Information and Motivated Reasoning: A Model of Selective Exposure

Tom (Hyeon Seok) Yu*

Last Update: March 13, 2022

Abstract

Previous research has documented the prevalence of selective exposure, the tendency to prefer and consume information that reinforces preexisting beliefs. Modeling individuals as motivated reasoners who face a tradeoff between accuracy ("getting it right") and 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 exposure. First, when the quality of information 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 increases the likelihood of information avoidance. 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 exposure. These results on exposure decisions yield direct implications for persuasion and polarization.

*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, since Lazarsfeld et al. (1948)'s study of the presidential election in 1940, scholars have noted people's tendency to expose themselves to information that confirms preexisting beliefs while dismissing contradictory information. One high profile account for such behavior is the motivated reasoning account (Kunda 1990), which considers the tradeoff between "getting it right" (the accuracy motive) and "desiring a certain conclusion to hold" (the directional motive) in people's minds to be the central driver.

For example, suppose a Republican voter cares deeply about family values – loyalty, for instance – and plans to vote for the Republican candidate in an upcoming election. But she comes across a headline hinting that the candidate 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 read the article? Would it matter if she does not recognize the source? And what if her perception of the source's credibility differs by its ideological slant? Addressing these questions on exposure is particularly relevant for related topics of persuasion and polarization since exposure to information often precedes a change in beliefs, which directly affects 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 accept or avoid action-relevant information from senders in the subsequent period. In addition to having 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 broadly applied to settings where strategic considerations affect the exchange and consumption of information, the exposition focuses on a setting closely related to the empirical literature: the sender is a news provider and the receiver is an individual with an action to take (e.g., a voter).

This paper contributes to the existing literature by making sense of seemingly inconsistent and disconnected stylized facts. On exposure, the model identifies the topic/issue as a critical part of individuals' 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 non-discriminatory (i.e., 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 generally 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 also reflects 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 determines the type of equilibrium sustainable, and in each, certain types of individuals are predicted to engage in selective exposure or complete information avoidance. This in turn reveals the difficulty of persuasion, which has been regularly documented in empirical studies (e.g., Guess et al. 2021; Peterson et al. 2021); 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 complete avoidance can impede, but not prevent, divergence in beliefs. Put another way, while selective exposure does lead to polarization, individuals' decisions to avoid information can lower the degree of divergence. This result 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 are 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 of exposure. The discrepancy results from the possibility of incurring a loss from exposing

[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).

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, another 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 the sender's information, as the key driver of exposure decisions. While a large gap necessarily requires the sender's information quality to be high, the receiver might choose to avoid information no matter how credible the sender is if one's information is deemed sufficiently accurate. Furthermore, the second construction addresses how the poor perception of the oppositional media affects polarization. The analysis shows that unfavorable perception alone may not be sufficient to induce divergence in beliefs. That is, convergence is possible as long as the aligned source sends unbiased signals to the receiver. However, this means a source that relays a biased set of information can cause 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; Hill 2017). 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 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 among motivated reasoners, 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) points to 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 expose themselves to 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 the 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 is an important gap to fill.[3]

[3] But how is "avoidance" defined? 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 some 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 the current model reveals subtleties consistent with these empirical studies.

A Model of Selective Exposure

The primary objective at hand is to construct a model that can account for individuals' information exposure/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 to a broad range 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.

The game proceeds as follows:

1. Period 0: Endowment Phase

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

    (b) Nature determines the directional motive (k_i) with $\Pr(k_i = 0) = \Pr(k_i = 1) =$

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

$\frac{1}{2}$ 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 $\mu^0_{(s_i,w)} = \Pr[w \in \{0, 1\}|s_i \in \{0, 1\}]$ 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 setting his signal.

(c) Upon observing the sender's type k_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 observes π_j without distortion, and she 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.

Before delving into specific utility structures that drive individuals' actions, there are at least three 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. In the real world, this could be a post she made in the past on social media about certain issues, which serves as a reference point when she considers taking another related action based on the newly received sender's information.[6] Lastly, another important point to note is the observability of the sender's type before observing his message. This was to allow the exposure decision ($a_{ij} = 1$) to yield further information about the SOW. In the news consumption setting, this would be equivalent to observing the headline from a recognized news provider (known k_j) and clicking the news link (exposure decision $a_{ij} = 1$) revealing the actual content of the article (π_j).

Separately, considering that the main focus of the model is receivers' exposure decisions, the role of senders might, at first, appear insignificant. Their role is important for two reasons: first, that the sender is also a motivated reasoner directly affects the receiver's exposure decision, which will become more evident in the latter's utility construction; the directional motive (i.e., type) of the sender matters when the receiver decides on her exposure decision. Second, more substantively, this allows the model to reflect a real-world setting where individuals obtain information from others in their network who likely possess information of comparable quality and similar motives (Mutz 2006). Besides, including a similarly motivated sender enables the model to capture the strategic aspect of the information sharing, the main benefit of adopting a game-theoretic framework.

[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] Mechanically, this intended action determines the baseline utility in the case the receiver does not expose herself to the sender's information at period 1. In other words, she decides to set $a_{ij} = 1$ if and only if the resulting utility is greater than the baseline utility determined by x_i^0.

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, λ_i is greater for the highly-committed individuals ($\lambda_H > \lambda_L$), meaning they incur greater loss from an action that contradicts their directional motives. Below specifies utility functions for each period and actor:[7]

Period 0:

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

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.

