high-type messenger is more able to attract consumer attention, making them useful when the price is high such that consumers’ default action is to not purchase the product. Therefore, when the deliberation cost $c$ is not too large, it is optimal to use a high-type messenger ad to attract consumer attention when the price is high, and dissuade consumer attention with a low-type messenger ad when the price is low.

On the other hand, when a high-type messenger overshadows the product information such that $\varepsilon_H(v) < \kappa \varepsilon_L(v)$ for all $v$, a low-type messengers signals can update consumers’ beliefs more significantly, making them more able to attract consumer attention. Thus, when the deliberation cost $c$ is not too large, it is optimal to use a low-type messenger ad to attract consumer attention when the price is high, and dissuade consumer attention with a high-type messenger ad when the price is low.

Figure 3: Optimal choice of messenger when pricing is exogenous