Information and Motivated Reasoning: A Model of Selective Exposure[*]

Tom (Hyeon Seok) Yu[†]

Last Update: February 8, 2022

Abstract

Previous research has consistently documented the prevalence of selective exposure, the tendency to seek out information that reinforces preexisting beliefs. Modeling individuals as motivated reasoners who face a tradeoff between the accuracy ("getting it right") and the directional ("reaching desired conclusions") motives, this paper develops a game-theoretic model that makes sense of seemingly inconsistent empirical findings by formally identifying conditions under which individuals, as receivers, engage in selective or cross-cutting exposure. When the information quality is uniform across individuals, selective exposure remains pervasive even in situations where the accuracy motive is high. Second, introducing uncertainty to the sender's directional motive lowers the likelihood of exposure. Finally, the size of the gap in the perceived quality of information between the sender and the receiver, rather than the high credibility of the sender, largely determines the possibility of cross-cutting exposure. These results on selective exposure yield direct implications for persuasion and polarization.

[*] I thank Neil Malhotra, Ken Shotts, Avidit Acharya, Steve Callander, and Dana Foarta for their helpful guidance and feedback.

[†] Stanford University Graduate School of Business. Email: tom.hs.yu@stanford.edu

Information assumes an integral role in any decision-making process, and the domain of politics has been no exception. As such, from principal-agent problems to electoral competitions, information has been featured prominently in canonical models of political economy. While most classical models presume that information is processed when provided (e.g., updating one's belief using Bayes's rule), individuals often face a preceding choice on whether to process information as given or, as regularly observed, ignore it. Indeed, ever since Lazarsfeld et al. (1948)'s study of the presidential election in 1940, scholars have long noted people's tendency to seek out information that confirms preexisting beliefs while dismissing contradictory ones. Among different accounts on why individuals exhibit such behavior, the motivated reasoning framework considers the tradeoff between "getting it right" (the accuracy motive) and "desiring a certain conclusion to hold" (the directional motive) in people's minds as a central driver (Kunda 1990).

As an example, suppose we have an individual, who cares deeply about family values – loyalty, for instance – and plans to vote in the upcoming election. With her party affiliation as a Republican, she might already have an idea on whom to select, but she comes across an unfavorable news headline about the Republican candidate: the latter has repeatedly engaged in extramarital affairs. The voter does not want to be "wrong" by voting for a person with questionable values. Meanwhile, her identity as a Republican might motivate her to believe that the news article/source is not trustworthy and that her preferred candidate upholds the same value that she cherishes. How would such a tradeoff between accuracy and directional motives affect the voter's exposure decision to the given information? Would it matter if she does not recognize the information source? And what if her perception of the source's credibility differs by its ideological slant? Addressing these questions on exposure is of particular importance to topics on persuasion and polarization, as exposure to information often necessarily precedes a change in beliefs, which has a direct implication on divergence.

Building on the motivated reasoning framework, this paper develops a game-theoretic model whose chief objective is to provide a systematic account of selective exposure. More specifically, it produces predictions on exposure decisions in a two-stage game where individuals initially form beliefs on a given issue based on an exogenous signal, and they can gather or avoid action-relevant information from senders in the subsequent period. In addition to holding heterogeneous directional motives, individuals are assumed to exhibit different degrees of commitment (i.e., the strength of conviction) to their motives. While the model can be generally applied to settings where strategic considerations affect the exchange and consumption of information, the bulk of exposition equates the sender to larger information sources like news providers and the receiver to individuals with actions to take like voters for the ease of connection to related empirical literature.

The main contribution of this paper is its ability to make sense of seemingly inconsistent and disconnected stylized facts. On selective exposure, the model identifies the topic/issue as a critical part of individuals' exposure decisions. In particular, the strength of one's directional motive and the degree of confidence in one's own information associated with a given topic determine whether an individual engages in selective or cross-cutting exposure. This result clarifies the seemingly mixed evidence of partisan selective exposure in the experimental and observational studies: the former generally documents a stronger pattern of selective exposure among partisans (e.g., Taber et al. 2009; Knobloch-Westerwick and Meng 2009; Benedictis-Kessner et al. 2019; Peterson and Iyengar 2021), while unobtrusive studies outside the laboratory settings find that even strong partisans from both sides significantly overlap in the selection of news providers (Gentzkow and Shapiro 2011; Guess 2021) and exposure to cross-cutting contents (Bakshy et al. 2015). The key difference between the two sets of studies is the range of topics considered. The experimental studies often employ a handful of salient partisan issues over which respondents likely hold strong directional motives, but observational studies often lack the control over topics to which individuals expose themselves, meaning the latter likely reflects patterns over a much wider range of political issues. Mummolo (2016)'s experimental study that shows topic relevance eclipsing the negative partisan cue when it comes to information source selection further highlights the importance of topics.[1]

Overall, the model proposed yields the following set of predictions on selective exposure and corresponding implications for persuasion and polarization. The baseline model, which assumes the identical quality of information across individuals and the sender's directional motive to be known, reveals the prevalence of selective exposure. In particular, the weight on the accuracy and directional motives determine the type of equilibrium sustainable, and in each, individuals are predicted to engage in different types of selective exposure. This in turn implies the difficulty of persuasion, which has been regularly documented in empirical studies (e.g., Guess et al. 2021; Peterson et al. 2021), since receivers do not find it appealing to expose themselves to information that contradicts either their preexisting beliefs or directional motives when they know that senders' information is no better than theirs. A notable implication is that selective exposure can impede, but not prevent, divergence in beliefs through avoidance. Put another way, while selective exposure does lead to polarization, individuals' decision to avoid even confirmatory information can lower the degree of divergence, which is similar in spirit to Arceneaux et al. (2012)'s finding that individuals' "tuning out" of political news blunts oppositional media hostility.

Though it is likely that people often know the motives held by those who share information with them, there exist situations where the source's directional motive might not be known (e.g., news headlines from unknown media outlets on social media). Extension 1 considers this effect of uncertainty on selective exposure to account for such a possibility. In contrast to an existing theory from Zaller (1992) that posits the increased likelihood of exposure with an unknown source, the model predicts uncertainty to decrease the likelihood

[1] As deliberated in the empirical discussion section of the baseline model, a similar line of reasoning can make sense of the seemingly inconsistent set of experimental results on partisan cheerleading (Bullock et al. 2015; Prior et al. 2015; Peterson and Iyengar 2021).

of exposure. The discrepancy results from the possibility of incurring a loss from exposing oneself to a source with opposite directional motives. This seems consistent with empirical findings that show individuals' tendency to prefer familiar sources over unrecognized ones in experiments (Iyengar and Hahn 2009) and concentrated visits to selected news providers despite many options available to users online (Schmidt et al. 2017; Peterson et al. 2021).

Lastly, an extension considers the setting where the sender's information quality is different from that of the receiver. Specifically, it constructs two different environments where the receiver perceives information quality to be higher either for all senders or only those with aligned directional motives. The analysis reveals the "gap" in the perceived quality of information between the sender and the receiver, rather than the high quality of sender's information, as the key driver of exposure decisions. While a large gap necessarily requires the sender's information quality to be high, if the receiver's information is good enough, the latter might choose to avoid information no matter how credible the sender is. The second construction allows analyzing the role of oppositional media distrust on polarization, and the result shows that unfavorable perception alone may not be sufficient to induce divergence in beliefs. That is, as long as the aligned source sends unbiased signals to the receiver, convergence is possible, but this means a source that only relays a biased set of information can result in polarization (Levendusky 2013; Martin and Yurukoglu 2017).

Related Literature

This section first situates the paper in relation to existing formal models and highlights its difference and contribution. Then, it briefly discusses existing theories' predictions on selective exposure and corresponding implications for persuasion and polarization.

Models of Information Processing and Motivated Reasoning

It is important to note that the focus of the model proposed is not on how individuals process information once consumed, but the exposure decision that precedes it. That said, individuals in the model still need to process information, and it is similar to the Bayesian learning models in this regard; individuals update in a standard Bayesian manner without bias should they choose to do so (Gerber and Green 1999; Bullock 2009). The model differs from these existing models in that individuals have an option of rejecting information as if they have never seen it. Furthermore, even after exposure, they can choose not to update their posterior beliefs, two actions of which are both behaviorally plausible and empirically observed (e.g., Zaller 1992; Barnes et al. 2018).

Given its reliance on the motivated reasoning framework stemming from Kunda (1990), the model here is much more closely related to existing formal models of motivated reasoning (Little 2019; Little 2022). In particular, the structure is similar to Little (2019)'s model of belief formation, where individuals also face the same tradeoff between the accuracy and the directional motives. A more closely related model is from Little (2022), whose extension allows the "rejection" of information, one of the key actions taken by individuals when making exposure decisions in the current model. An important difference between the model proposed and these existing models pertains to the game-theoretic nature of the current endeavor.[2] Modeling exposure decisions, which precede belief formation, in a signaling game set up not only captures the strategic nature of information sharing but also allows one to explicitly model the sender's behavior, whose action is just as important as that of the receiver when it comes to persuasion (Kamenica and Gentzkow 2011). In addition, Little (2022) identifies understanding information-seeking behavior as another apt venue for applying the motivated reasoning framework, and this paper precisely takes this next step on individuals' exposure decisions.

[2] Another existing work that develops a game-theoretic model is Bracha and Brown (2012), but their analysis focuses on cognitive processes in an "intrapersonal" setting, unlike the current model that analyzes exposure decisions in an interpersonal setting.

Motivated Reasoning, Selective Exposure, and Avoidance

Why do individuals show a tendency to seek out confirmatory information while avoiding contradictory information? Among many strands of research that grapple with this question, the motivated reasoning account based on Kunda (1990) has been particularly influential in political science. Its central observation that people are motivated not just to be correct, but also by the innate desire to arrive at certain conclusions has been applied in both theoretical (e.g., Little 2019; Bénabou and Tirole 2016) and experimental studies addressing related questions (e.g., Redlawsk 2002; Taber et al. 2009; Knobloch-Westerwick and Meng 2009; Prior et al. 2015; Peterson and Iyengar 2021).

One of the major predictions from existing information processing theories pertains to exposure. Both the motivated reasoning framework and the theory of cognitive dissonance from Festinger (1957) generally predict the dominance of confirmation bias: when given a choice, people will seek out information confirming their preexisting beliefs while dismissing contradictory ones. Zaller (1992) and Taber and Lodge (2006), in particular, provide similar accounts on the exposure decision and its corresponding effect on attitudes. However, both presume exposure, and the option of "avoidance altogether" is not explicitly modeled. Unlike laboratory settings where respondents are usually forced to select different types of information without the possibility of not choosing at all, people have relatively strong control over their information exposure given the myriad of information sources available in the contemporary media landscape (Prior 2013). Considering that avoidance is an exposure decision that can directly affect beliefs, accounting for such a possibility constitutes an important gap to fill.[3]

[3] It is important to clearly define "avoidance." Here and throughout this paper, avoidance is equated to "rejection" when information is provided, which is different from not seeking out information in the first place. As information is provided to individuals in this model, avoidance necessarily means rejection.

Persuasion and Polarization

An immediate implication of selective exposure is persuasion, as in whether exposure to information can change beliefs and induce associated actions. Existing theories have identified specific conditions under which persuasion can be feasible. In Zaller (1992)'s Receive-Accept-Sample ("RAS") model, for example, persuasion to counter-attitudinal message occurs if (1) an individual is not politically sophisticated and (2) the source of information is unknown. The motivated reasoning framework of Taber and Lodge (2006) emphasizes "weak prior attitudes" as an important condition. Aside from formally analyzing the relevance of these conditions, the model here provides more systematic accounts of when persuasion is feasible.

The possibility of change in beliefs further yields implications for polarization. Both Zaller (1992) and Taber and Lodge (2006) predict that even balanced news (without partisan slant) can polarize attitudes through selective exposure to like-minded information. While some empirical studies find results consistent with this prediction (e.g., Levendusky 2013), Arceneaux et al. (2012)'s experimental study that allows avoidance of political information finds that individuals often tune out political programming altogether, which in turn blunts the polarizing effect of slanted information. Put another way, as noted by Benedictis-Kessner et al. (2019), the relationship between selective exposure and polarization may not be as straightforward as some existing studies have theorized (e.g., Stroud 2010), and analyzing the current model reveals subtleties that might have been overlooked.

A Model of Selective Exposure

The primary objective at hand is to construct a model that can account for individuals' information-seeking/avoidance behavior. In particular, the game-theoretic model developed here relies on the motivated reasoning framework of Kunda (1990) to explain why and how people engage in selective exposure: given some state of the world ("SOW") $w \in \{0, 1\}$, individuals in the model care not just about accuracy (i.e., taking an action that matches the SOW), but also their directional motives (i.e., taking an action congruent with their identities). These sometimes conflicting motives drive their decisions on exposure to information and updating beliefs, which have direct implications for their susceptibility to persuasion and polarization in beliefs.

While the model can be applied in a broad set of settings, suppose, for the sake of exposition, we have an individual i who needs to take an action (e.g., vote) based on (1) information about the SOW she gathers through exposure to others' opinions and (2) her "type," which determines one's directional motive.[4] Individuals are endowed with a binary type represented by $k_i \in \{0, 1\}$, assumed to be equiprobable and private, but its distribution is common knowledge. k_i also marks i's preference over beliefs about the SOW. For example, $k_i = 0$ might represent being a Republican, and in an electoral setting, such a party affiliation might drive her to believe a Republican candidate to be the right choice even if the information she holds might suggest otherwise.

In addition to the heterogeneity in types, individuals can exhibit varying degrees of conviction based on their directional motives. A natural example would be "strong" and "weak" partisans, who show different levels of commitment to their partisan ideals. One way of capturing this heterogeneity is by allowing the tradeoff between the accuracy and the directional motive to differ. Specifically, define $m_i \in \{m_H, m_L\}$ as an indicator for whether one holds the high or low degree of conviction. The high-type individuals incur greater psychological costs if they take actions that do not align with their directional motives.

But where do individuals obtain their information in this model? The sequence of the game below describes the process:

1. Period 0: Endowment Phase

   (a) Nature determines the SOW (w), assumed to be equiprobable, and its distribution is common knowledge.

[4] This section employs a female pronoun for receivers and a male pronoun for senders for the ease of exposition.

(b) Nature determines the directional motive (k_i) with $\Pr(k_i = 0) = \Pr(k_i = 1) = \frac{1}{2}$ and the degree of conviction (m_i) with $\Pr(m_i = m_H) = \rho \in [0, 1]$ for all individuals.

(c) Individuals receive independently drawn informative but imperfect signal $s_i \in \{0, 1\}$ with accuracy $q \in (\frac{1}{2}, 1)$.

(d) Based on their signals, individuals form initial posterior beliefs about the SOW $\mu^0_{(s_i,w)}$ and select the intended action $x^0_i \in \{0, 1\}$.

2. Period 1: Exposure and Action Phase

(a) Individuals retain information on k_i, s_i, and x^0_i from the endowment phase, all of which are private knowledge.

(b) A sender j, chosen at random, selects a signal $\pi_j \in \{0, 1\}$ for a receiver i. The sender does not know who the receiver will be when choosing his signal for the latter.

(c) Upon observing the signal π_j, the receiver i can choose whether to expose herself to j's information. Define her action as $a_{ij} \in \{0, 1\}$.

- If $a_{ij} = 1$, the receiver subsequently decides whether to update her belief about the SOW; define $b_i \in \{0, 1\}$ as the belief update decision, and if $b_i = 1$, her posterior gets updated from $\mu^0_{(s_i,w)} \rightarrow \mu^1_{(\pi_j,w)}$, using Bayes's rule.
- If $a_{ij} = 0$, the receiver ignores the sender's signal, and both her intended action and posterior beliefs remain unchanged.

(d) Based on her exposure and belief update decisions, i selects her final action $x^1_i \in \{0, 1\}$.

Before delving into specific utility structures that drive individuals' actions, there are at least two aspects of the information in the model that deserve clarifications. First, how should we think about the "exogenously" provided initial signal s_i? This can be perceived as an initial cue with a slant. Consider, for instance, a candidate with low name recognition. An advertisement for the candidate shared over social media platforms could serve as the initial signal.[5] Second, the intended action at period 0 can be construed as an "inclination" that an individual forms for herself. Mechanically, it becomes a reference point when she makes her exposure decision later in period 1.