Period 1, Sender Utility:

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

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.

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

Period 1, Receiver Utility:

- Exposure decision (a_{ij}):

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

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

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

Receiver i's utility depends on her exposure decision a_{ij} and belief update decision b_i. The key 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.[8] This separation of the exposure and the belief update decisions sets the model apart from existing models that generally assume exposure to automatically entail a change in beliefs; such a construction better reflects an empirical regularity that exposure does not always imply persuasion (e.g., Knobloch-Westerwick and Meng 2009; Barnes et al. 2018).

Given a comparatively nontraditional setup of the game, some 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,

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

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.

Baseline: 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 trade-off 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.[9] The solution concept is perfect Bayesian.

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

[9] 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.

exposure at the subsequent stage. The type of pure-strategy equilibrium sustainable based on this threshold is defined as follows:[10]

Definition 2 (Types of Pure-Strategy Equilibrium)

(1) ***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.

(2) ***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.

(3) ***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 informative initial signals s_i. 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 opposite 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 formally define the selective exposure and different types of individuals for the exposition.

Definition 3 (Selective Exposure)

A receiver i sets $a_{ij} = 1$ and observes the sender's information π_j if and only if $k_j = k_i$.

[10]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.

Definition 4 (Types of Individuals)

(1) ***Confirmed Types:*** *Individuals who receive the aligned signal ($s_i = k_i$) at period 0.*

(2) ***Conflicted Types:*** *Individuals who receive the misaligned signal ($s_i \neq k_i$) at period 0.*

Most existing works equate selective exposure to prior selective exposure in the context of confirmation bias; individuals expose themselves to information that confirms their pre-existing beliefs. In the current model, by construction, the main criterion for exposure is the alignment in directional motives rather than the prior beliefs because the receiver does not observe the sender's message when making her exposure decision. While it is possible that individuals might choose to engage in selective exposure solely based on prior beliefs, existing empirical studies show the pattern of individuals consuming information and forming beliefs based on partisan and other ideological alignment cues rather than actual contents of the information (e.g., Cohen 2003; Peterson et al. 2021).

Result 1: Selective Exposure is Pervasive across Equilibria

In a setting where everyone receives the same quality of the signal at period 0, individuals are more likely to expose themselves to the sender's information (i.e., set $a_{ij} = 1$) if the sender's directional motive matches that of the receiver ($k_j = k_i$). The magnitude of λ_i determines the type of equilibrium sustainable.[11] Formal characterizations and proofs of all propositions are relegated to the Appendix.

Proposition 1 (Exposure Decision when Information Source is Known)

(1) *Accuracy equilibrium ($\lambda_L < \lambda_H < \bar{\lambda}^0$): if λ_i is sufficiently low ($\lambda_H < \bar{\lambda}^0$), confirmed types engage in selective exposure, and conflicted types set $a_{ij} = 1$.*

(2) *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 selective exposure.*

[11] Results below assume that individuals opt to set $a_{ij} = 1$ when they are indifferent. This assumption has a negligible impact on the results of the subsequent analysis.

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

- *High-types* (m_H): *confirmed types set* $a_{ij} = 0$; *conflicted types engage in selective exposure.*

- *Low-types* (m_L): *confirmed types set* $a_{ij} = 0$; *if* λ_L *is sufficiently high* ($\lambda_L > \underline{\lambda}_L^1 \equiv \frac{\rho(2q-1)}{\rho(2q-1)+1}$), *conflicted types engage in 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 (confirmed types) 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 loss from accepting information sent by the sender of the opposite type renders the exposure undesirable, and the sender's quality of information is simply not good enough. In expectation, the loss from getting the SOW wrong is determined by the sender's information quality, which is identical to that of the receiver. Therefore, without directly observing the sender's message π_j, the sure loss from exposing oneself to the wrong type eclipses any potential gain in posterior beliefs on the SOW. Separately, the lower bound on the low conflicted types' λ_i results from the fact that they need to care sufficiently about directional motives to expose themselves to directionally-aligned sender's message even if they might eventually see a misaligned message from the latter.

Now consider the individuals' belief update decisions given their exposure decisions stated in proposition 1.

Proposition 2 (Belief Update Decision when Information Source is Known)

(1) *Accuracy equilibrium* ($\lambda_L < \lambda_H < \bar{\lambda}^0$):, *confirmed types set* $b_i = 1$ *if and only of* $\pi_j = k_i$; *conflicted types set* $b_i = 1$ *if* λ_i *is sufficiently high* ($\lambda_L > \frac{2q-1}{2q+1}$).

(2) *Directional motive equilibrium* ($\bar{\lambda}^0 < \lambda_L < \lambda_H$): *individuals' beliefs are not updated.*