Individual Utilities and Actions

The influence of the motivated reasoning framework on this model is most evident in the construction of individuals' utilities. For individual i,

$$u_i(x_i) = -\lambda_i \mathbb{1}(x_i \neq k_i) - (1-\lambda_i)\mathbb{1}(x_i \neq w) \tag{1}$$

where $\lambda_i \in [0,1]$ is the key weight variable that represents the aforementioned trade-off between accuracy and directional motives. By design, a higher λ_i means a greater weight on the directional motive as opposed to the accuracy motive. Naturally, for the highly-committed individuals ($m_i = m_H$), $\lambda_H > \lambda_L$, reflecting the greater loss from an action that contradicts their directional motives. Then, how is an individual's specific action as a sender and as a receiver associated with one's utilities? Below specifies utility functions for each period and actor:[6]

1. Period 0:

$$u_i^0(x_i^0, s_i) = -\lambda_i \mathbb{1}\{x_i^0(s_i) \neq k_i\} - (1-\lambda_i)\mathbb{1}\{x_i^0(s_i) \neq w\} \tag{2}$$

[5] As shown in a number of empirical analyses of online platforms, individuals do often get exposed to a cross-cutting set of opinions and news without their conscious decisions to do so (e.g., Yang et al. 2020; Bakshy et al. 2015).

[6] The description below does not distinguish the utility for the degree of conviction; for the high- and the low-type individuals, simply replace λ_i with λ_H and λ_L, respectively.

Individual i, knowing her own type (k_i), forms posterior beliefs on the SOW based on the signal s_i and selects x_i^0 that minimizes the loss. Note that all individuals are myopic: when selecting their initial actions, they do not take possible actions in the subsequent periods into account.

2. Period 1:

   - Sender utility:

$$u_{S,j}^1(\pi_j, a_{ij}) = -\lambda_i \mathbb{1}\{\pi_j \neq k_j\} - (1-\lambda_i)\mathbb{1}\{\pi_j \neq w\} - \mathbb{1}\{a_{ij} = 0 | \pi_j\} \tag{3}$$

     A sender j suffers loss if he (1) sends a signal π_j that does not correspond to his own directional motive and if (2) it does not match the SOW. Finally, the last term represents the relational damage in the case of rejection by the receiver.

   - Receiver utility:

     - Exposure decision (a_{ij}):

$$u_{R,i}^1(\pi_j, a_{ij}) = \begin{cases} -\lambda_i \mathbb{1}\{k_j \neq k_i\} - (1-\lambda_i)\mathbb{1}\{\pi_j \neq w\}, & \text{if } a_{ij} = 1 \\ -\lambda_i \mathbb{1}\{x_i^0 \neq k_i\} - (1-\lambda_i)\mathbb{1}\{x_i^0 \neq w\}, & \text{otherwise.} \end{cases} \tag{4}$$

     - Belief update decision (b_i, requires $a_{ij} = 1$):

$$u_{R,i}^1(\pi_j, b_i) = \begin{cases} -\lambda_i \mathbb{1}\{\pi_j \neq k_i\} - (1-\lambda_i)\mathbb{1}\{\pi_j \neq w\}, & \text{if } b_i = 1 \\ -\lambda_i \mathbb{1}\{x_i^0 \neq k_i\} - (1-\lambda_i)\mathbb{1}\{x_i^0 \neq w\}, & \text{otherwise.} \end{cases} \tag{5}$$

     Receiver i's utility depends on her exposure decision a_{ij} and belief update decision b_i. The only difference between the two utility functions is the directional motive portion. Depending on the sender's type and message, if $k_j \neq \pi_j$, then the receiver may choose to expose herself to sender's information but does not update her belief.[7]

- Final action (x_i^1):

$$u_i^1(x_i^1) = -\lambda_i \mathbb{1}\{x_i^1 \neq k_i\} - (1-\lambda_i)\mathbb{1}\{x_i^1 \neq w\} \tag{6}$$

  Finally, after sending and receiving information take place, i selects x_i^1 based on her final posterior beliefs and her type to minimize the loss.

Given a comparatively nontraditional setup of the game, a number of modeling decisions call for justifications. First, beginning with the sender's utility, why does his utility depend on the receiver's exposure decision (a_{ij}) but not the update decision (b_i)? The observability is the issue: it is easier to check whether someone has allowed herself to be exposed to the given information (e.g., clicking on a given news article online) than to confirm persuasion. Besides, as a sender, while persuasion might be important, rejection/avoidance by the receiver should incur sufficient cost, as it likely precludes attention and persuasion.

On the receiver's utility, why does one incur loss from exposing herself to the opposite type? This, in part, captures the strong emotional response individuals often exhibit toward the "out-group members," a prominent example of which includes strong partisans in the US (see Iyengar et al. 2019 for a summary). For instance, a long-time Democrat might incur psychological discomfort from being exposed to a news clip from a right-leaning medium. The separation of the exposure and the belief update decisions might appear odd at first, but this reflects an empirical regularity that exposure does not always imply persuasion (e.g., Knobloch-Westerwick and Meng 2009; Barnes et al. 2018).

[7] The construction excludes the possibility of updating one's belief while choosing to avoid information (i.e., setting $b_i = 1$ after choosing $a_{ij} = 0$). In other words, avoidance precludes persuasion, which is not entirely unrealistic.

Modeling Assumptions

Bayesian information processing In this model, individuals update their beliefs based on new information in a Bayesian manner, should they choose to do so. This might at first seem far from reality, as examples of biased information processing abound (e.g., Rabin 1998; Lord et al. 1979). However, when it comes to factual information, existing studies by Gerber and Green (1999), Bullock (2009), and Hill (2017) find that, albeit imperfect, individuals do update their beliefs in ways consistent with the Bayesian learning models. Besides, though certainly important, the focus of the model is not about the role of bias in information processing, but the role of motivated reasoning in exposure decisions, which precede information processing.

No backfire effect? Among many, one set of behavior commonly observed in empirical studies yet not allowed in the model pertains to the backfire effect, where exposure to counter-attitudinal information reinforces existing beliefs rather than correcting it (Nyhan and Reifler 2010). Taber and Lodge (2006)'s motivated reasoning model includes a prediction consistent with such an effect, but it is omitted in this model based on recent empirical investigations by Wood and Porter (2019), who find tenuous evidence of the backfire effect even on highly partisan issues. Furthermore, the model's primary focus on the exposure decision renders this aspect of information processing of secondary importance.

Uniform quality of information The current baseline assumes that individuals receive signals with identical quality at period 0 (i.e., q is not heterogeneous), which might be unrealistic. Indeed, especially when the source of information is known to the receiver, individuals likely perceive the quality of information to be different across sources. As an individual's faith in the source accuracy can directly affect one's exposure and belief update decisions, an extension in the later section considers the effect of allowing (1) either uniformly greater perceived accuracy of senders or (2) greater faith in the congruent information source but lack of it for oppositional sources.

Selective Exposure with Known Information Source

The baseline of the model assumes that the sender's directional motive is known to the receiver. An example of applicable settings includes the receiver obtaining action-relevant information (e.g., candidate quality before voting) from a long-time friend who has access to information of similar quality. Given the construction that builds on the tradeoff between accuracy and directional motives, the weight variable λ_i largely determines the type of equilibrium. Specifically, the magnitude of λ_i relative to the threshold $\bar{\lambda}^0$ drives individuals' behavior upon receiving the signal s_i at period 0.[8]

Definition 1 (Period 0 Threshold, $\bar{\lambda}^0$)
There exists a threshold $\bar{\lambda}^0 = \frac{2q-1}{2q}$ below which individuals prefer to select the intended action $x_i^0 = s_i$ regardless of their types k_i. Conversely, individuals choose $x_i^0 = k_i$ regardless of the initial signal s_i if $\lambda_i > \bar{\lambda}^0$.

Simply put, if λ_i is sufficiently low, one selects the intended action x_i^0, one's "inclination," identical to the signal received even if the latter does not match one's directional motive. This intended action is crucial, as it serves as a point of comparison when deciding on exposure at the subsequent stage. The type of pure-strategy equilibrium sustainable based on this threshold is as follows, and the solution concept is perfect Bayesian.[9]

Definition 2 (Types of Pure-Strategy Equilibrium)

[8] An important point to remember here is that individuals are myopic, so when selecting their intended actions at period 0, they do not take their actions in subsequent periods into account.

[9] There also exist mixed-strategy equilibria at knife-edge cases where the low or high-type (m_i) individuals' weight variable matches the threshold. These cases are omitted given their lack of substantive contribution to the discussion.

- *Accuracy equilibrium:*

  The equilibrium in which all individuals, regardless of their true types k_i, select $x_i^0 = s_i$ at period 0 and set $\pi_j = s_j$ as senders at period 1.

- *Directional motive equilibrium:*

  The equilibrium in which all individuals, regardless of signals received at period 0 s_i, select $x_i^0 = k_i$, and set $\pi_j = k_j$ as senders at period 1.

- *Separating equilibrium:*

  The equilibrium in which the high-type individuals ($m_i = m_H$) behave as in the directional motive equilibrium, while the low-type ($m_i = m_L$) behave as in the accuracy equilibrium.

That the magnitude of λ_i determines the type of equilibrium sustainable is intuitive: if individuals care a lot about the accuracy on a given issue (i.e., low λ_i), for instance, they will likely heed to their initial signals. Moreover, when sharing information, they are incentivized to share what they believe to be correct, as they know that others place a greater weight on accuracy as well. The converse holds for the directional motive equilibrium, and the separating equilibrium accounts for the case when the high- and the low-type individuals diverge. Before delving into the result, it is helpful to discriminate the type of selective exposure individuals, as receivers, exhibit.

Definition 3 (Types of Selective Exposure)

- *Prior Selective Exposure: a receiver i chooses to expose herself to the sender j's message π_j if and only if $\pi_j = s_i$.*

- *Direction Selective Exposure: i chooses to expose herself to π_j if and only if $\pi_j = k_i$.*

- *Complete Selective Exposure: i chooses to expose herself to π_j if and only if $\pi_j = s_i$ and $\pi_j = k_i$.*

Most existing works equate selective exposure to prior selective exposure in the context of confirmation bias; individuals seek or expose themselves to information that confirms their preexisting beliefs. Different types of exposure are possible in this model, as it allows an individual to hold a prior conflicting with her directional motive, a segment of the population that appears understudied. Note, by definition, only an individual who receives a signal that aligns with her directional motive at period 0 can engage in complete selective exposure, the most restrictive form of exposure that only accepts both prior and directionally aligned messages from a sender.

Result 1: Individuals Always Engage in a Type of Selective Exposure

Put another way, individuals do not expose themselves to information that contradicts both their priors and directional motives in a setting where everyone receives an equally informative signal at period 0. The magnitude of λ_i determines the type of equilibrium, in which individuals exhibit different types of selective exposure. Formal characterizations and proofs of all propositions are relegated to the Appendix.

Proposition 1 (Selective Exposure when Information Source is Known)

- *Accuracy equilibrium ($\lambda_L < \lambda_H < \bar{\lambda}^0$): if λ is sufficiently low ($\lambda_H < \bar{\lambda}_H^1 \equiv \frac{q-3q^2+2q^3}{-1+3q-5q^2+3q^3}$), all individuals engage in prior selective exposure.*
- *Directional motive equilibrium ($\bar{\lambda}^0 < \lambda_L < \lambda_H$): if λ is sufficiently high ($\bar{\lambda}^0 < \lambda_L$), all individuals engage in direction selective exposure.*
- *Separating equilibrium ($\lambda_L < \bar{\lambda}^0 < \lambda_H$):*
  - *High-types (m_H):*

    if $k_i = s_i$, individuals only engage in complete selective exposure; if $k_i \neq s_i$ and λ_H is sufficiently low ($\lambda_H < \bar{\lambda}_{H,1}^1 \equiv \frac{1-3q+3q^2-2q^3+q\rho-3q^2\rho+2q^3\rho}{-q+q^2-2q^3-q\rho-q^2\rho+2q^3\rho}$), individuals engage in both prior and direction selective exposure.

- *Low-types (m_L):*

  if λ_L is sufficiently low ($\lambda_L < \bar{\lambda}_L^1 \equiv \frac{q-3q^2+2q^3}{-1+3q-5q^2+3q^3}$), all individuals engage in prior selective exposure.

The conditions specified above place additional constraints on the magnitude of λ_i. While most results seem intuitive, the result on the accuracy equilibrium might not appear obvious. Why do individuals avoid information in the accuracy equilibrium? That is, if they do care a lot about getting the SOW right, why would they reject additional information? Simply put, the sender's quality of information is not good enough. Consider the case when $\lambda_i = 0$ (i.e., individuals solely care about the accuracy, which would be close to standard Bayesian learning models that do not consider directional motives). Even in such an extreme case, a receiver rejects a sender's message that does not confirm her prior because she knows that his information quality is as just good as her own, and accepting a contradictory message means increased uncertainty, as it pushes her posterior back to the prior.

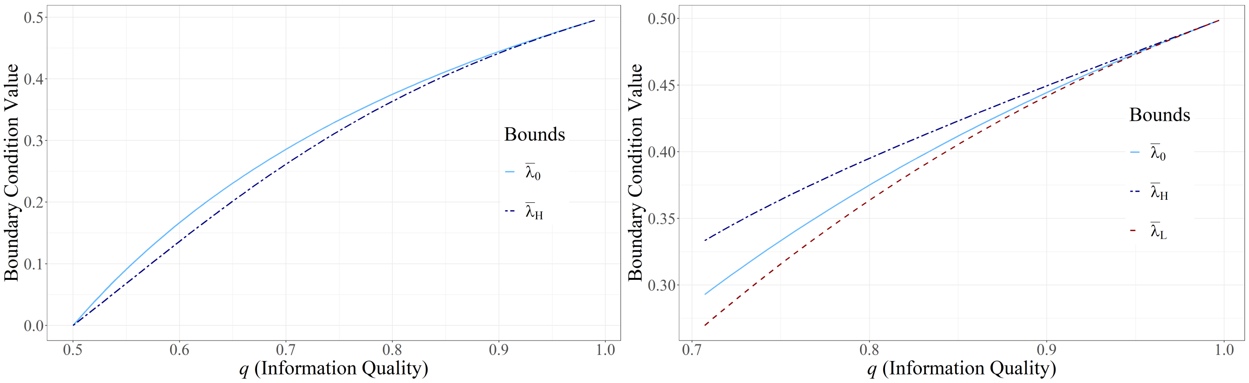
Another seemingly odd result pertains to the conflicted (i.e., $s_i \neq k_i$) high-type individuals exposing themselves more to different types of messages even compared to the low-types. This results from their conflicted prior and the "just-right" magnitude of λ_i. The latter leads them to expose themselves to the aligned sender ($k_j = k_i$) even if the message itself contradicts one's prior ($\pi_j \neq s_i$), as she cares sufficiently enough about her directional motive. Conversely, with an additional upper bound on λ_i, she also accepts a prior-confirming message ($\pi_j = s_i$) from a misaligned sender ($k_j \neq k_i$), as it reduces uncertainty on the SOW. For the low-type individuals, the analogous line of reasoning from the accuracy equilibrium applies. Then, what do these results on exposure imply about the possibility of persuasion?

Corollary 1 (Persuasion when Information Source is Known)

If q is identical between senders and receivers, persuasion is only possible among conflicted high-type individuals ($k_i \neq s_i \wedge m_i = m_H$), which further requires $\lambda_H \in (\bar{\lambda}^0, \bar{\lambda}_H)$, where $\bar{\lambda}_H \equiv \frac{1-3q+3q^2-2q^3+q\rho-3q^2\rho+2q^3\rho}{-q+q^2-2q^3-q\rho-q^2\rho+2q^3\rho}$).

Note that the persuasion here means not just updating beliefs, but also being convinced to take an action that contradicts (1) one's preexisting belief (i.e., initial posterior) or (2) the intended action x_i^0 from period 0. The corollary states that the conflicted high-type individuals, who receive a signal s_i that does not align with their types k_i, are the only ones capable of being persuaded. Moreover, the sender's message must be aligned with either the receiver's directional motive to persuade the latter. A close gap between the boundary conditions on λ_i in figure 1 (b), however, shows the difficulty of exposure, hence persuasion, to take place. In particular, a necessary condition for the conflicted high-types to expose themselves to a directionally misaligned but prior-confirming message is $\lambda_H \in (\bar{\lambda}_0, \bar{\lambda}_H)$ (a gap between a dash-dot and a solid line in the right panel).