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

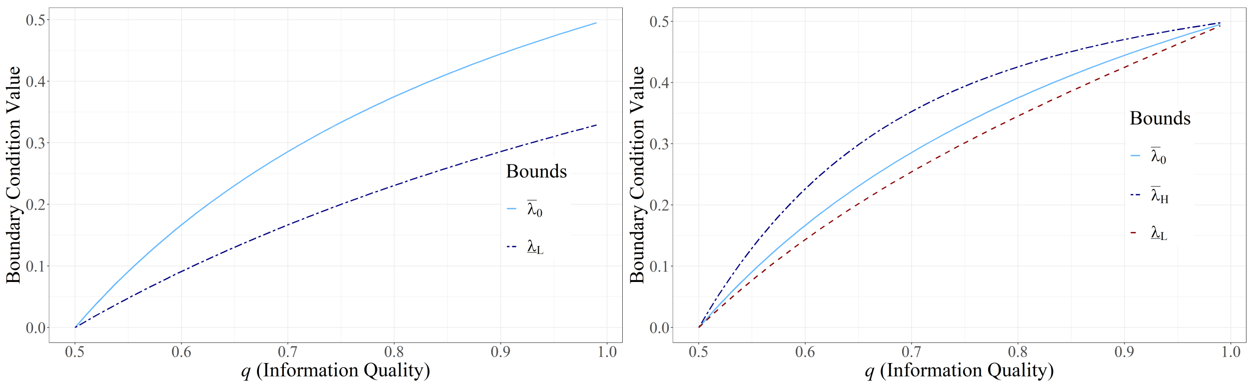
- *High-types (m_H): confirmed types set $b_i = 0$; conflicted types set $b_i = 1$ if λ_H is sufficiently low ($\lambda_H < \bar{\lambda}_H \equiv \frac{-1+3q-3q^2+2q^3}{q-q^2+2q^3}$).*
- *Low-types (m_L): confirmed types set $b_i = 0$; if conflicted types set $b_i = 1$ if λ_L is sufficiently high ($\lambda_L > \underline{\lambda}_L^1 \equiv \frac{-q+3q^2-2q^3-\rho+3q\rho-3q^2\rho+2q^3\rho}{q+q^2-2q^3+q\rho+q^2\rho+2q^3\rho}$).*

Similar to the exposure decision, the conflicted types need to care sufficiently about directional motives to update their beliefs according to the message that contradicts their initial signal ($\pi_j \neq s_i$) in the accuracy equilibrium. Individuals' posterior beliefs on the SOW do not change in the directional motive equilibrium, as the sender's information does not provide any meaningful information; recall that senders set $\pi_j = k_j$ regardless of the signal s_j they receive at period 0. Lastly, in the separating equilibrium, the confirmed types do not update their beliefs given their preceding no exposure decisions. The additional boundary conditions on the conflicted high and low types are necessary for them to update beliefs based on messages $\pi_j \neq k_i$ and $\pi_j \neq s_i$, respectively. These belief update decisions yield direct implications for persuasion.

Corollary 1 (Persuasion when Information Source is Known)
If q is identical across individuals, persuasion is only possible among the conflicted types.

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 individuals, who receive signals s_i that do not align with their types k_i, are the only ones capable of being persuaded. In the separating equilibrium, the condition is more stringent compared to the accuracy equilibrium in that the message needs to be from an aligned sender. A close gap between the boundary conditions on λ_i in figure 1 (b) shows the difficulty of belief update, hence persuasion to take place: for the high-type individuals, λ_H needs to be such that $\lambda_H \in (\bar{\lambda}_0, \bar{\lambda}_H)$ and the low-type individuals, $\lambda_L \in (\underline{\lambda}_L, \bar{\lambda}_0)$. As shown in the figure, the area between these curves is not large.

Figure 1: Comparative Statics on the Boundary Conditions by Equilibria



(a) Accuracy Equilibrium

(b) Separating Equilibrium

Note: λ_L needs to be above the depicted lower bound $\underline{\lambda}_L$ and below the cutpoint $\bar{\lambda}^0$ for the equilibria to be sustainable as stated. Conversely, λ_H needs to be below the upper bound $\bar{\lambda}_H$ for the separating equilibrium to be sustainable.

Result 2: Selective Exposure Leads to Polarization, but Complete Avoidance Can Impede Polarization in Beliefs

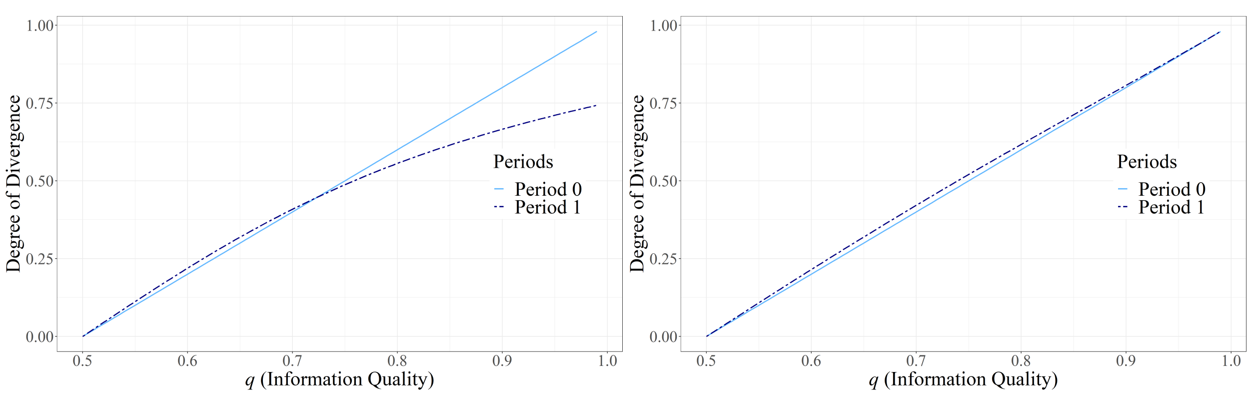
Propositions 1 and 2 revealed the prevalence of selective exposure and the difficulty of persuasion. What do these results, then, imply about the polarization in beliefs? This subsection carries out a simple analysis that addresses this question in a setup where a representative set of receivers' expected posterior beliefs after one interaction with a sender are computed to measure the degree of divergence ex-post predicted exposure.[12]