Figure 1: Comparative Statics on the Boundary Conditions by Equilibria



(a) Accuracy Equilibrium

(b) Separating Equilibrium

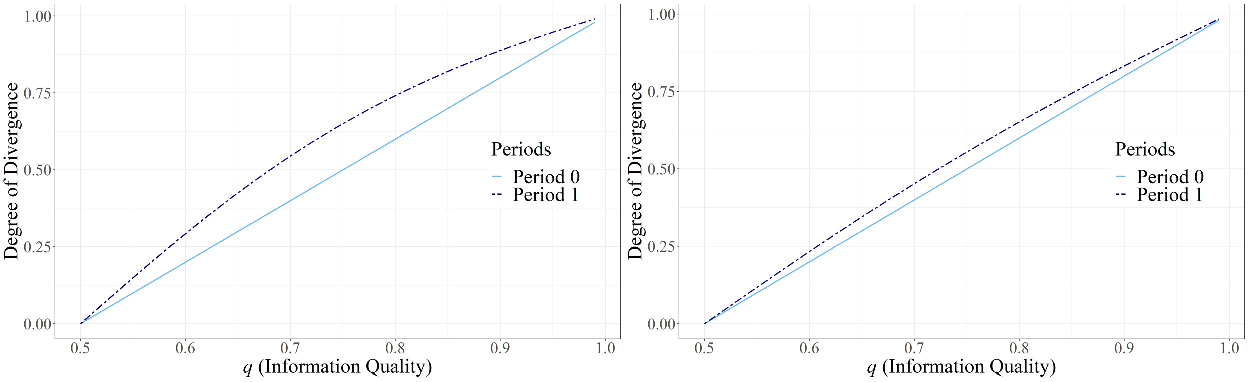
Note: As all boundary conditions are upper bounds, λ_H and λ_L need to be below the dot-dash and dashed curves, respectively, for each of the accuracy and the separating equilibrium to be sustainable as conjectured.

Result 2: Selective Exposure Can Impede Polarization in Beliefs

The first result shows the prevalence of selective exposure and the difficulty of persuasion. What do these results, then, imply about the polarization in beliefs? This section carries out a simple analysis that addresses this question in a setup where a representative set of receivers' expected posterior beliefs after one interaction with a sender are computed to measure the degree of divergence ex-post expected exposure decisions.[10]

Figure 2 shows the comparison of the expected degree of divergence in the accuracy and the separating equilibria. As indicated in the subtitles, the expected divergence is actually greater in the accuracy rather than the separating equilibrium. Figure 6 in the Appendix, which depicts the actual positions of expected posteriors after one interaction, reveals that this lower degree of divergence results from individuals avoiding exposure – namely, those who received a message that aligned with their directional motives in the first period (i.e., $k_i = s_i$). Accordingly, the separation of the low- and the high types posterior beliefs shown in figure 7 in the Appendix reveals that the least amount of divergence can be expected among the high-types, especially those who are only predicted to engage in complete selective exposure.

Figure 2: Degree of Divergence in Posterior Beliefs at Period 1



(a) Accuracy Equilibrium (0.587)

(b) Separating Equilibrium (0.540)

[10] More concretely, the degree of divergence is computed as the difference between the average of those who receive $s_i = 1$ – types $(k_i, s_i) = (1, 1), (0, 1)$ – at period 0 and the corresponding average of those who receive $s_i = 0$ – types $(k_i, s_i) = (1, 0), (0, 0)$. Then, on the belief that $w = 1$, the initial gap at period 0 is: $q - (1 - q) = 2q - 1$, which is plotted as solid lines in figure 2.

Empirical Discussion: Persuasion, Partisan Cheerleading, Avoidance

Among different types of selective exposure explored in the analysis, the strong influence of preexisting beliefs theorized in earlier works is noteworthy (Rabin and Schrag 1999). Except for the directional motive equilibrium in which exposure requires the alignment of directional motives, a receiver generally exposes herself to a sender's message if it matches her prior belief in most equilibria. In addition, the slim possibility for persuasion stated in corollary 1 reflects the general difficulty of persuading others when the quality of information is uniform across individuals. Indeed, recent studies that analyze the online news consumption behavior generally find that exposure does not lead to modification of preexisting beliefs (Peterson et al. 2021; Guess et al. 2021).

Albeit not explicitly modeled, a behavior that can be explained from the model is partisan cheerleading, in which partisans intentionally express views revealed to be wrong. The evidence on whether individuals genuinely believe in given misinformation and adopt inaccurate beliefs or they are merely exhibiting partisan cheerleading while actually knowing it to be wrong seems mixed: Bullock et al. (2015) and Prior et al. (2015) find financial incentives to be sufficient in suppressing partisan cheerleading on factual questions, while Peterson and Iyengar (2021) report an opposite result that financial incentives do not close the partisan gap in responses to factual questions. If one takes the "response to factual questions" as the main action (x_i) in this framework, the seeming inconsistency might be attributable to the difference in the set of questions/topics to which individuals assign different weights (λ_i). In Bullock et al. (2015)'s experiment, the financial incentives likely lowered λ_i (i.e., increased the accuracy motive) and induced respondents to truthfully state their beliefs, whereas λ_i might have been simply too high for the respondents to not engage in partisan cheerleading in Peterson and Iyengar (2021)'s study.[11]

Third, the analysis on the implication of selective exposure on polarization in beliefs calls

[11] Taking a closer look at the set of questions adopted in the two studies, it appears Peterson and Iyengar (2021) relied on questions where partisan respondents might assign greater

for a subtler version of an existing claim that selective exposure leads to polarization (e.g., Stroud 2010). While it is true that most types of selective exposure reinforce preexisting beliefs, which in turn lead to divergence in beliefs, avoidance of information that often accompanies selective exposure can impede such a divergence. This result is in a similar spirit to that of Arceneaux et al. (2012), whose experimental study shows that avoidance of political information blunts oppositional media hostility. Benedictis-Kessner et al. (2019)'s experiment that employs both forced exposure and free choice designs also speaks to the relevance of avoidance: partisan media has a particularly strong effect on attitudes among "inadvertent audiences" (some of whom include partisans) who would otherwise not consume it in the free-choice setting. This indirectly shows how avoidance can impede divergence.

Extension 1: Unknown Source of Information

What if the sender's directional motive is not known to the receiver? In the real world, especially online where individuals' choice set in terms of news provider is large, it is possible that individuals do not know the ideological leaning or directional motive of a given information source. This section reanalyzes the model assuming that the receiver might not observe or know the sender's type, which means the receiver now needs to infer both the SOW and the sender's type based on the sender's message.

Result 3: Uncertainty Renders the Exposure Less Likely

Removing the visibility of the sender's directional motive makes the receiver weakly more likely to avoid exposure even to the confirmatory information. While the general pattern

weights on directional motives (higher λ_i); the maximum differences in the partisan divide in the control groups of the two studies are 0.52 and 0.24, respectively. As noted by the authors themselves, the "psychic" rewards for cheerleading might have exceeded financial rewards in their studies.

across equilibria remain largely unchanged from the baseline, the possibility of rejection increases for both the high- and the low-type individuals.

Proposition 2 (Selective Exposure when Information Source is Unknown)

- *Accuracy equilibrium* ($\lambda_L < \lambda_H < \bar{\lambda}^0$)*: if* λ_i *is sufficiently low* ($\lambda_H < \frac{2q-6q^2+4q^3}{-1+4q-8q^2+4q^3}$)*, all individuals engage in prior selective exposure.*

- *Directional motive equilibrium* ($\bar{\lambda}^0 < \lambda_L < \lambda_H$)*: if* λ_i *is sufficiently high* ($\bar{\lambda}^0 < \lambda_L$)*, all individuals engage in direction selective exposure.*

- *Separating equilibrium :* ($\lambda_L < \bar{\lambda}^0 < \lambda_H$)

  - *High-types* (m_H)*:*

    if $k_i = s_i$*, individuals always avoid exposure; if* $k_i \neq s_i$ *and* λ_H *is sufficiently low* ($\lambda_H < \bar{\lambda}_H \equiv \frac{2-6q+6q^2-4q^3-\rho+4p\rho-6q^2\rho+4q^3\rho}{1-4q+4q^2-4q^3-\rho+2q\rho-4q^2\rho+4q^3\rho}$)*, individuals engage in prior selective exposure.*

  - *Low-types* (m_L)*:*

    if λ_L *is sufficiently low* ($\lambda_L < \bar{\lambda}_L \equiv \frac{2q-6q^2+4q^3}{-1+4q-8q^2+4q^3}$)*, all individuals engage in prior selective exposure.*

Except the directional motive equilibrium where removing the certainty over the sender's directional motive does not have any bite, the reduced likelihood of exposure is attributable to the increased potential loss from exposing oneself to the "wrong" type of senders. For example, a receiver who has seen a confirmatory signal at period 0 might not expose herself even to a confirmatory message from a sender due to the potential loss from exposing oneself to the opposite type eclipsing the gain in posterior belief on the SOW. This is evident in the depiction of the boundary condition of the accuracy equilibrium in figure 3: unlike its counterpart in the baseline, the increase in q does not render the equilibrium more likely. It peaks at around 0.75 and then decreases, which results from the fact that individuals' signals at period 0 are "already good enough," and the additional confirmatory signal from a sender is insufficient to risk potentially exposing oneself to the opposite type.

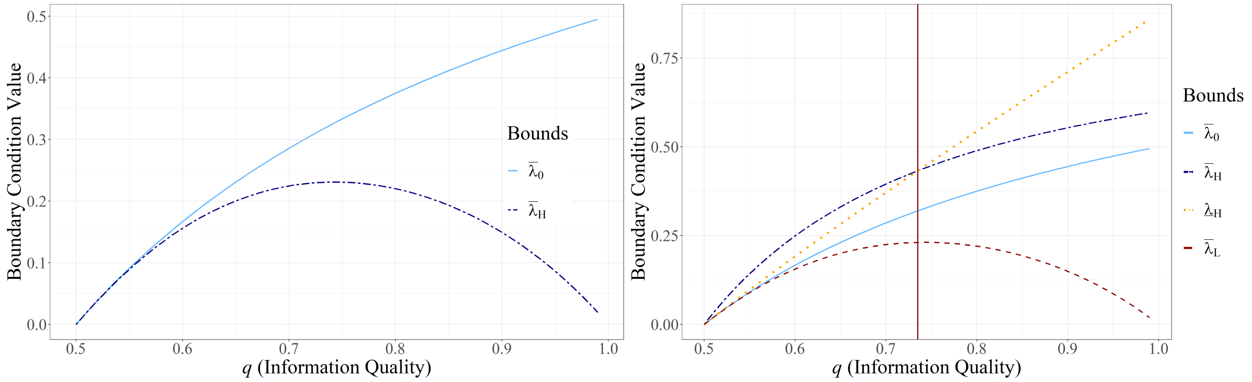
In the case of the separating equilibrium, a similar line of reasoning applies to the conflicted high-type individuals, who were most prone to exposure in the baseline. As shown in figure 3 (b), the condition of engaging in prior selective exposure is stringent, reflected as a small area between $\bar{\lambda}_H$ and $\underline{\lambda}_H$. Perhaps unsurprisingly, such results on the lower likelihood of exposure translate to the difficulty of persuasion.

Corollary 2 (Persuasion when Information Source is Unkown)
If the sender's directional motive (k_j) is unknown, the persuasion is only possible among conflicted high-type individuals ($k_i \neq s_i \wedge m_i = m_H$), which further requires $\lambda_H \in (\underline{\lambda}_H, \bar{\lambda}_{H,2})$, where $\underline{\lambda}_H \equiv \frac{-4q+8q^2-\rho+4q\rho-4q^2\rho}{4q}$ and $\bar{\lambda}_{H,2} \equiv \frac{2-6q+6q^2-4q^3-\rho+4q\rho-6q^2\rho+4q^3\rho}{4\rho q^3-4q^3-2\rho q^2+2q^2-2q}$.

In the baseline, the sender having the same quality of information as the receiver prevented exposure to contradictory information. This corollary on persuasion shows that uncertainty does not render exposure more feasible. The conflicted high-type individuals no longer expose themselves to prior-contradicting information, and they can only be persuaded to take a different action following exposure to prior-confirming information. While they might have selected directionally-congruent action at period 0, their posterior belief after prior-confirming information from a sender can move sufficiently farther to induce them to take an action consistent with their prior (but inconsistent with their directional motives). Nevertheless, a comparison shows $\bar{\lambda}_{H,2} < \bar{\lambda}_H$, which attests to the low feasibility of persuasion even among such individuals.[12]

[12] Figure 9 in the Appendix illustrates this tight boundary.

Figure 3: Comparative Statics on the Boundary Conditions of Extension 1



(a) Accuracy Equilibrium

(b) Separating Equilibrium

Note: In figure (b), the vertical line marks the point at which the lower bound $\underline{\lambda}_H$ exceeds the upper bound $\bar{\lambda}_H$, thereby rendering the separating equilibrium no longer sustainable as conjectured. The given boundary conditions assume $\rho = \frac{1}{2}$.

Empirical Discussion: the Role of Uncertainty

The stated result on the negative effect of uncertainty in sender's directional motive on exposure contrasts with a prediction from Zaller (1992)'s RAS model, which argues that not knowing the source of information can increase the likelihood of exposure to counter-attitudinal information, as individuals will be less likely to question the credibility of the source. The discrepancy in the prediction of the current model results from the potential loss from exposing oneself to someone or source with opposite directional motives; even if the sender's message confirms the receiver's prior beliefs or is consistent with her directional motives, the possibility of the sender being an opposite type makes her balk.

Iyengar and Hahn (2009) find a consistent result with this analysis on uncertainty: "the presence of a news organization label increases the appeal of news stories across all subject matter dimensions," meaning when given a choice, respondents preferred a familiar information source over those unrecognized. Observational studies on social media news consumption also reflect this preference for known sources. Schmidt et al. (2017), for example, show that despite the myriad of news providers available on social media platforms, users often limit their exposure to a selected few news providers.

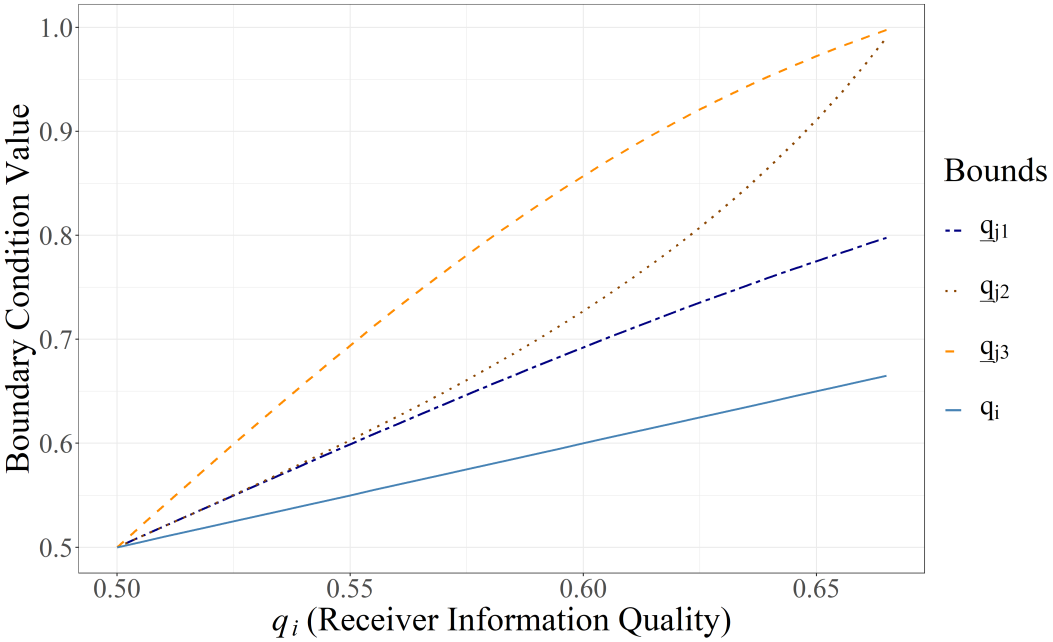
Extension 2: Difference in Perceived Quality of Information

The baseline assumed an equal quality of information among individuals. But there certainly exist cases where the receiver might believe the sender's information quality to be better (i.e., a higher probability of getting the SOW right). An applicable case would be an individual deciding on whether to click on a news article headline from a source she perceives to be credible. This extension allows the assessment of existing claims on the effect of the information quality and the lack of faith in "oppositional" media on selective exposure. The receiver is assumed to know the sender's directional motive as in the baseline, and the analysis focuses on the accuracy equilibrium.

Proposition 3 (Selective Exposure when Senders are Uniformly Credible)
If the gap between the receiver and the sender's information quality is sufficiently large ($q_i < \frac{2}{3}, q_j > \frac{-q_i+3q_i^2}{2-6q_i+6q_i^2}$), the receiver always expose herself to the sender's message.

At first, this might appear as a straightforward confirmation of the existing claim that the higher quality of information increases the likelihood of exposure. The statement, however, is more nuanced, as it emphasizes the "gap" between the accuracy of the receiver's information and that of the sender. Put another way, for a receiver to expose oneself to all types of information, strong faith in the sender's information may not be sufficient. This is evident in figure 4 that shows different lower-bounds on the sender's information to sustain the equilibrium. As shown, once the receiver's information accuracy surpasses a certain point (0.67 in this setup), she no longer exposes herself to contradictory information from those with contradictory directional motives (i.e., $\pi_j \neq s_i$ and $k_j \neq k_i$). In other words, if the receiver knows that her information quality is good enough, she is not willing to expose herself to uncomfortable information that suggests she had a wrong idea, especially from a source known to have opposite directional motives.

Figure 4: Comparative Statics on Boundary Conditions given Senders Uniformly Credible



Note: The solid line represents the receiver's information quality (q_i) for comparison. $\underline{q}_{j1}$ represents the lower bound on the sender's information quality for a receiver to be willing to expose herself a contradictory message ($\pi_j \neq s_i$) from an aligned source ($k_j = k_i$). $\underline{q}_{j2}$ represents the lower bound for an exposure to a confirmatory message ($\pi_j = s_i$) from a misaligned source ($k_j \neq k_i$). $\underline{q}_{j3}$ represents the lower bound for an exposure to a contradictory message ($\pi_j \neq s_i$) from a misaligned source ($k_j \neq k_i$).

Perceiving Oppositional Media as Not Credible

As existing studies have shown, people's perception of the information provider's credibility can vary, and one's ideological or partisan alignment with a given source is often predictive of the former's perception of the latter (Arceneaux and Johnson 2013). Therefore, another natural way to model heterogeneity in the quality of information is by taking the oppositional media into account. More specifically, this part of the analysis now assumes that if the given source's directional motive differs from that of the receiver (i.e., $k_j \neq k_i$), the receiver perceives the sender's information quality to be poor, hence not credible (i.e., $q_j < \frac{1}{2} < q_i$). On the contrary, if the media source shares the same directional motive, the receiver continues to perceive the source's message to be more informative than her own signal (i.e.,

$q_{j,k_j=k_i} > q_i > \frac{1}{2}$). Perhaps unsurprisingly, adopting this structure returns the result similar to the directional motive equilibrium, except that individuals update their beliefs on the SOW upon exposure to directionally-aligned sources.

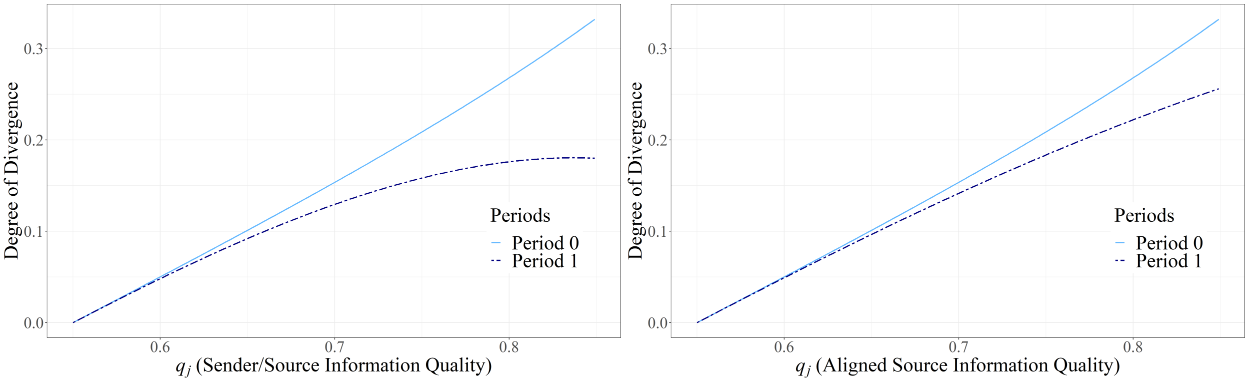
Proposition 4 (Selective Exposure when Opposite Types are Not Credible)
If the receiver perceives the opposite type sender's information quality to be low ($q_{j,k_j \neq k_i} < \frac{1}{2}$, while $q_{j,k_j=k_i} > q_i > \frac{1}{2}$), the receiver engages in direction selective exposure even in the accuracy equilibrium.

Result 4: Avoiding Oppositional Source Not Sufficient for Divergence

This part now addresses whether the receiver perceiving the information quality of the opposite sources to be low leads to polarization in beliefs. Figure 5, which also shows the degree of divergence at period 1 for the first construction assuming the uniformly greater quality of information for comparison, reveals that avoiding information from misaligned sources is not sufficient to induce polarization: the expected degree of divergence at period 1 curve (dotted-dash) remains below that of the initial period.

Compared to the construction where the receiver perceives all types of senders to be credible, the degree of convergence is noticeably lower when she considers the opposite type's information quality to be poor. Nevertheless, the fact that the posteriors are expected to converge even when individuals only expose themselves to information from aligned sources might be surprising. The convergence results from the possibility of an aligned source sending a contradictory message that does not abide by their types. If, however, the sender is biased and only shares the aligned message, the posteriors can diverge relative to the initial period.

Figure 5: Degree of Divergence in Posterior Beliefs at Period 1 in Extension 2



(a) q_j Uniformly Greater (0.121)

(b) Opposite Types Not Credible (0.143)

Empirical Discussion: Source Credibility, Topics, Oppositional Media

This extension analyzed the effect of differentiating the sender's information quality by setting it either uniformly greater than that of the receiver or greater only in the case of directional alignment. While it mostly discussed q_j as information quality, a more apt interpretation for empirical applications might be "perceived source credibility," as it is unlikely that individuals keep statistical records on how accurate a given source has been. Rather, people are more likely to assess the quality of given information based on the credibility of its source, especially when they are not familiar with a given topic (e.g., Petty and Cacioppo 1986). What the analysis adds is that high perceived credibility alone is not sufficient to induce exposure. The receiver's perception of her own information quality should be low, meaning the perceived "gap" is the key.

But this result on the gap in perceived information quality constitutes just one necessary factor for cross-cutting exposure. As the analysis revolved around the accuracy equilibrium, a relatively low directional motive is another necessary condition. A critical point to note is that both the information quality gap and the weight variable can be topic-specific. That is, for an issue on which the receiver believes her information to be fairly accurate and holds a strong directional motive, the model shows that exposure to contradictory information and/or directionally misaligned sources is unlikely. The topic-specific nature of the exposure decision brings clarity to the seemingly mixed set of evidence of partisan selective exposure from experimental (e.g., Taber et al. 2009; Knobloch-Westerwick and Meng 2009; Benedictis-Kessner et al. 2019; Peterson and Iyengar 2021) and unobtrusive observational studies (e.g., Gentzkow and Shapiro 2011; Guess 2021; Bakshy et al. 2015). An important difference is the possibility of selecting specific topics in the experimental settings, while the observational works mostly analyze the general news consumption patterns online without direct control over the topics.[13]

The analysis also yields a notable result on whether perceiving the oppositional media to be not credible can lead to polarization. Such perception impedes convergence of beliefs, and as shown in Peterson and Kagalwala (2021)'s survey experiments, selective exposure can sustain the unfavorable perception. However, the analysis suggests that avoiding information from the oppositional media alone may not result in divergence of beliefs. As long as the aligned media sources relay truth most of the time (i.e., be willing to share information that goes against one's directional motive), the divergence may not be severe.

Concluding Remarks

This paper proposes a game-theoretic model that identifies specific conditions under which individuals engage in selective exposure when they are assumed to be motivated reasoners. Returning to our hypothetical voter from the introduction, the model predicts that she will likely reject the damaging information about her preferred candidate (i.e., avoid exposure)

[13] In particular, those experimental studies that find strong patterns of selective exposure appear to have selected topics over which respondents might hold strong directional motives. Peterson and Iyengar (2021), for instance, adopt noticeably partisan issues such as voter fraud, immigrant crime, and Obama wiretap. A number of other experimental works is carried out during the election cycle (Knobloch-Westerwick and Meng 2009) and relies on topics related to candidates in presidential elections (Iyengar et al. 2008).

if one of the following holds: (1) her directional motive (λ_i) is particularly high for the given election/candidate, (2) the news source is not known to her, or (3) if the source is known, her perceived gap in the quality of information about the candidate between herself and the news provider is not large. Among these conditions, the first and the third factors shed light on the seeming inconsistencies in the empirical literature on selective exposure and partisan cheerleading. Individuals presumably show different degrees of directional motive and confidence in their information compared to those from other sources for different issues or topics. Then, the discrepancy among empirical findings on the prevalence of selective exposure and partisan cheerleading might be attributable to the difference in the set of topics considered in these studies.

Analyzing the model reveals two notable subtleties in the relationship between selective exposure and polarization. First, certain types of selective exposure can impede, but not prevent, divergence in beliefs. Although selective exposure generally widens the gap, individuals' decision to avoid/reject even prior-confirming information can translate to no update in beliefs, thereby preventing the divergence. This seems consistent with empirical findings that report partisans preferring to expose themselves to apolitical information (e.g., entertainment shows) even when given an option to select a source with aligned ideological leanings or prior-reinforcing information (Arceneaux et al. 2012; Benedictis-Kessner et al. 2019). Next, the extension that considers heterogeneity in information quality reveals that the receiver perceiving the oppositional source's quality of information to be poor is not sufficient to induce polarization. Indeed, beliefs are expected to converge as long as the preferred sources are not biased in their sharing of information, but this is why findings on bias in news coverage by prominent media are concerning (Levendusky 2013; Martin and Yurukoglu 2017).

There are several ways in which this game-theoretic approach to explaining selective exposure can be extended. On the theoretical front, an important question remains as to what determines the trade-off parameter λ_i.[14] That is, which factors would lead one to place a heavier weight on accuracy over directional motives, or vice versa? Some evidence of suppressing partisan cheerleading behavior using financial incentives suggests the relevance of contexts (e.g., Bullock et al. 2015), but there could be more concrete factors such as individuals' perception of stake associated with a given issue: what could one lose from getting the action wrong? Solidifying the parameter's micro-foundations would be beneficial not just for this particular model, but for the motivated reasoning framework as a whole.

One important factor not explicitly modeled in this paper is the topic salience. The relevance of an issue plays a critical role in one's information consumption decision (Entman 1989; Mummolo 2016). The current model implicitly assumes that an issue calling for action is either salient or important enough for an individual to seek information, but the salience of a given topic often exerts sizeable influence over exposure decisions. Aside from salience, there likely exist factors that await analysts' exploration as the next steps. This paper will have served its purpose if such an endeavor materializes to extend our understanding of exposure decisions.

[14] This is related to the idea of endogenizing the function v, the directional motive, in Little (2019)'s framework.

References

Arceneaux, Kevin and Martin Johnson (2013). *Changing Minds or Changing Channels?: Partisan News in an Age of Choice*. University of Chicago Press.

Arceneaux, Kevin, Martin Johnson, and Chad Murphy (2012). "Polarized Political Communication, Oppositional Media Hostility, and Selective Exposure". *The Journal of Politics* 74 (1): pp. 174–186.

Bakshy, Eytan, Solomon Messing, and Lada A. Adamic (2015). "Exposure to ideologically diverse news and opinion on Facebook". *Science* 348 (6239): pp. 1130–1132.

Barnes, Lucy, Avi Feller, Jake Haselswerdt, and Ethan Porter (2018). "Information, Knowledge, and Attitudes: An Evaluation of the Taxpayer Receipt". *The Journal of Politics* 80 (2): pp. 701–706.

Bénabou, Roland and Jean Tirole (2016). "Mindful Economics: The Production, Consumption, and Value of Beliefs". *Journal of Economic Perspectives* 30 (3): pp. 141–164.

Benedictis-Kessner, Justin De, Matthew A. Baum, Adam J. Berinsky, and Teppei Yamamoto (2019). "Persuading the Enemy: Estimating the Persuasive Effects of Partisan Media with the Preference-Incorporating Choice and Assignment Design". *American Political Science Review* 113 (4): pp. 902–916.

Bracha, Anat and Donald J. Brown (2012). "Affective decision making: A theory of optimism bias". *Games and Economic Behavior* 75 (1): pp. 67–80.

Bullock, John G. (2009). "Partisan Bias and the Bayesian Ideal in the Study of Public Opinion". *The Journal of Politics* 71 (3): pp. 1109–1124.

Bullock, John G., Alan S. Gerber, Seth J. Hill, and Gregory A. Huber (2015). "Partisan Bias in Factual Beliefs about Politics". *Quarterly Journal of Political Science* 10 (4): pp. 519–578.

Entman, Robert M. (1989). "How the Media Affect What People Think: An Information Processing Approach". *The Journal of Politics* 51 (2): pp. 347–370.

Festinger, Leon (1957). *A Theory of Cognitive Dissonance*. Stanford University Press.

Gentzkow, Matthew and Jesse M. Shapiro (2011). "Ideological Segregation Online and Offline". *The Quarterly Journal of Economics* 126 (4): pp. 1799–1839.

Gerber, Alan and Donald Green (1999). "Misperceptions About Perceptual Bias". *Annual Review of Political Science* 2 (1): pp. 189–210.

Guess, Andrew M. (2021). "(Almost) Everything in Moderation: New Evidence on Americans' Online Media Diets". *American Journal of Political Science* 65 (4): pp. 1007–1022.

Guess, Andrew M., Pablo Barberá, Simon Munzert, and JungHwan Yang (2021). "The consequences of online partisan media". *Proceedings of the National Academy of Sciences* 118 (14).

Hill, Seth J. (2017). "Learning Together Slowly: Bayesian Learning about Political Facts". *The Journal of Politics* 79 (4): pp. 1403–1418.

Iyengar, Shanto and Kyu S Hahn (2009). "Red Media, Blue Media: Evidence of Ideological Selectivity in Media Use". *Journal of Communication* 59 (1): pp. 19–39.

Iyengar, Shanto, Kyu S. Hahn, Jon A. Krosnick, and John Walker (2008). "Selective Exposure to Campaign Communication: The Role of Anticipated Agreement and Issue Public Membership". *The Journal of Politics* 70 (1): pp. 186–200.

Iyengar, Shanto, Yphtach Lelkes, Matthew Levendusky, Neil Malhotra, and Sean J. Westwood (2019). "The Origins and Consequences of Affective Polarization in the United States". *Annual Review of Political Science* 22 (1): pp. 129–146.

Kamenica, Emir and Matthew Gentzkow (2011). "Bayesian Persuasion". *American Economic Review* 101 (6): pp. 2590–2615.

Knobloch-Westerwick, Silvia and Jingbo Meng (2009). "Looking the Other Way: Selective Exposure to Attitude-Consistent and Counterattitudinal Political Information". *Communication Research* 36 (3): pp. 426–448.

Kunda, Ziva (1990). "The case for motivated reasoning". *Psychological Bulletin* 108 (3): pp. 480–498.

Lazarsfeld, Paul F., Bernard Berelson, and Hazel Gaudet (1948). *The People's Choice: How the Voter Makes Up His Mind in a Presidential Campaign, Legacy Edition*. Columbia University Press.

Levendusky, Matthew S. (2013). "Why Do Partisan Media Polarize Viewers?" *American Journal of Political Science* 57 (3): pp. 611–623.

Little, Andrew T. (2019). "The Distortion of Related Beliefs". *American Journal of Political Science* 63 (3): pp. 675–689.

— (2022). "Detecting Motivated Reasoning". *Working Paper*.