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 greater

[12] 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.

in the separating than the accuracy equilibrium. Each types' contribution to polarization flips in these two equilibria: in the accuracy equilibrium, the confirmed types who only expose themselves to aligned sender's confirmatory message (i.e., $k_j = \pi_j = k_i$) results in updating their beliefs toward one direction, while the conflicted types' exposure and belief update decisions lead them to converge, thereby contributing to closing the gap in beliefs. On the contrary, in the separating equilibrium, figure 5 (b) in the Appendix, which depicts the actual positions of expected posteriors after one interaction, reveals that a greater degree of divergence results from the conflicted types engaging in selective exposure and not the confirmed types avoiding exposure altogether. This results from the conflicted types only exposing themselves to the aligned sender; although they can update their beliefs to both confirmatory and contradictory messages, the degree of the update is greater when $\pi_j \neq k_j$, as it reveals that the sender is a low-type, which means the latter's signal is more informative about the SOW. Consequently, these conflicted types update beliefs more toward the direction consistent with their priors, thereby contributing to a wider divergence. In sum, individuals who engage in selective exposure cause further polarization, and complete avoidance (i.e., no exposure) can impede the divergence.

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



(a) Accuracy Equilibrium (0.441)   (b) Separating Equilibrium (0.502)

Empirical Discussion: Persuasion, Partisan Cheerleading, Avoidance

The analysis reveals the prevalence of selective exposure when the directional motive is present. Although the current construction assumes individuals to be selective on the directional motive dimension at the exposure stage, the strong influence of preexisting beliefs theorized in earlier works (Rabin and Schrag 1999) is also present in the model, reflected by the additional boundary conditions needed for individuals to update beliefs based on messages that contradict their initial signals. In addition, the slim possibility for persuasion stated in corollary 1 translates to 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). In particular, the difficulty of persuading the confirmed types – those who receive confirmatory signals at period 0 – is consistent with the empirical finding that partisans whose initial attitudes are anchored on party positions are generally not susceptible to persuasion (Cobb and Kuklinski 1997).[13]

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 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

[13]With its construction that allows even strong partisans to have "conflicting priors," the model highlights the importance of analyzing the initial beliefs held by individuals. Specifically, the analysis predicts strong but conflicted partisans to be susceptible to persuasion based on information that contradicts their priors but aligns with their partisan motives. Although Cohen (2003)'s experiment provides a seemingly consistent result, more direct empirical testing should be conducted to validate this prediction.

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.[14]

Third, the analysis on the implication of selective exposure for polarization in beliefs confirms the existing claim that selective exposure leads to polarization (e.g., Stroud 2010). In different types of equilibria analyzed, individuals who engaged in selective exposure contributed the most to the divergence in beliefs. Another notable result is that avoidance of information can impede such a divergence. In the separating equilibrium, for example, confirmed individuals who prefer not to expose themselves even to the aligned source due to the possibility of receiving a contradictory message do not contribute to the convergence with their unmoving beliefs, but they do not cause further 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

[14]Taking a closer look at the set of topics adopted in the two studies, Peterson and Iyengar (2021) seem to have adopted questions where partisan respondents might assign greater 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, the "psychic" rewards for cheerleading might have exceeded financial rewards in their studies.

"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.

Result 3: Uncertainty Renders the Exposure Less Likely

Removing the visibility of the sender's directional motive leads the receiver – the confirmed types, in particular – to avoid exposure. While the general pattern across equilibria remains similar to that of the baseline, the confirmed types no longer expose themselves to the sender's information in all equilibria.

Proposition 3 (Exposure when the Information Source is Unknown)

(1) *Accuracy Equilibrium ($\lambda_L < \lambda_H < \bar{\lambda}^0$): the confirmed types set $a_{ij} = 0$, and the conflicted types set $a_{ij} = 1$.*

(2) *Directional motive Equilibrium ($\bar{\lambda}^0 < \lambda_L < \lambda_H$): all types set $a_{ij} = 0$.*

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

- *High-types (m_H): the confirmed types set $a_{ij} = 0$, and the conflicted types set $a_{ij} = 1$ if and only if λ_H is sufficiently low ($\lambda_H < \bar{\lambda}_H \equiv \frac{2-4q-\rho+2q\rho}{1-4q-\rho+2q\rho}$).*

- *Low-types (m_L): the confirmed types set $a_{ij} = 0$, and the conflicted types set $a_{ij} = 1$ if and only if λ_L is sufficiently high ($\lambda_L > \underline{\lambda}_L \equiv \frac{\rho(2q-1)}{1+\rho(2q-1)}$).*

The confirmed types' decision of no 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 prefers not to expose herself to a sender's message due to the potential loss from exposing herself to the opposite type eclipsing the potential gain in posterior beliefs on the SOW. The conflicted types also no longer expose themselves to the sender's information in the directional motive equilibrium. Uncertainty can induce more exposure for such types in the separating equilibrium compared to the baseline, but relatively tight boundary conditions must be satisfied.

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 a 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 document 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 the "oppositional" media on selective exposure. The receiver is assumed to know the sender's directional motive as in the baseline, and the analysis continues to focus on the accuracy equilibrium.