Lord, Charles G., Lee Ross, and Mark R. Lepper (1979). "Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence". *Journal of Personality and Social Psychology* 37 (11): pp. 2098–2109.

Martin, Gregory J. and Ali Yurukoglu (2017). "Bias in Cable News: Persuasion and Polarization". *American Economic Review* 107 (9): pp. 2565–2599.

Mummolo, Jonathan (2016). "News from the Other Side: How Topic Relevance Limits the Prevalence of Partisan Selective Exposure". *The Journal of Politics* 78 (3): pp. 763–773.

Nyhan, Brendan and Jason Reifler (2010). "When Corrections Fail: The Persistence of Political Misperceptions". *Political Behavior* 32 (2): pp. 303–330.

Peterson, Erik, Sharad Goel, and Shanto Iyengar (2021). "Partisan selective exposure in online news consumption: evidence from the 2016 presidential campaign". *Political Science Research and Methods* 9 (2): pp. 242–258.

Peterson, Erik and Shanto Iyengar (2021). "Partisan Gaps in Political Information and Information-Seeking Behavior: Motivated Reasoning or Cheerleading?" *American Journal of Political Science* 65 (1): pp. 133–147.

Peterson, Erik and Ali Kagalwala (2021). "When Unfamiliarity Breeds Contempt: How Partisan Selective Exposure Sustains Oppositional Media Hostility". *American Political Science Review* 115 (2): pp. 585–598.

Petty, Richard E. and John T. Cacioppo (1986). “The Elaboration Likelihood Model of Persuasion”. *Advances in Experimental Social Psychology.* Ed. by Leonard Berkowitz. Vol. 19. Academic Press, pp. 123–205.

Prior, Markus (2013). “Media and Political Polarization”. *Annual Review of Political Science* 16 (1): pp. 101–127.

Prior, Markus, Gaurav Sood, and Kabir Khanna (2015). “You Cannot be Serious: The Impact of Accuracy Incentives on Partisan Bias in Reports of Economic Perceptions”. *Quarterly Journal of Political Science* 10 (4): pp. 489–518.

Rabin, Matthew (1998). “Psychology and Economics”. *Journal of Economic Literature* 36 (1): pp. 11–46.

Rabin, Matthew and Joel L. Schrag (1999). “First Impressions Matter: A Model of Confirmatory Bias”. *The Quarterly Journal of Economics* 114 (1): pp. 37–82.

Redlawsk, David P. (2002). “Hot Cognition or Cool Consideration? Testing the Effects of Motivated Reasoning on Political Decision Making”. *Journal of Politics* 64 (4): pp. 1021–1044.

Schmidt, Ana Lucía, Fabiana Zollo, Michela Del Vicario, Alessandro Bessi, Antonio Scala, Guido Caldarelli, H. Eugene Stanley, and Walter Quattrociocchi (2017). “Anatomy of news consumption on Facebook”. *Proceedings of the National Academy of Sciences* 114 (12): pp. 3035–3039.

Stroud, Natalie Jomini (2010). “Polarization and Partisan Selective Exposure”. *Journal of Communication* 60 (3): pp. 556–576.

Taber, Charles S., Damon Cann, and Simona Kucsova (2009). “The Motivated Processing of Political Arguments”. *Political Behavior* 31 (2): pp. 137–155.

Taber, Charles S. and Milton Lodge (2006). “Motivated Skepticism in the Evaluation of Political Beliefs”. *American Journal of Political Science* 50 (3): pp. 755–769.

Wood, Thomas and Ethan Porter (2019). “The Elusive Backfire Effect: Mass Attitudes’ Steadfast Factual Adherence”. *Political Behavior* 41 (1): pp. 135–163.

Yang, Tian, Sílvia Majó-Vázquez, Rasmus K. Nielsen, and Sandra González-Bailón (2020). "Exposure to news grows less fragmented with an increase in mobile access". *Proceedings of the National Academy of Sciences* 117 (46): pp. 28678–28683.

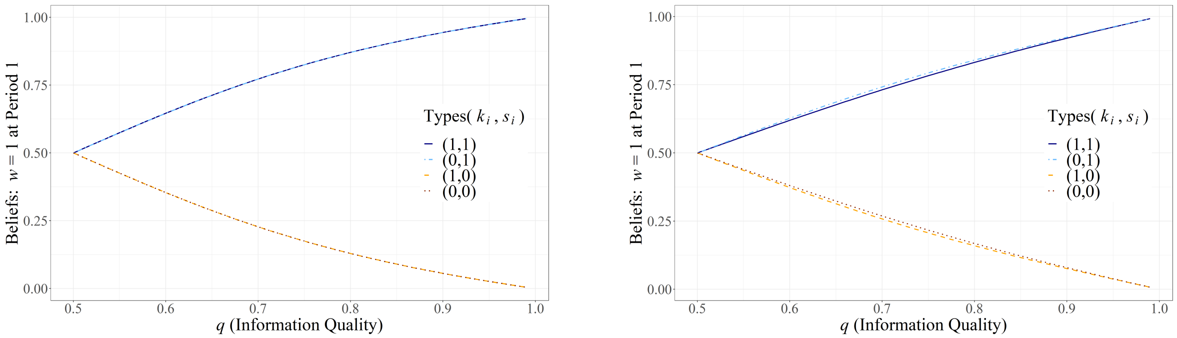
Zaller, John R. (1992). *The Nature and Origins of Mass Opinion.* Cambridge University Press.

Supporting Information for Information and Motivated Reasoning: A Model of Selective Exposure

Table of Contents

A Additional Figures

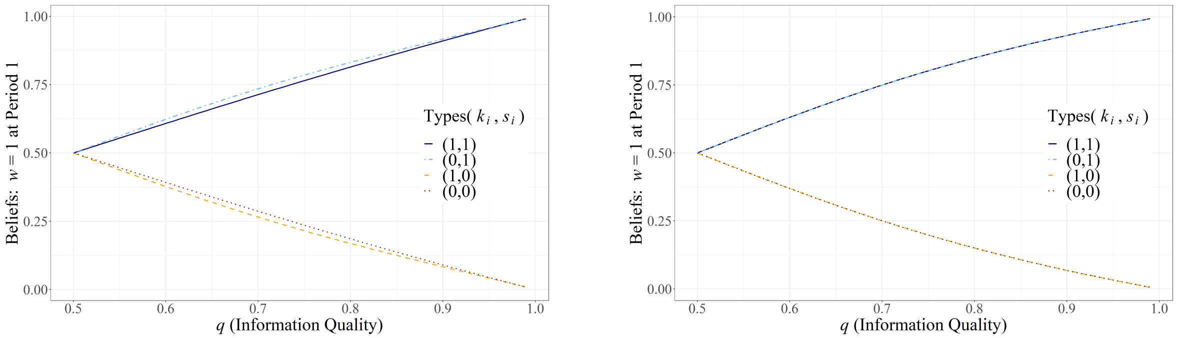
Figure 6: Posterior Beliefs at Period 1 by Types



(a) Accuracy Equilibrium

(b) Separating Equilibrium

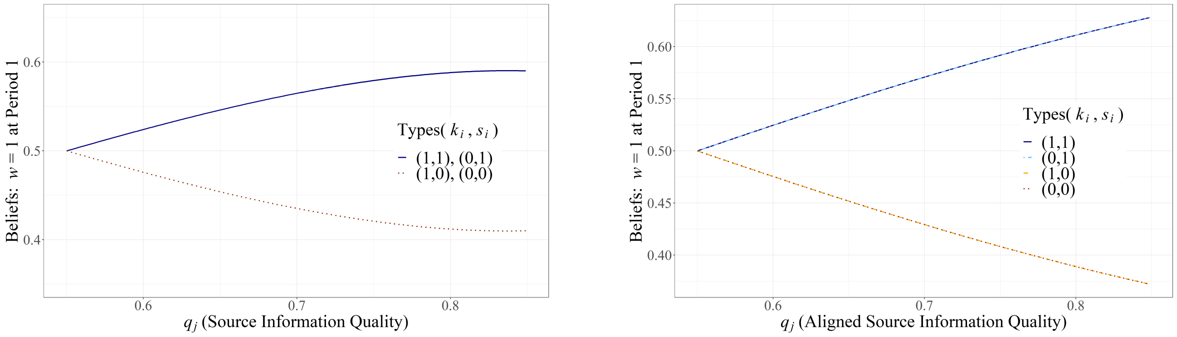
Figure 7: Posterior Beliefs at Period 1 in Separating Equilibrium by Types



(a) High-Types m_H (0.521)

(b) Low-Types m_L (0.558)

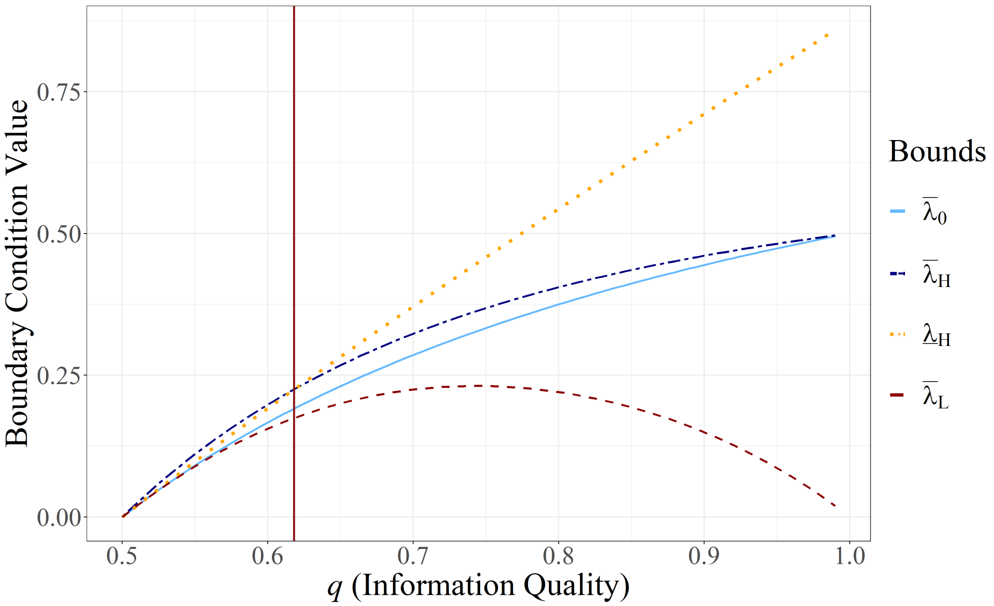
Figure 8: Extension 2: Posterior Beliefs at Period 1 by Varying Degrees of Credibility



(a) q_j Uniformly Greater

(b) Opposite Types Not Credible

Figure 9: Comparative Statics on the Boundary Conditions of Extension 1 Corollary 2



Note: the vertical line marks the point at which the lower bound $\underline{\lambda}_H$ exceeds the upper bound for persuasion $\bar{\lambda}_H$, thereby rendering the persuasion no longer feasible. The given boundary conditions assume $\rho = \frac{1}{2}$.

B Proof and Formal Characterizations of Propositions

Below provides detailed proof of proposition 1 stated in the main text. The proof is followed by formal characterizations of propositions 2 and 3. Detailed proof of propositions 2 and 3 follow identical steps to that of proposition 1, hence omitted.[1]

Proof of Definition 1 (λ threshold)

$$\bar{\lambda}^0 \equiv \frac{2q-1}{2q}$$

Proof

Note that it suffices to consider the payoffs of an individual i who receives a signal s_i that does not align with one's identity k_i. For those who receive the aligned signal (i.e., $s_i = k_i$), setting $x_i^0 \neq s_i$ is strictly dominated. Consider the payoffs for the former case ($s_i \neq k_i$):

$$u_i[x_i^0 = s_i] = -\lambda - (1-\lambda)\Pr(x_i^0 \neq w|s_i) = -\lambda - (1-\lambda)(1-q)$$

$$u_i[x_i^0 \neq s_i] = -(1-\lambda)q$$

Setting the two payoffs equal to each other and solving for λ returns the desired threshold $\bar{\lambda}^0$. □

[1] More detailed proof with derivations of relevant boundary conditions are available upon request.

B.1 Proof of Proposition 1 (Known Information Source Equilibria)

B.1.1 Accuracy Equilibrium ($\lambda_L < \lambda_H < \bar{\lambda}^0$)

1. Receiver equilibrium strategies (a^*_{ij} and b^*_i):

$(k_j, \pi_j) \backslash (k_i, s_i)$	$(1,1)$	$(1,0)$	$(0,1)$	$(0,0)$
$(1,1)$	1	0^*	1	0
$(1,0)$	0	1	0	1^*
$(0,1)$	1^*	0	1	0
$(0,0)$	0	1	0^*	1

   where

   - $1^* = 1$ and $0^* = 0$ if $\lambda_H < \bar{\lambda}^1_H \equiv \frac{q-3q^2+2q^3}{-1+3q-5q^2+3q^3}$;
   - $1^* = 0$ and $0^* = 1$ if $\lambda_L > \underline{\lambda}^1_H \equiv \frac{2q-1}{2q+1}$.[2]

2. Sender equilibrium strategy: $\pi^*_j = s_j$ for all types.

3. Receiver posterior beliefs (based on π_j) on the SOW:

   - $s_i = 1$:
     - $\pi_j = 1$: $\mu^{1*}_{(\pi_j=1,w=0)} = \frac{(1-q)^2}{(1-q)^2+q^2}$
     - $\pi_j = 0$: $\mu^{1*}_{(\pi_j=0,w=1)} = \frac{1}{2}$
   - $s_i = 0$:
     - $\pi_j = 1$: $\mu^{1*}_{(\pi_j=1,w=0)} = \frac{1}{2}$
     - $\pi_j = 0$: $\mu^{1*}_{(\pi_j=0,w=1)} = \frac{(1-q)^2}{(1-q)^2+q^2}$

Proof

Conjecture an equilibrium where all agents, regardless of their types, prefer to set $\pi_j = s_j$ as senders. This requires those with misaligned signals (i.e., $s_i \neq k_i$) to be willing to set their intended actions $x^0_i = s_i \neq k_i$, meaning $\lambda_H < \bar{\lambda}^0$.

Before considering the receiver's expected payoffs, note that considering the equilibrium behavior for one type (e.g., $k_i = 1$) is WLOG given the symmetry in type realization. Then, assuming $k_i = 1$, we characterize the equilibrium by first considering the receiver's equilibrium behavior and then checking the sender's no-deviation conditions. Given the perfect information receivers have about the senders' types, there are four cases to consider for both high and low types, but Lemma 1 below simplifies our analysis by showing the need to consider the case only for the high-types.

[2] The assumption on $q \in (\frac{1}{2}, 1)$ implies $\bar{\lambda}^1_H < \underline{\lambda}^1_H$, meaning both conditions cannot simultaneously be satisfied.

Lemma 1
In the accuracy equilibrium, low-types ($m_i = m_L$) equilibrium behavior as receivers match those of high-types ($m_i = m_H$). For the senders' equilibrium behavior, it suffices to consider the high-types' incentives to deviate.

Proof This is almost immediate from the fact that both types are conjectured to set the initial intended actions x_i^0 and π_i equal to s_i in this equilibrium. This in turn results in an identical set of expected payoffs when facing senders' messages as receivers, just with different λ_i. When considering possible deviations on selecting π_i as senders, note that $\lambda_L < \lambda_H$ necessarily renders the high-types as more likely ones to deviate to setting $\pi_i = k_i \neq s_i$. Therefore, no deviation by the high-types implies the same for the low-types. □

We first begin with the receiver's equilibrium behavior given senders' messages. Denote the receiver as i and the sender j.