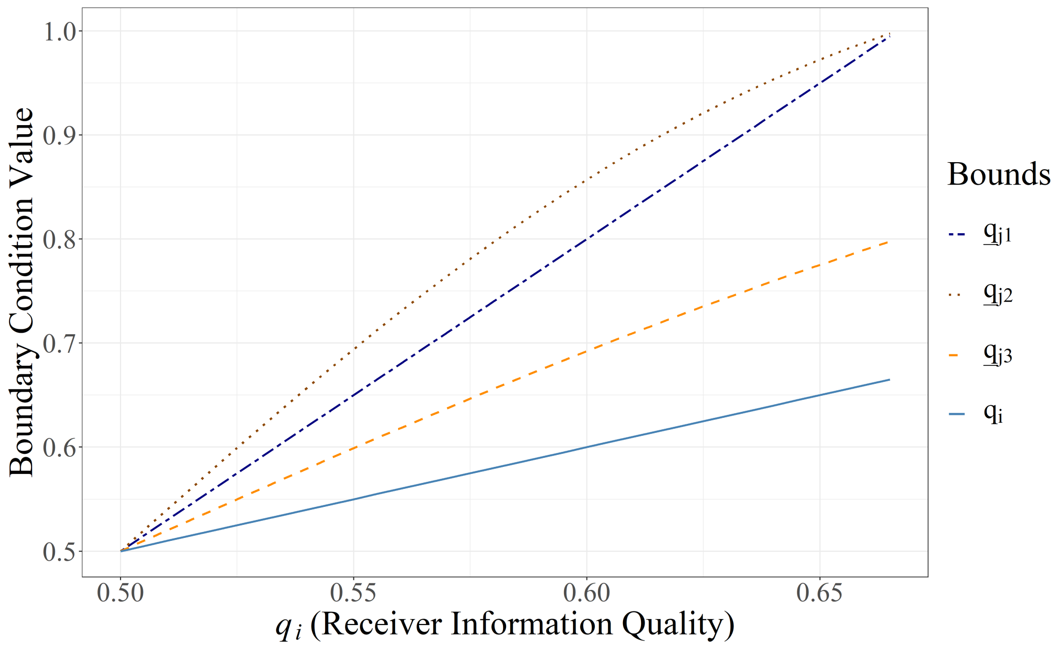
Result 4: A Sufficiently Large Gap in Information Quality Induces Exposure

At first, this result might appear as a straightforward confirmation of the existing claim that the higher quality of information increases the likelihood of exposure, hence the change in beliefs. 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.

Proposition 4 (Exposure and Belief Update when Senders are Uniformly More Credible) *If the gap between the receiver and the sender's information quality is sufficiently large ($q_i < \frac{2}{3}, q_j > \underline{q}_{j,1} \equiv 3q_i - 1$), the receiver sets $a_{ij} = 1$. If $q_j > \underline{q}_{j,2} \equiv \frac{-q_i+3q_i^2}{2-6q_i+6q_i^2}$, then the receiver sets $b_i = 1$.*

Note that $q_j > \underline{q}_{j,2}$ is a sufficient condition for both $a_{ij} = b_i = 1$, since $\underline{q}_{j,2} > \underline{q}_{j,1}$. Put another way, a sufficiently large gap in the quality of information can induce complete exposure and update in beliefs toward the direction suggested by the sender's message. The importance of the gap, rather than the magnitude of the sender's information quality, is evident in figure 3 that shows different lower-bounds on the sender's information quality q_j to sustain the equilibrium. As shown, once the receiver's information accuracy surpasses a certain point ($\frac{2}{3}$ in this setup), she no longer exposes herself to information from those with misaligned directional motives (i.e., $k_j \neq k_i$). In other words, if the receiver knows that her information quality is good enough, she is not willing to expose herself to information from a source known to have opposite directional motives.

Figure 3: Comparative Statics on Boundary Conditions given Senders Uniformly Credible



Note: The solid line represents the receiver's information quality (q_i) for comparison. $\underline{q}_{j,1}$ represents the lower bound on the sender's information quality for a confirmed-type receiver to be willing to expose herself to information from a misaligned source ($k_j \neq k_i$). $\underline{q}_{j,2}$ represents the lower bound for updating belief based on a contradictory message ($\pi_j \neq s_i$) for the confirmed type. $\underline{q}_{j,3}$ represents the lower bound for updating belief based on a contradictory message ($\pi_j \neq s_i$) for the conflicted type.

Perceiving Oppositional Media to be Less 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 less credible (i.e., $q_j < q_i$). 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 assumption returns the result similar to the directional motive equilibrium from the baseline, except that individuals update their beliefs on the SOW upon exposure to information from the aligned sources.

Proposition 5 (Exposure and Belief Update when Opposite Types are Less Credible) *If the receiver perceives the opposite type sender's information quality to be low ($q_{j,k_j \neq k_i} < q_i$, while $q_{j,k_j=k_i} > q_i > \frac{1}{2}$), the receiver sets $a_{ij} = 1$ if and only if $k_j = k_i$. If $q_j > \underline{q}_{j,2} \equiv \frac{-q_i+3q_i^2}{2-6q_i+6q_i^2}$, then the receiver sets $b_i = 1$.*