1. $(k_i, s_i, m_i) = (1, 1, m_H)$:

   - $(k_j, \pi_j) = (1, 1)$:
     First, denote the receiver's posterior on SOW upon seeing the sender's message as $\mu^1_{\pi_j,w}$ where super script 1 denotes the period. Then, upon observing $(k_j, \pi_j) = (1, 1)$,

     $$\begin{aligned}\mu^1_{(\pi_j=1,w=0)} = \Pr(w=0|\pi_j=1) &= \frac{\Pr(w=0)\Pr(\pi_j=1|w=0)}{\Pr(\pi_j=1)}\\ &= \frac{(1-q)(\frac{1}{2}(1-q)+\frac{1}{2}(1-q))}{(1-q)(\frac{1}{2}(1-q)+\frac{1}{2}(1-q))+q(\frac{1}{2}(q)+\frac{1}{2}(q))} = \frac{(1-q)^2}{(1-q)^2+q^2} < 1-q\end{aligned}$$

     where the last inequality follows from $q > \frac{1}{2}$. Then, her expected payoffs for each interaction decision a_{ij} is as follows:

     $$\begin{aligned}&E[u_i|a_{ij}=1] = -(1-\lambda_H)\mu^1_{(\pi_j=1,w=0)}\\ &E[u_i|a_{ij}=0] = -(1-\lambda_H)\mu^0_{(s_i=1,w=0)} = -(1-\lambda_H)(1-q)\end{aligned}$$

     As shown with the inequality above, the second expression is strictly dominated, hence $a_{ij} = 1$ always.

   - $(k_j, \pi_j) = (1, 0)$:
     Following the same procedure as before,

     $$\begin{aligned}\mu^1_{(\pi_j=0,w=1)} = \Pr(w=1|\pi_j=0) &= \frac{\Pr(w=1)\Pr(\pi_j=0|w=1)}{\Pr(\pi_j=0)}\\ &= \frac{q(\frac{1}{2}(1-q)+\frac{1}{2}(1-q))}{q(\frac{1}{2}(1-q)+\frac{1}{2}(1-q))+(1-q)(\frac{1}{2}(q)+\frac{1}{2}(q))} = \frac{1}{2}\end{aligned}$$

Comparing her expected payoffs,

$$E[u_i|a_{ij}=1,\pi_j=0] = -(1-\lambda_H)\mu^1_{(\pi_j=0,w=1)}$$
$$E[u_i|a_{ij}=0,x_i^0=1] = -(1-\lambda_H)(1-q)$$

Since $\frac{1}{2} > 1-q$, $a_{ij}=0$ always.

- $(k_j,\pi_j)=(0,1)$:
  Now the receiver observes that the sender's type is directly opposite to that of her type. Since she still observes $\pi_j=1$, the posterior remains identical to the case when $(k_j,\pi_j)=(1,1)$. Then, her expected payoffs are:

  $$E[u_i|a_{ij}=1] = -\lambda_H-(1-\lambda_H)\mu^1_{(\pi_j=1,w=0)}$$
  $$E[u_i|a_{ij}=0] = -(1-\lambda_H)\mu^0_{(s_i=1,w=0)} = -(1-\lambda_H)(1-q)$$

  Then, $a_{ij}=1$ iff

  $$-\lambda_H-(1-\lambda_H)\frac{(1-q)^2}{(1-q)^2+q^2} > -(1-\lambda_H)(1-q) \Leftrightarrow \lambda_H < \frac{q-3q^2+2q^3}{-1+3q-5q^2+3q^3} \equiv \bar{\lambda}^1_H$$

- $(k_j,\pi_j)=(0,0)$:
  $a_{ij}=0$ always, which directly follows from the second case $(k_j,\pi_j)=(1,0)$ above; the exposure decision remains strictly dominated.

2. $(k_i,s_i,m_i)=(1,0,m_H)$: Now the receiver has seen a different signal at Period 0. This affects the "prior" on SOW, so her calculation of posteriors differs from above.

   - $(k_j,\pi_j)=(1,1)$:

     $$\mu^1_{(\pi_j=1,w=0)} = \frac{(q)(\frac{1}{2}(1-q)+\frac{1}{2}(1-q))}{(q)(\frac{1}{2}(1-q)+\frac{1}{2}(1-q))+(1-q)(\frac{1}{2}(q)+\frac{1}{2}(q))} = \frac{1}{2}$$

     Then, her expected payoffs for each interaction decision a_{ij} is as follows:

     $$E[u_i|a_{ij}=1] = -(1-\lambda_H)\frac{1}{2}$$
     $$E[u_i|a_{ij}=0] = -\lambda_H-(1-\lambda_H)(1-q)$$

     $a_{ij}=1$ iff $\lambda_H > \frac{2q-1}{2q+1} \equiv \underline{\lambda}^1_H$.

   - $(k_j,\pi_j)=(1,0)$:

     $$\mu^1_{(\pi_j=0,w=1)} = \frac{(1-q)^2}{(1-q)^2+q^2}$$

     $$E[u_i|a_{ij}=1] = -(1-\lambda_H)\mu^1_{(\pi_j=0,w=1)}$$
     $$E[u_i|a_{ij}=0] = -\lambda_H-(1-\lambda_H)(1-q)$$

Since the second expression is strictly dominated, $a_{ij} = 1$ always.

- $(k_j, \pi_j) = (0, 1)$:

  $$E[u_i|a_{ij} = 1] = -\lambda_H - (1 - \lambda_H)\frac{1}{2}$$
  $$E[u_i|a_{ij} = 0] = -\lambda_H - (1 - \lambda_H)(1 - q)$$

  Since $\frac{1}{2} > 1 - q$, $a_{ij} = 0$ always.

- $(k_j, \pi_j) = (0, 0)$:

  $$E[u_i|a_{ij} = 1] = -\lambda_H - (1 - \lambda_H)\mu^1_{(\pi_j=0,w=1)}$$
  $$E[u_i|a_{ij} = 0] = -\lambda_H - (1 - \lambda_H)(1 - q)$$

  No exposure is strictly dominated, so $a_{ij} = 1$ always.

Now we confirm the sender's incentives to deviate. By Lemma 1, it suffices to consider the non-deviating conditions for the "conflicted" high-types (i.e., those who receive an initial signal not aligned with their own types). Also, comparing the two additional boundary conditions on λ_H above, the assumption on $q \in (\frac{1}{2}, 1)$ leads to $\bar{\lambda}^1_H < \underline{\lambda}^1_H$, meaning we have two cases to consider: (1) $\lambda_H < \bar{\lambda}^1_H$ and (2) $\lambda_H > \underline{\lambda}^1_H$. WLOG, suppose $(k_j, s_j) = (1, 0)$, who is conjectured to set $\pi_j = 0$.

1. $\lambda_H < \bar{\lambda}^1_H$:

   - $\pi_j = 1$:

     $$E[u_j|\pi_j = 1] = -\lambda_H \cdot 0 - (1 - \lambda_H)\Pr(w = 0|s_j = 0) - \Pr(a_{ij} = 0|\pi_j = 1)$$

     where the last term represents the potential loss from relational damage (i.e., rejection by the receiver). Then, based on the receiver's equilibrium action,

     $$\begin{aligned}\Pr(a_{ij} = 0|\pi_j = 1) &= q[\frac{1}{2}\{q\} + \frac{1}{2}\{q\}] + (1 - q)[\frac{1}{2}\{1 - q\} + \frac{1}{2}\{1 - q\}] \\ &= q^2 + (1 - q)^2\end{aligned}$$

   - $\pi_j = 0$:

     $$E[u_j|\pi_j = 0] = -\lambda_H \cdot 1 - (1 - \lambda_H)\Pr(w = 1|s_j = 0) - \Pr(a_{ij} = 0|\pi_j = 0)$$

     where

     $$\begin{aligned}\Pr(a_{ij} = 0|\pi_j = 0) &= q[\frac{1}{2}\{1 - q\} + \frac{1}{2}\{1 - q\}] + (1 - q)[\frac{1}{2}\{q\} + \frac{1}{2}\{q\}] \\ &= 2q(1 - q)\end{aligned}$$

   Then, comparing the expected payoffs reveals that $\pi_j = 1$ is strictly dominated, so no individuals would want to deviate under this condition.

2. $\lambda_H > \underline{\lambda}_H^1$:[3]

   - $\pi_j = 1$:

     $$E[u_j|\pi_j = 1] = -\lambda_H \cdot 0 - (1-\lambda_H)\Pr(w = 0|s_j = 0) - \Pr(a_{ij} = 0|\pi_j = 1)$$

     where the latter probability now becomes:

     $$\begin{aligned}\Pr(a_{ij} = 0|\pi_j = 1) &= q[\frac{1}{2}\{0\} + \frac{1}{2}\{q\}] + (1-q)[\frac{1}{2}\{0\} + \frac{1}{2}\{1-q\}] \\ &= \frac{1}{2}(q^2 + (1-q)^2)\end{aligned}$$

   - $\pi_j = 0$:

     $$E[u_j|\pi_j = 0] = -\lambda_H \cdot 1 - (1-\lambda_H)\Pr(w = 1|s_j = 0) - \Pr(a_{ij} = 0|\pi_j = 0)$$

     where

     $$\begin{aligned}\Pr(a_{ij} = 0|\pi_j = 0) &= q[\frac{1}{2}\{1-q\} + \frac{1}{2}\{1\}] + (1-q)[\frac{1}{2}\{q\} + \frac{1}{2}\{1\}] \\ &= q(1-q) + \frac{1}{2}\end{aligned}$$

   Then, comparing the expected payoffs reveals that

   $$\pi_j = 0 \text{ iff } \lambda_H < \frac{2q^2-1}{2q} \equiv \bar{\lambda}_{H,2}^1 \text{ and } q > \hat{q} \approx 0.848$$

Assuming that boundary conditions on λ are met for both types, then, the equilibrium is sustainable as conjectured. □

Corollary 3
In the accuracy equilibrium of the known information source identity setting, exposure ($a_{ij} = 1$) implies persuasion ($b_i = 1$).

Proof This is immediate from the consideration of expected payoffs for the cases of $a_{ij} = 1$. For those choosing to expose themselves, no persuasion (i.e., maintaining their prior beliefs) is strictly dominated. □

B.1.2 Directional Motive Equilibrium ($\bar{\lambda}^0 < \lambda_L < \lambda_H$)

1. Receiver equilibrium strategies (a_{ij}^*, b_i^*):

[3] This is a particularly difficult condition to meet, considering that λ_H is assumed to be particularly high, making it more appealing for the conflicted high-types to set $\pi_j = k_j$. To be more complete, we also need to consider separate cases where the low-types' weights λ_L falls into a different condition, but for the sake of brevity, the derivation below assumes $\lambda_H > \lambda_L > \underline{\lambda}_H^1$.

- Exposure decisions (a^*_{ij}):

$(k_j, \pi_j) \backslash (k_i, s_i)$	$(1,1)$	$(1,0)$	$(0,1)$	$(0,0)$
$(1,1)$	1^*	1^*	0	0
$(1,0)$	1^*	1^*	0	0
$(0,1)$	0	0	1^*	1^*
$(0,0)$	0	0	1^*	1^*

where $1^* \in [0,1]$.

- Belief update decisions: $b^*_i = 0$ for all types.

2. Sender equilibrium strategy: $\pi^*_j = k_j$ for all types.
3. Receiver posterior beliefs (based on π_j) on the SOW: $\mu^{1*}_{\pi_j,w} = \mu^0_w \quad \forall \pi_j$
4. Receiver off-path beliefs (upon observing $\pi_j \neq k_j$): $k_j \neq \pi_j$.

Proof
Now conjecture an equilibrium where all individuals, regardless of their types, select intended actions at period 0 equal to their directional motives and send signals identical to their types as senders (i.e., $x^0_i = k_i$ and $\pi_j = k_j$).

Lemma 2
In the directional motive equilibrium, low-types ($m_i = m_L$) equilibrium behavior as receivers match those of high-types ($m_i = m_H$). For the senders' equilibrium behavior, it suffices to consider the low-types' incentives to deviate.

Proof An analogous line of reasoning from lemma 1 applies for both parts. For the sender's equilibrium behavior, the "conflicted" low-types are the ones more likely to deviate by setting $\pi_j = s_j \neq k_j$. Therefore, no deviation by the low-types implies the same for the high-types. □

Lemma 3
Sender's signal π_j does not relay any additional information about the SOW in the directional motive equilibrium.

Proof This is immediate from the fact that every individual, regardless of their types, is conjectured to set the signal equal to their types. Consider a receiver's posterior on SOW upon observing π_j:

$$\mu^1_{(\pi_j,w)} = \Pr(w|\pi_j) = \frac{(\mu^0)(\frac{1}{2}\cdot 1 + \frac{1}{2}\cdot 0)}{(\mu^0)(\frac{1}{2}\cdot 1 + \frac{1}{2}\cdot 0) + (1-\mu^0)(\frac{1}{2}\cdot 1 + \frac{1}{2}\cdot 0)} = \mu^0$$

□

First, consider the receiver's equilibrium behavior. Based on lemma 2 and 3, it follows that receiver never interacts with those holding opposite directional motives, while indifferent when seeing that the sender is of the same type. More formally,

1. $k_j = k_i$:

   $E[u_i|a_{ij} = 1] = -(1-\lambda)\mu^1_{(\pi_j,w)} = -(1-\lambda)\mu^0$
   $E[u_i|a_{ij} = 0] = -(1-\lambda)\mu^0$

   where μ_0 is the initial posterior belief from period 0 that the action taken in period 0 is wrong (i.e., $x_i^0 \neq w$). Given the equality, the receiver is indifferent about exposure.

2. $k_j \neq k_i$:

   $E[u_i|a_{ij} = 1] = -\lambda - (1-\lambda)\mu^1_{(\pi_j,w)} = -\lambda - (1-\lambda)\mu^0$
   $E[u_i|a_{ij} = 0] = -(1-\lambda)\mu^0$

   As setting $a_{ij} = 1$ is strictly dominated, the receiver chooses not to expose herself when the sender's directional motive is not aligned.

Now we confirm the sender's incentives to deviate. Based on Lemma 2, it suffices to consider the case for the conflicted low-types. Unlike the accuracy equilibrium, it is important to note that the receiver's rejection probability $\Pr[a_{ij} = 0]$ is independent of the sender's message π_j but depends solely on k_j. Then, the sender's expected payoffs are as follows:

1. $(k_j, s_j) = (1, 0)$: expected to set $\pi_j = 1$.

   - $\pi_j = 1$:

     $E[u_j|\pi_j = 1] = -\lambda_L \cdot 0 - (1-\lambda_L)\Pr(w = 0|s_j = 0) - \Pr(a_{ij} = 0|k_j)$

   - $\pi_j = 0$:

     $E[u_j|\pi_j = 0] = -\lambda_L - (1-\lambda_L)(1-q) - \Pr(a_{ij} = 0|k_j)$

   By the initial assumption on $\lambda_L > \bar{\lambda}^0$, deviation to setting $\pi_j = s_j \neq k_j$ is strictly dominated. An analogous line of reasoning applies for types $k_j = 0$.

Note that this means the receivers can adopt any exposure decision $a_{ij} \in [0, 1]$. As we have confirmed that senders would not deviate, the equilibrium is sustainable as characterized. □

Corollary 4
Receivers' posterior beliefs on SOW do not change following the exposure; $b_i^* = 0$, *which implies no further divergence in beliefs at period 1.*

Proof
This is immediate from Lemma 3. With the sender's signal not relaying any information, there is no persuasion to be done for receivers. Accordingly, receivers' beliefs on SOW remain identical to those from period 0, hence no further divergence in beliefs. □

B.1.3 Separating Equilibrium ($\lambda_L < \bar{\lambda}^0 < \lambda_H$)

1. Receiver equilibrium strategies:

   - Exposure decision (a^*_{ij}):

m_i	m_H				m_L			
$(k_j, \pi_j)\ \backslash(k_i, s_i)$	$(1,1)$	$(1,0)$	$(0,1)$	$(0,0)$	$(1,1)$	$(1,0)$	$(0,1)$	$(0,0)$
$(1,1)$	1	1	1^*	0	1	0^*	1	0
$(1,0)$	0	1	0	0	0	1	0	1^+
$(0,1)$	0	0	1	0	1^+	0	1	0
$(0,0)$	0	1^*	1	1	0	1	0^*	1

where

- $1^* = 1$ if $\lambda_H < \bar{\lambda}^1_{H,1} \equiv \frac{1-3q+3q^2-2q^3+q\rho-3q^2\rho+2q^3\rho}{-q+q^2-2q^3-q\rho-q^2\rho+2q^3\rho}$;
- $1^+ = 1$ if $\lambda_L < \bar{\lambda}^1_L \equiv \frac{q-3q^2+2q^3}{-1+3q-5q^2+3q^3}$;
- $0^* = 0$ if $\lambda_L < \underline{\lambda}^1_L \equiv \frac{-q+3q^2-2q^3-\rho+3q\rho-3q^2\rho+2q^3\rho}{q+q^2-2q^3+q\rho-q^2\rho+2q^3\rho}$.[4]

   - Belief update decision (b^*_i):

m_i	m_H				m_L			
$(k_j, \pi_j)\ \backslash(k_i, s_i)$	$(1,1)$	$(1,0)$	$(0,1)$	$(0,0)$	$(1,1)$	$(1,0)$	$(0,1)$	$(0,0)$
$(1,1)$	1	1	1	0	1	0	1	0
$(1,0)$	0	1^{**}	0	0	0	1	0	1
$(0,1)$	0	0	1^{**}	0	1	0	1	0
$(0,0)$	0	1	1	1	0	1	0	1

where $1^{**} = 1$ if $\lambda_H < \bar{\lambda}^1_{H,2} \equiv \frac{-1+3q-3q^2+2q^3}{q-q^2+2q^3}$.