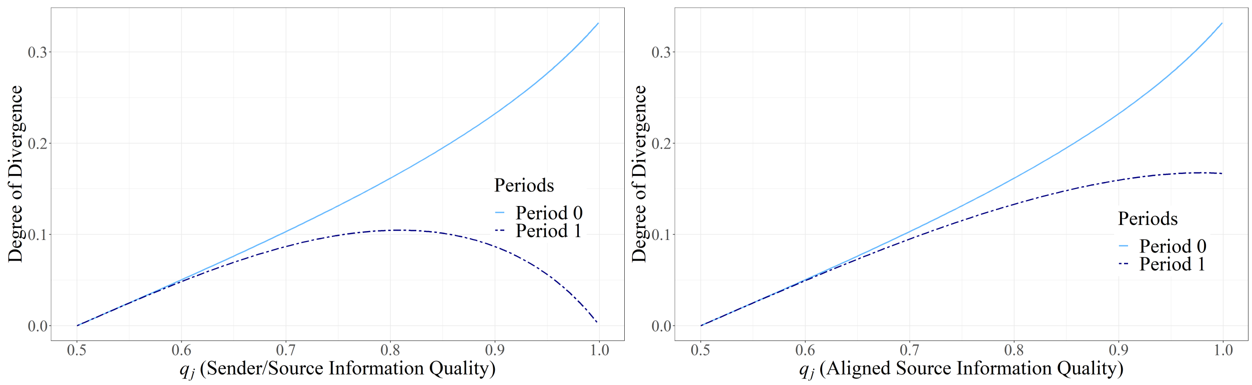
Once the receiver sets $a_{ij} = 1$ after observing that the sender is of an aligned type, the identical boundary condition on q_j from the previous construction applies for the receiver to always update her belief according to the sender's message π_j.

Result 5: 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 4, which also shows the degree of divergence at period 1 for the first construction that assumes 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 poorer than her own. 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 4: Degree of Divergence in Posterior Beliefs at Period 1 of the Extension



(a) q_j Uniformly Greater (0.067) (b) Opposite Types Not Credible (0.117)

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 complete 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 a 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.[15]

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 a 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.

[15] 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 rely on topics related to candidates in presidential elections (Iyengar et al. 2008).

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) 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 hold varying 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, complete avoidance can impede, but not prevent, divergence in beliefs. Although selective exposure generally widens the gap, individuals' decision to avoid/reject information altogether can translate to no update in beliefs, thereby not causing further 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.[16] 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.

[16]This is related to the idea of endogenizing the directional motive parameter, v, 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.

Cobb, Michael D. and James H. Kuklinski (1997). "Changing Minds: Political Arguments and Political Persuasion". *American Journal of Political Science* 41 (1): pp. 88–121.

Cohen, Geoffrey L. (2003). “Party Over Policy: The Dominating Impact of Group Influence on Political Beliefs”. *Journal of Personality and Social Psychology* 85 (5): pp. 808–822.

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*.

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.

Mutz, Diana C. (2006). *Hearing the Other Side: Deliberative versus Participatory Democracy*. Cambridge University Press.

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 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.

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.

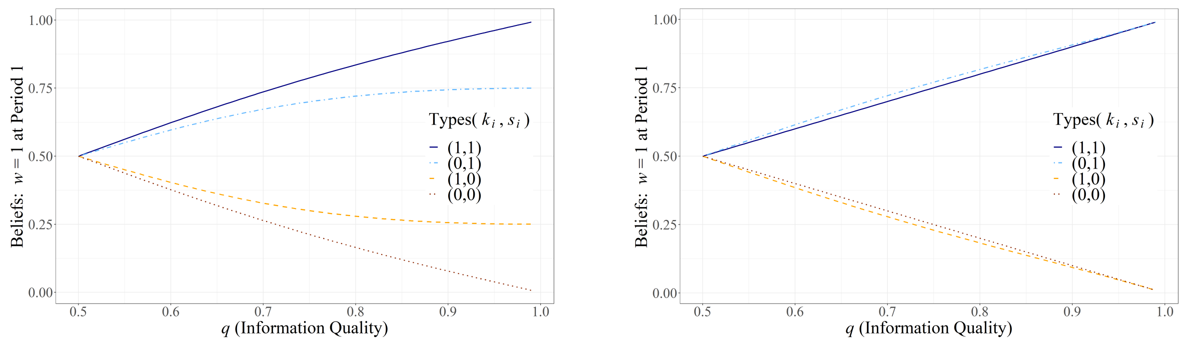
Supporting Information for
Information and Motivated Reasoning:
A Model of Selective Exposure

Table of Contents

*Stanford University Graduate School of Business. Email: tom.hs.yu@stanford.edu

A Additional Figures

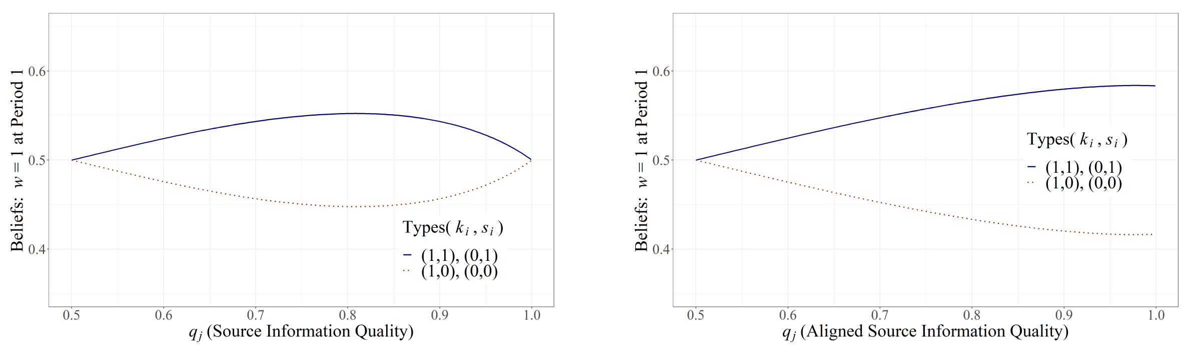
Figure 5: Posterior Beliefs at Period 1 by Types



(a) Accuracy Equilibrium

(b) Separating Equilibrium

Figure 6: Extension 2 Posterior Beliefs at Period 1 by Varying Degrees of Credibility



(a) q_j Uniformly Greater

(b) Opposite Types Not Credible

B Proof and Formal Characterizations of Propositions

Below provides detailed proofs of propositions 1 and 2 stated in the main text. These proofs are followed by formal characterizations of propositions 3 and 4. Proofs of propositions 3 and 4 follow identical procedures to those of propositions 1 and 2, 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

[1]Detailed proofs with derivations of relevant boundary conditions are available upon request.

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 ($s_i \neq k_i$):

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

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

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

B.1 Proofs of Propositions 1 & 2 (Known Information Source Equilibria)

Note that the receiver's exposure decision a_{ij}^* and belief update decision b_i^* in each equilibrium corresponds to propositions 1 and 2, respectively.

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

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

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

     where 1^* marks the indifference.
   - Belief update decision (b_i^*)

$\pi_j \setminus (k_i, s_i)$	$(1,1)$	$(1,0)$	$(0,1)$	$(0,0)$
1	1	1^+	1	0
0	0	1	1^+	1

     where $1^+ = 1$ if $\lambda_L > \underline{\lambda}_L \equiv \frac{2q-1}{2q+1}$
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 $x_i^0 = s_i \neq k_i$, 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 the 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.

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 conflicted 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, but with different λ_i. When considering possible deviations on selecting π_i as senders, note that $\lambda_L < \lambda_H$ necessarily renders the conflicted 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. $\square$

We first begin with the receiver's equilibrium behavior upon observing the paired sender's type. Denote the receiver as i and the sender j.

(1) $(k_i, s_i, m_i) = (1, 1, m_H)$, $k_j = 1$:
First, the receiver needs to form expectations on the probability of the sender's message being wrong given the sender's type:

$$\Pr[\pi_j \neq w|k_j] = \Pr[\pi_j \neq w] = \Pr[\pi_j = 1] \cdot \Pr[w = 0|\pi_j = 1] + \Pr[\pi_j = 0] \cdot \Pr[w = 1|\pi_j = 0]$$

where

$$\begin{aligned}
\Pr[\pi_j = 1|k_j = 1] &= \Pr[\pi_j = 1] \quad (\text{since } s_j \text{ is indept. of } k_j \ \&\ \pi_j = s_j) \\
&= \Pr[\pi_j = 1|w = 0] \cdot \Pr[w = 0] + \Pr[\pi_j = 1|w = 1] \cdot \Pr[w = 1] \\
&= (1 - q)^2 + q^2 \\
\Pr[\pi_j = 0|k_j = 1] &= 1 - ((1 - q)^2 + q^2) = 2q(1 - q)
\end{aligned}$$

Define $\mu^1_{(\pi_j,w)}$ as the receiver's posterior belief on the SOW upon observing π_j. Then,

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

Collecting the terms,

$$\Pr[\pi_j \neq w] = ((1-q)^2+q^2)\cdot\frac{(1-q)^2}{(1-q)^2+q^2} + 2q(1-q)\cdot\frac{1}{2} = 1-q$$

Now we compute the expected payoffs of the exposure decision a_{ij}.

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

Therefore, the receiver is indifferent between $a_{ij}=1$ and $a_{ij}=0$. If the receiver sets $a_{ij}=1$, her expected payoffs for each belief update decision b_i for each π_j are as follows:

$$E[u_i|b_i=1,\pi_j=1] = -(1-\lambda_H)\mu^1_{(\pi_j=1,w=0)} = -(1-\lambda_H)\frac{(1-q)^2}{(1-q)^2+q^2}$$
$$E[u_i|b_i=0] = -(1-\lambda_H)(1-q)$$

Since $\frac{(1-q)^2}{(1-q)^2+q^2} < 1-q$, $b_i^*=1$ if $\pi_j=1$.

$$E[u_i|b_i=1,\pi_j=0] = -(1-\lambda_H)\mu^1_{(\pi_j=0,w=1)} = -(1-\lambda_H)\frac{1}{2}$$
$$E[u_i|b_i=0] = -(1-\lambda_H)(1-q)$$

Since $q>\frac{1}{2}$, $b_i^*=0$ if $\pi_j=0$.

(2) $(k_i,s_i,m_i)=(1,1,m_H)$, $k_j=0$:
The only part that changes from (1) is the first part of the receiver's expected payoffs, since the sender is known to be of an opposite type:

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

Therefore, $a_{ij}^*=0$ if $k_j=0$. Since $a_{ij}^*=0$, $b_i^*=0$.

(3) $(k_i, s_i, m_i) = (1, 0, m_H)$, $k_j = 1$:
Now this "conflicted" type observes a signal that contradicts her type at period 0, the initial posterior belief on the SOW ($\mu^0_{(s_i,w)}$) changes from (1) and (2) above. The derivation is identical as before.