2. Sender equilibrium strategies:

$$\pi^*_j = \begin{cases} k_j & \text{if } m_i = m_H \\ s_j & \text{if } m_i = m_L \end{cases}$$

which requires $\rho > \underline{\rho} \equiv \frac{1-4q+4q^2}{1-2q+2q^2}$, and $q > \underline{q} \equiv \frac{1}{\sqrt{2}}$

3. Receiver posterior beliefs (based on π_j) on the SOW:

   - $s_i = 1$:
     - $(k_j, \pi_j) = (1,1)$: $\mu^{1*}_{(\pi_j=1, w=0)} = \frac{(1-q)(\rho+(1-\rho)(1-q))}{(1-q)(\rho+(1-\rho)(1-q))+q(\rho+(1-\rho)q)}$
     - $(k_j, \pi_j) = (1,0)$: $\mu^{1*}_{(\pi_j=0, w=1)} = \frac{1}{2}$

[4] By the assumptions on q and ρ, $\underline{\lambda}^1_L > \bar{\lambda}^1_L$, which necessarily means that both figures cannot simultaneously be equal to 1.

  - $(k_j, \pi_j) = (0, 1)$: $\mu^{1*}_{(\pi_j=1,w=0)} = \frac{(1-q)^2}{(1-q)^2+q^2)}$
  - $(k_j, \pi_j) = (0, 0)$: $\mu^{1*}_{(\pi_j=0,w=1)} = \frac{q(\rho+(1-\rho)(1-q))}{q(\rho+(1-\rho)(1-q))+(1-q)(\rho+(1-\rho)q)}$
- $s_i = 0$:
  - $(k_j, \pi_j) = (1, 1)$: $\mu^{1*}_{(\pi_j=1,w=0)} = \frac{(q)(\rho+(1-\rho)(1-q))}{(q)(\rho+(1-\rho)(1-q))+(1-q)(\rho+(1-\rho)q)}$
  - $(k_j, \pi_j) = (1, 0)$: $\mu^{1*}_{(\pi_j=0,w=1)} = \frac{(1-q)^2}{(1-q)^2+q^2)}$
  - $(k_j, \pi_j) = (0, 1)$: $\mu^{1*}_{(\pi_j=1,w=0)} = \frac{1}{2}$
  - $(k_j, \pi_j) = (0, 0)$: $\mu^{1*}_{(\pi_j=0,w=1)} = \frac{(1-q)(\rho+(1-\rho)(1-q))}{(1-q)(\rho+(1-\rho)(1-q))+(q)(\rho+(1-\rho)q)}$

Proof
Conjecture an equilibrium where high-types adopt the pure strategy of setting $\pi_j = k_j$ regardless of their initial signals s_j, while low-types set $\pi_j = s_j$ regardless of their types k_j. The assumption on λ can induce such a divergence in behaviors. Now that we have another dimension in identity, we have four different cases to consider.

Before the derivation, note that observing $k_j \neq \pi_j$ perfectly reveals that the sender is a low-type individual, which simplifies the derivation of expected payoffs. Below specifically derives b_i only when $a_{ij} = 1$ does not necessarily imply updating of beliefs (i.e., when omitted, $a_{ij} = 1$ implies $b_i = 1$).

1. $(k_i, s_i, m_i) = (1, 1, m_H)$

   - $(k_j, \pi_j) = (1, 1)$:

     $$\mu^1_{(\pi_j=1,w=0)} = \frac{(1-q)(\rho+(1-\rho)(1-q))}{(1-q)(\rho+(1-\rho)(1-q))+q(\rho+(1-\rho)q)}$$

     $E[u_i|a_{ij} = 1] = -(1-\lambda_H)\mu^1_{(\pi_j=1,w=0)}$
     $E[u_i|a_{ij} = 0] = -(1-\lambda_H)(1-q)$

     Comparing the payoffs shows that $a_{ij} = 0$ is strictly dominated, hence $a_{ij} = 1$.
   - $(k_j, \pi_j) = (1, 0)$:

     $$\mu^1_{(\pi_j=0,w=1)} = \frac{(1-q)(q)}{(1-q)(q)+q(1-q)} = \frac{1}{2}$$

     $E[u_i|a_{ij} = 1] = -(1-\lambda_H)\mu^1_{(\pi_j=0,w=1)}$
     $E[u_i|a_{ij} = 0] = -(1-\lambda_H)(1-q)$

     Since $(1-q) < \frac{1}{2}$, $a_{ij} = 0$.
   - $(k_j, \pi_j) = (0, 1)$:

     $$\mu^1_{(\pi_j=1,w=0)} = \frac{(1-q)(1-q)}{(1-q)(1-q)+q^2)}$$

$$E[u_i|a_{ij}=1] = -\lambda_H - (1-\lambda_H)\mu^1_{(\pi_j=1,w=0)}$$
$$E[u_i|a_{ij}=0] = -(1-\lambda_H)(1-q)$$

$a_{ij}=0$, as the former payoff is strictly dominated.

- $(k_j, \pi_j) = (0,0)$:

$$\mu^1_{(\pi_j=0,w=1)} = \frac{q(\rho+(1-\rho)(1-q))}{q(\rho+(1-\rho)(1-q))+(1-q)(\rho+(1-\rho)q)}$$

$$E[u_i|a_{ij}=1] = -\lambda_H - (1-\lambda_H)\mu^1_{(\pi_j=0,w=1)}$$
$$E[u_i|a_{ij}=0] = -(1-\lambda_H)(1-q)$$

$a_{ij}=0$, as the former payoff is strictly dominated.

2. $(k_i, s_i, m_i) = (1, 0, m_H)$

- $(k_j, \pi_j) = (1,1)$:

$$\mu^1_{(\pi_j=1,w=0)} = \frac{(q)(\rho+(1-\rho)(1-q))}{(q)(\rho+(1-\rho)(1-q))+(1-q)(\rho+(1-\rho)q)}$$

$$E[u_i|a_{ij}=1] = -(1-\lambda_H)\mu^1_{(\pi_j=1,w=0)}$$
$$E[u_i|a_{ij}=0] = -(1-\lambda_H)(q)$$

Comparing the payoffs shows that $a_{ij}=0$ is strictly dominated, hence $a_{ij}=1$.

- $(k_j, \pi_j) = (1,0)$:

$$\mu^1_{(\pi_j=0,w=1)} = \frac{(1-q)(1-q)}{(1-q)(1-q)+q(q)}$$

$$E[u_i|a_{ij}=1] = -(1-\lambda_H)\mu^1_{(\pi_j=0,w=1)}$$
$$E[u_i|a_{ij}=0] = -(1-\lambda_H)(q)$$

As the second payoff is strictly dominated, $a_{ij}=1$. On the actual updating of belief, the first payoff term incurs an additional cost of $-\lambda_H$ since $k_i = 1$. Then,

$$b_i = 1 \text{ iff } \lambda_H < \frac{-1+3q-3q^2+2q^3}{q-q^2+2q^3} \equiv \bar{\lambda}^1_{H,2}$$

.

- $(k_j, \pi_j) = (0, 1)$:

  $$\mu^1_{(\pi_j=1,w=0)} = \frac{(q)(1-q)}{(q)(1-q)+(1-q)q} = \frac{1}{2}$$

  $$E[u_i|a_{ij}=1] = -\lambda_H - (1-\lambda_H)\mu^1_{(\pi_j=1,w=0)}$$
  $$E[u_i|a_{ij}=0] = -(1-\lambda_H)(q)$$

  $a_{ij} = 0$, as the former payoff is strictly dominated.

- $(k_j, \pi_j) = (0, 0)$:

  $$\mu^1_{(\pi_j=0,w=1)} = \frac{(1-q)(\rho+(1-\rho)(1-q))}{(1-q)(\rho+(1-\rho)(1-q))+(q)(\rho+(1-\rho)q)}$$

  $$E[u_i|a_{ij}=1] = -\lambda_H - (1-\lambda_H)\mu^1_{(\pi_j=0,w=1)}$$
  $$E[u_i|a_{ij}=0] = -(1-\lambda_H)(q)$$

  Comparison shows that

  $$a_{ij} = 1 \text{ iff } \lambda_H < \frac{1-3q+3q^2-2q^3+q\rho-3q^2\rho+2q^3\rho}{-q+q^2-2q^3-q\rho-q^2\rho+2q^3\rho} \equiv \bar{\lambda}^1_{H,1}$$

3. $(k_i, s_i, m_i) = (1, 1, m_L)$
   For these individuals, the posteriors are identical to those of $(k_i, s_i, m_i) = (1, 1, m_H)$.

   - $(k_j, \pi_j) = (1, 1)$:

     $$E[u_i|a_{ij}=1] = -(1-\lambda_L)\frac{(1-q)(\rho+(1-\rho)(1-q))}{(1-q)(\rho+(1-\rho)(1-q))+q(\rho+(1-\rho)q)}$$
     $$E[u_i|a_{ij}=0] = -(1-\lambda_L)(1-q)$$

     Comparing the payoffs shows that $a_{ij} = 0$ is strictly dominated, hence $a_{ij} = 1$.

   - $(k_j, \pi_j) = (1, 0)$:

     $$E[u_i|a_{ij}=1] = -(1-\lambda_L)\frac{1}{2}$$
     $$E[u_i|a_{ij}=0] = -(1-\lambda_L)(1-q)$$

     Since $(1-q) < \frac{1}{2}$, $a_{ij} = 0$.

- $(k_j, \pi_j) = (0,1)$:

  $$E[u_i|a_{ij} = 1] = -\lambda_L - (1-\lambda_L)\frac{(1-q)^2}{(1-q)^2 + q^2)}$$
  $$E[u_i|a_{ij} = 0] = -(1-\lambda_L)(1-q)$$

  $$a_{ij} = 1 \text{ iff } \lambda_L < \frac{q - 3q^2 + 2q^3}{-1 + 3q - 5q^2 + 2q^3} \equiv \bar{\lambda}_L^1$$

- $(k_j, \pi_j) = (0,0)$:

  $$E[u_i|a_{ij} = 1] = -\lambda_L - (1-\lambda_L)\frac{q(\rho + (1-\rho)(1-q))}{q(\rho + (1-\rho)(1-q)) + (1-q)(\rho + (1-\rho)q)}$$
  $$E[u_i|a_{ij} = 0] = -(1-\lambda_L)(1-q)$$

  $a_{ij} = 0$, as the former payoff is strictly dominated.

4. $(k_i, s_i, m_i) = (1, 0, m_L)$
   For these individuals, the posteriors are identical to those of $(k_i, s_i, m_i) = (1, 0, m_H)$.

   - $(k_j, \pi_j) = (1,1)$:

     $$E[u_i|a_{ij} = 1] = -(1-\lambda_L)\frac{(q)(\rho + (1-\rho)(1-q))}{(q)(\rho + (1-\rho)(1-q)) + (1-q)(\rho + (1-\rho)q)}$$
     $$E[u_i|a_{ij} = 0] = -\lambda_L - (1-\lambda_H)(1-q)$$

     Comparing the payoffs shows that

     $$a_{ij} = 1 \text{ iff } \lambda_L > \frac{-q + 3q^2 - 2q^3 - \rho + 3q\rho - 3q^2\rho + 2q^3\rho}{q + q^2 - 2q^3 + q\rho - q^2\rho + 2q^3\rho} \equiv \underline{\lambda}_L^1$$

   - $(k_j, \pi_j) = (1,0)$:

     $$E[u_i|a_{ij} = 1] = -(1-\lambda_L)\frac{(1-q)^2}{(1-q)^2 + q^2}$$
     $$E[u_i|a_{ij} = 0] = -\lambda_L - (1-\lambda_L)(1-q)$$

     As the second payoff is strictly dominated, $a_{ij} = 1$.

   - $(k_j, \pi_j) = (0,1)$:

     $$E[u_i|a_{ij} = 1] = -\lambda_L - (1-\lambda_L)\frac{1}{2}$$
     $$E[u_i|a_{ij} = 0] = -\lambda_L - (1-\lambda_L)(1-q)$$

     $a_{ij} = 0$ as the first expected payoff is strictly dominated.

- $(k_j, \pi_j) = (0, 0)$:

  $$E[u_i|a_{ij} = 1] = -\lambda_L - (1 - \lambda_L)\frac{(1 - q)(\rho + (1 - \rho)(1 - q))}{(1 - q)(\rho + (1 - \rho)(1 - q)) + (q)(\rho + (1 - \rho)q)}$$
  $$E[u_i|a_{ij} = 0] = -\lambda_L - (1 - \lambda_L)(1 - q)$$

  $a_{ij} = 1$ as the first expected payoff is strictly dominated.

Based on the receiver's equilibrium behavior, we now confirm the sender's incentives to deviate. Given the divergence in behavior by the degree of conviction, there are two types of "conflicted" types who are more likely to deviate: $(k_j, s_j, m_j) = (1, 0, m_H)$ and $(1, 0, m_L)$. In addition, a number of boundary conditions required for exposure decisions necessarily means we have multiple cases to consider. The derivation below considers (1) the case outlined in the characterization and (2) the "toughest" cases for each type to sustain its conjectured equilibrium behavior.

1. Conjectured case for both types: $\lambda_H < \bar{\lambda}^1_{H,1}, \lambda_L < \bar{\lambda}^1_L$:

   - Type $(k_j, s_j, m_j) = (1, 0, m_H)$:

     $$E[u_j|\pi_j = 1] = -\lambda_H \cdot 0 - (1 - \lambda_H)\Pr(w = 1|s_j = 0) - \Pr(a_{ij} = 0|\pi_j = 1)$$

     where

     $$\Pr(a_{ij} = 0|\pi_j = 1) = \frac{1}{2}(q^2 + (1 - q)^2)(1 + (1 - \rho))$$

     which yields,

     $$u_j[\pi_j = 1] = -(1 - \lambda_H)q - (\frac{1}{2}(q^2 + (1 - q)^2)(1 + (1 - \rho)))$$

     Now consider his expected payoff from deviating to $\pi_j = 0$:

     $$E[u_j|\pi_j = 1] = -\lambda_H \cdot 1 - (1 - \lambda_H)\Pr(w = 0|s_j = 0) - \Pr(a_{ij} = 0|\pi_j = 0)$$

     where

     $$\Pr(a_{ij} = 0|\pi_j = 1) = \frac{\rho}{2} + q(1 - q)(1 + (1 - \rho))$$

     which yields,

     $$u_j[\pi_j = 0] = -\lambda_H - (1 - \lambda_H)(1 - q) - (\frac{\rho}{2} + q(1 - q)(1 + (1 - \rho)))$$

     Solving for $E[u_j[\pi_j = 1]] > E[u_j[\pi_j = 0]]$ along with the boundary condition on λ_H returns a lower bound on ρ as a sufficient condition for non-deviation: $\rho > \frac{1-4q+4q^2}{1-2q+2q^2} \equiv \underline{\rho}_1$.

   - Type $(k_j, s_j, m_j) = (1, 0, m_L)$:

Note that the expected probability of rejection for either type of message is identical for low-types. Carrying out the same analysis for this type returns a lower-bound on q, along with the same condition on ρ from above, as a condition for non-deviation: $q > \frac{1}{\sqrt{2}}$.

2. Toughest case for $(k_j, s_j, m_j) = (1, 0, m_H)$: $\lambda_H > \bar{\lambda}^1_{H,1}, \lambda_L < \underline{\lambda}_{L,1}$:

- Type $(k_j, s_j, m_j) = (1, 0, m_H)$:

  $$E[u_j|\pi_j = 1] = -\lambda_H \cdot 0 - (1 - \lambda_H)\Pr(w = 1|s_j = 0) - \Pr(a_{ij} = 0|\pi_j = 1)$$

  where

  $$\begin{aligned}\Pr(a_{ij} = 0|\pi_j = 1) = (1-q)&\left[\frac{1}{2}(1-\rho)(1-q) + \frac{1}{2}(\rho + (1-\rho)(1-q))\right] \\ &+ q\left[\frac{1}{2}(1-\rho)(q) + \frac{1}{2}(\rho + (1-\rho)(q))\right] = \frac{\rho}{2} + (1-\rho)(q^2 + (1-q)^2)\end{aligned}$$

  which yields,

  $$u_j[\pi_j = 1] = -(1 - \lambda_H)q - (\frac{\rho}{2} + (1-\rho)(q^2 + (1-q)^2))$$

  Now consider his expected payoff from deviating to $\pi_j = 0$:

  $$E[u_j|\pi_j = 1] = -\lambda_H \cdot 1 - (1 - \lambda_H)\Pr(w = 0|s_j = 0) - \Pr(a_{ij} = 0|\pi_j = 0)$$

  where

  $$\begin{aligned}\Pr(a_{ij} = 0|\pi_j = 1) = (1-q)&\left[\frac{1}{2}(\rho q + (1-\rho)(q)) + \frac{1}{2}(\rho + (1-\rho)(q))\right] \\ &+ q\left[\frac{1}{2}(\rho(1-q) + (1-\rho)(1-q)) + \frac{1}{2}(\rho + (1-\rho)(1-q))\right] \\ &= \frac{\rho}{2} + q(1-q) + (1-\rho)q(1-q)\end{aligned}$$

  which yields,

  $$u_j[\pi_j = 0] = -\lambda_H - (1 - \lambda_H)(1-q) - (\frac{\rho}{2} + q(1-q) + (1-\rho)q(1-q))$$

  Solving for $E[u_j[\pi_j = 1]] > E[u_j[\pi_j = 0]]$ along with the boundary condition returns an upper bound on q as a sufficient condition.

- Type $(k_j, s_j, m_j) = (1, 0, m_L)$: Carrying out the same computation as above,

  $$\begin{aligned}E[u_j|\pi_j = 1] &= -\lambda_L \cdot 0 - (1 - \lambda_L)\Pr(w = 1|s_j = 0) - \Pr(a_{ij} = 0|\pi_j = 1) \\ &= -(1 - \lambda_L)q - (\frac{\rho}{2} + (1-\rho)(q^2 + (1-q)^2))\end{aligned}$$

$$E[u_j|\pi_j = 0] = -\lambda_L - (1-\lambda_L)(1-q) - (\frac{\rho}{2} + q(1-q) + (1-\rho)q(1-q))$$

Solving the analogous set of inequality returns (1) upper bound on $q < \frac{-1+\sqrt{5}}{2}$ and (2) lower bound on $\rho > \frac{-2q+4q^3}{1-3q+3q^2}$ as additional conditions.[5]

3. Toughest case for $(k_j, s_j, m_j) = (1, 0, m_L)$: $\lambda_H < \bar{\lambda}^1_{H,1}, \lambda_L > \underline{\lambda}^1_{L,1}$:

- Type $(k_j, s_j, m_j) = (1, 0, m_H)$:

$$E[u_j|\pi_j = 1] = -(1-\lambda_H)q - (\frac{1}{2}(q^2 + (1-q)^2))$$
$$E[u_j|\pi_j = 0] = -\lambda_H - (1-\lambda_H)(1-q) - (\frac{1}{2} + q(1-q))$$

Solving the inequality shows that the only additional condition for $E[u_j|\pi_j = 1] > E[u_j|\pi_j = 0]$ is the initial boundary condition that $\lambda_H < \bar{\lambda}_{H,1}$.

- Type $(k_j, s_j, m_j) = (1, 0, m_L)$:

$$E[u_j|\pi_j = 1] = -(1-\lambda_L)q - (\frac{1}{2}(q^2 + (1-q)^2))$$
$$E[u_j|\pi_j = 0] = -\lambda_L - (1-\lambda_L)(1-q) - (\frac{1}{2} + q(1-q))$$

Solving the inequality shows that additional conditions for $E[u_j|\pi_j = 1] < E[u_j|\pi_j = 0]$ are (1) a lower bound on q, (2) upper bound on $\rho < \frac{-1-2q^2+4q^3}{-1+4q-2q^2+4q^3}$, and (3) upper bound on $\lambda_L < \frac{-1+2q^2}{2q}$.

Then, assuming these boundary conditions hold for each case, the equilibrium is sustainable as characterized. □

B.2 Formal Characterization of Proposition 2 (Unknown Info Source)

B.2.1 Accuracy Equilibrium ($\lambda_L < \lambda_H < \bar{\lambda}^0$)

1. Receiver equilibrium strategies (a^*_{ij} and b^*_i):

$\pi_j \backslash (k_i, s_i)$	$(1,1)$	$(1,0)$	$(0,0)$	$(0,1)$
1	1^*	0	1	0
0	0	1	0	1^*

where $1^* = 1$ if $\lambda_H < \bar{\lambda}^1_H \equiv \frac{2q-6q^2+4q^3}{-1+4q-8q^2+4q^3}$.

2. Sender equilibrium strategy: $\pi^*_j = s_j$ for all types.

[5] Other conditions, including those for the case below, are omitted for brevity, which are available upon request.

3. Receiver posterior beliefs (based on π_j):
   - Posterior belief on sender type: $\tilde{k}_j^*(\pi_j) = \frac{1}{2}$ for all π_j.
   - Posterior belief on SOW ($\mu^1_{\pi_j,w}$):
     - $s_i = 1$: $\mu^{1*}_{\pi_j=1,w=0} = \frac{(1-q)^2}{(1-q)^2+q^2}$ and $\mu^{1*}_{\pi_j=0,w=1} = \frac{1}{2}$.
     - $s_i = 0$: $\mu^{1*}_{\pi_j=1,w=0} = \frac{1}{2}$ and $\mu^{1*}_{\pi_j=0,w=1} = \frac{(1-q)^2}{(1-q)^2+q^2}$.

B.2.2 Directional Motive Equilibrium ($\bar{\lambda}^0 < \lambda_L < \lambda_H$)

1. Receiver equilibrium strategies (a^*_{ij} and b^*_i):

$\pi_j \backslash(k_i, s_i)$	$(1,1)$	$(1,0)$	$(0,0)$	$(0,1)$
1	1^*	1^*	0	0
0	0	0	1^*	1^*

   where $1^* \in [0,1]$.

2. Sender equilibrium strategy: $\pi^*_j = k_j$ for all types.
3. Receiver posterior beliefs (based on π_j):
   - Posterior belief on sender type: $\tilde{k}_j^*(\pi_j) = \pi_j$ for all π_j.
   - Posterior belief on SOW: $\mu^{1*}_{\pi_j,w=1} = \frac{1}{2}$ for all π_j.

B.2.3 Separating Equilibrium ($\lambda_L < \bar{\lambda}^0 < \lambda_H$)

1. Receiver equilibrium strategies:
   - Exposure decision (a^*_{ij}):

m_i	m_H				m_L			
$\pi_j \backslash(k_i, s_i)$	$(1,1)$	$(1,0)$	$(0,1)$	$(0,0)$	$(1,1)$	$(1,0)$	$(0,1)$	$(0,0)$
1	0	0	1^*	0	1^+	0	1	0
0	0	1^*	0	0	0	1	0	1^+

   where
     - $1^* = 1$ if $\lambda_H < \bar{\lambda}^1_{H,1} \equiv \frac{2-6q+6q^2-4q^3-\rho+4q\rho-6q^2\rho+4q^3\rho}{1-4q+4q^2-4q^3-\rho+2q\rho-4q^2\rho+4q^3\rho}$
     - $1^+ = 1$ if $\lambda_L < \bar{\lambda}^1_L \equiv \frac{2q-6q^2+4q^3}{-1+4q-8q^2+4q^3}$
   - Belief update decision (b^*_i):

m_i	m_H				m_L			
$\pi_j \setminus (k_i, s_i)$	$(1,1)$	$(1,0)$	$(0,1)$	$(0,0)$	$(1,1)$	$(1,0)$	$(0,1)$	$(0,0)$
1	0	0	1^{**}	0	1	0	1	0
0	0	1^{**}	0	0	0	1	0	1

where $1^{**} = 1$ if $\lambda_H < \bar{\lambda}^1_{H,2} \equiv \frac{2-6q+6q^2-4q^3-\rho+4q\rho-6q^2\rho+4q^3\rho}{4\rho q^3-4q^3-2\rho q^2+2q^2-2q}$.

2. Sender equilibrium strategy:

$$\pi_j^* = \begin{cases} k_j & \text{if } m_i = m_H \\ s_j & \text{if } m_i = m_L \end{cases}$$

   Non-deviating condition for the "conflicted" high-type individuals ($k_i \neq s_i \wedge m_i = m_H$):[6]

$$\lambda_H > \frac{-4q + 8q^2 - \rho + 4q\rho - 4q^2\rho}{4q} \equiv \underline{\lambda}^1_H$$

3. Receiver posterior beliefs (based on π_j):

   - Posterior belief on sender type: $\tilde{k}_j(\pi_j)$
     - $(s_i, \pi_j) = (1,1)$:
       $\tilde{k}^*_j(1) = \Pr(k_j = 0 | \pi_j = 1) = \frac{\frac{1}{2}[q(1-\rho)q+(1-q)((1-\rho)(1-q))]}{\frac{1}{2}[q(1-\rho)q+(1-q)((1-\rho)(1-q))]+\frac{1}{2}[q(\rho+(1-\rho)(1-q))+(1-q)(\rho+(1-\rho)(1-q))]}$
     - $(s_i, \pi_j) = (1,0)$:
       $\tilde{k}^*_j(0) = \Pr(k_j = 0 | \pi_j = 0) = \frac{\frac{1}{2}[q\{\rho+(1-\rho)(1-q)\}+(1-q)\{\rho+(1-\rho)(q)\}]}{\frac{1}{2}[q\{\rho+(1-\rho)(1-q)\}+(1-q)\{\rho+(1-\rho)(q)\}]+\frac{1}{2}[q\{(1-\rho)(1-q)\}+(1-q)\{(1-\rho)q\}]}$
     - $(s_i, \pi_j) = (0,1)$ and $(s_i, \pi_j) = (0,0)$: replace the initial posterior (i.e., $\mu^0_{s_i,w=1} = 1 - q$) in the corresponding expressions above.
   - Posterior belief on SOW: $\mu^1_{\pi_j,w}$
     - $(s_i, \pi_j) = (1,1)$:
       $\mu^{1*}_{1,w=0} = \frac{(1-q)[\frac{1}{2}\{\rho+(1-\rho)(1-q)\}\frac{1}{2}\{(1-\rho)(1-q)\}]}{(1-q)[\frac{1}{2}\{\rho+(1-\rho)(1-q)\}\frac{1}{2}\{(1-\rho)(1-q)\}]+(q)[\frac{1}{2}\{\rho+(1-\rho)(q)\}\frac{1}{2}\{(1-\rho)(q)\}]}$
     - $(s_i, \pi_j) = (1,0)$:
       $\mu^{1*}_{0,w=1} = \frac{(q)[\frac{1}{2}\{(1-\rho)(1-q)\}+\frac{1}{2}\{(\rho)+(1-\rho)(1-q)\}]}{(q)[\frac{1}{2}\{(1-\rho)(1-q)\}+\frac{1}{2}\{(\rho)+(1-\rho)(1-q)\}]+(1-q)[\frac{1}{2}\{(1-\rho)q\}+\frac{1}{2}\{\rho+(1-\rho)(q)\}]}$
     - $(s_i, \pi_j) = (0,1)$ and $(s_i, \pi_j) = (0,0)$: replace the initial posterior in the corresponding expressions above.

B.3 Formal Characterization of Proposition 3 (Heterogeneous q)

1. Receiver equilibrium strategies (a^*_{ij}):

   - Exposure decision (a^*_{ij}):

[6] Additionally, a non-deviation requires that $\underline{\lambda}^1_H < \bar{\lambda}^1_{H,1}$, which returns an upper bound on q.

Perception of q_j	Uniformly Credible				Opposite Types Not Credible			
$(k_j, \pi_j) \backslash (k_i, s_i)$	$(1,1)$	$(1,0)$	$(0,1)$	$(0,0)$	$(1,1)$	$(1,0)$	$(0,1)$	$(0,0)$
$(1,1)$	1	$1'$	1	1^{***}	1	$1^{'}$	0	0
$(1,0)$	1^*	1	1^*	1^{**}	1^*	1	0	0
$(0,1)$	1^{**}	1^*	1	1^*	0	0	1	1^*
$(0,0)$	1^{***}	1	$1^{'}$	1	0	0	$1^{'}$	1

Characterization of conditions:

- $1^* = 1$ if $q_j > \frac{q_i^2}{1-2q_i+2q_i^2}$
- $1^{**} = 1$ if (1) $q_i < \frac{2}{3}$ and $q_j > \frac{1-4q_i+3q_i^2}{1-6q_i+6q_i^2}$ or (2) $q_i \geq \frac{2}{3}$ and $\lambda_H < \frac{q_i-q_i^2-2q_iq_j+2q_i^2q_j}{-1+2q_i-q_i^2+q_j-4q_iq_j+2q_i^2q_j}$
- $1^{***} = 1$ if (1) $q_i < \frac{2}{3}$ and $q_j > \frac{-q_i+3q_i^2}{2-6q_i+6q_i^2}$ or (2) $q_i \geq \frac{2}{3}$, $q_j > \frac{q_i^2}{1-2q_i+2q_i^2}$, and $\lambda_H < \frac{-q_i^2+q_j-2q_iq_j+2q_i^2q_j}{q_i-q_i^2+2q_j-4q_iq_j+2q_i^2q_j}$
- $1' = 1$ if (1) $q_j \leq \frac{q_i^2}{1-2q_i+2q_i^2}$ and $\lambda_H > \frac{-q_i^2+q_j-2q_iq_j+2q_i^2q_j}{-q_i-q_i^2+2q_i^2q_j}$ or (2) $q_j > \frac{q_i^2}{1-2q_i+2q_i^2}$

- Belief update decision (b_i^*):

Perception of q_j	Uniformly Credible				Opposite Types Not Credible			
$(k_j, \pi_j) \backslash (k_i, s_i)$	$(1,1)$	$(1,0)$	$(0,0)$	$(0,1)$	$(1,1)$	$(1,0)$	$(0,0)$	$(0,1)$
$(1,1)$	1	1	1	1	1	1	0	0
$(1,0)$	1^+	1	1	1	1^+	1	0	0
$(0,1)$	1	1	1	1^+	0	0	1	1^+
$(0,0)$	1	1	1	1	0	0	1	1

where $1^+ = 1^{***}$ above.

2. Sender equilibrium strategy: $\pi_j^* = s_j$ for all types.

3. Receiver posterior beliefs (based on π_j) on the SOW:

   - Construction 1 – uniformly greater $q_j > q_i$:
     - $(s_i, \pi_j) = (1,1)$: $\mu^{1*}_{(\pi_j=1,w=0)} = \frac{(1-q_i)(1-q_j)}{(1-q_i)(1-q_j)+q_iq_j}$
     - $(s_i, \pi_j) = (1,0)$: $\mu^{1*}_{(\pi_j=0,w=1)} = \frac{q_i(1-q_j)}{q_i(1-q_j)+(1-q_i)q_j}$
     - $(s_i, \pi_j) = (0,1)$: $\mu^{1*}_{(\pi_j=1,w=0)} = \frac{q_i(1-q_j)}{q_i(1-q_j)+(1-q_i)q_j}$
     - $(s_i, \pi_j) = (0,0)$: $\mu^{1*}_{(\pi_j=0,w=1)} = \frac{(1-q_i)(1-q_j)}{(1-q_i)(1-q_j)+q_iq_j}$
   - Construction 2 – the misaligned source not credible $q_{j,a} > q_i > q_{j,b}$:
     - $k_j = k_i$: replace q_j in the posterior beliefs of construction 1 with $q_{j,a}$.
     - $k_j \neq k_i$: replace q_j in the posterior beliefs of construction 1 with $q_{j,b}$.