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

Therefore, $a^*_{ij} = 1$ if $k_j = 1$. Now consider her belief update decision for each π_j:

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

Comparison to the no update shows that $b^*_i = 1$ if and only if $\lambda_H > \frac{2q-1}{2q+1} \equiv \underline{\lambda}_L$ when $\pi_j = 1$.[2]
If $\pi_j = 0$, $b^*_i = 1$, since $\frac{(1-q)^2}{(1-q)^2+q^2} < 1 - q$.

(4) $(k_i, s_i, m_i) = (1, 0, m_H)$, $k_j = 0$:

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

Therefore, the receiver is indifferent. When she sets $a_{ij} = 1$, her expected payoffs of her belief update decision remain identical to the (3) above.

Now we confirm that the sender would not deviate from setting $\pi_j = s_j$. An important point to note here is that the receiver's exposure decision does not depend on π_j; rather, the sender's type determines the possibility of acceptance, so the sender's hands are, in a sense, tied. Consider the conflicted high type $(k_j, s_j, m_j) = (1, 0, m_H)$'s expected payoffs:

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

Since the last part is identical, the initial assumption that $\lambda_H < \bar{\lambda}^0$ leads the first expected payoff strictly dominated, hence $\pi^*_j = 0$, no deviation. Assuming that the additional boundary condition on λ_i is satisfied, then, the equilibrium is sustainable as conjectured. □

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

1. Receiver equilibrium strategies:

[2] By assumption, $\lambda_L < \lambda_H$. To ensure that both high and low types pool together, the boundary condition is defined in terms of the low-type.

- Exposure decisions (a_{ij}^*)

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

where 1^* marks the indifference.

- Belief update decisions: $b_i^* = 0$ for all types.[3]

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

3. Receiver posterior beliefs (based on π_j) on the SOW: $\mu^{1*}_{(\pi_j,w)} = \mu^0_{(s_i,w)} \quad \forall \pi_j$.

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., $x_i^0 = k_i$ 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

A symmetric 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:

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

□

[3]To be more precise, the receiver is indifferent about setting b_i upon observing $\pi_j = k_i$. However, even if she sets $b_i = 1$, her posterior belief does not change, hence equivalent to no update in beliefs. Lemma 3 formally establishes this claim.

First, consider the receiver's equilibrium behavior. Based on lemmas 2 and 3, it follows that the receiver never exposes herself to a message from the sender holding opposite directional motives, while indifferent when seeing that the sender is of the same type. More formally,

1. $k_j = k_i$:

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

   where $\mu^0_{(s_i,w)}$ is the initial posterior belief from period 0 that the action taken in period 0 is wrong (i.e., $x_i^0 \neq w$) based on observing the signal s_i. Given the equality, the receiver is indifferent about exposure.

2. $k_j \neq k_i$:

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

   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. WLOG, consider the case of $(k_j, s_j) = (1, 0)$, who is expected to set $\pi_j = 1$. The sender's expected payoffs are as follows:

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

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 $k_j = 0$. As we have confirmed that senders would not deviate, the equilibrium is sustainable as characterized. □

Corollary 2
Receivers' posterior beliefs on SOW do not change following the exposure, 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 update in beliefs 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			
$k_j \setminus (k_i, s_i)$	$(1,1)$	$(1,0)$	$(0,1)$	$(0,0)$	$(1,1)$	$(1,0)$	$(0,1)$	$(0,0)$
1	0	1	0	0	0	1^*	0	0
0	0	0	1	0	0	0	1^*	0

where $1^* = 1$ if and only if $\lambda_L > \underline{\lambda}_{L,1} \equiv \frac{\rho(2q-1)}{\rho(2q-1)+1}$;

- 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	1	1^+	0	0	1^{++}	1	0
0	0	1^+	1	0	0	1	1^{++}	0

where $1^+ = 1$ if and only if $\lambda_H < \bar{\lambda}^1_{H,1} \equiv \frac{-1+3q-3q^2+2q^3}{q-q^2+2q^3}$; $1^{++} = 1$ if and only if $\lambda_L > \underline{\lambda}_{L,2} \equiv \frac{-q+3q^2-2q^3-\rho+3q\rho-3q^2\rho+2q^3\rho}{q+q^2-2q^3+q\rho+q^2\rho+2q^3\rho}$

2. Sender equilibrium strategies:

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

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

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

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. Before the derivation, note that after setting $a_{ij} = 1$, observing $k_j \neq \pi_j$ perfectly reveals that the sender is a low-type individual, which simplifies the derivation of expected payoffs.

We have four different cases to consider.

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

$$\begin{aligned}
\Pr[\pi_j = 1 | k_j = 1] &= \Pr[m_i = m_H] \cdot \Pr[\pi_j = 1 | k_j = 1, m_i = m_H] \\
&\quad + \Pr[m_i = m_L] \cdot \Pr[\pi_j = 1 | k_j = 1, m_i = m_L] \\
&= \rho + (1-\rho)\{(1-q)^2 + q^2\} \\
\Pr[\pi_j = 0 | k_j = 1] &= 1 - (\rho + (1-\rho)\{(1-q)^2 + q^2\}) = (1-\rho)(2q(1-q)) \\
\mu^1_{(\pi_j=1,w=0)} &= \frac{(1-q)(\rho + (1-\rho)(1-q))}{(1-q)(\rho + (1-\rho)(1-q)) + q(\rho + (1-\rho)q)} \\
\mu^1_{(\pi_j=0,w=1)} &= \frac{(1-q)(q)}{(1-q)(q) + q(1-q)} = \frac{1}{2}
\end{aligned}$$

Then, this type's expected payoffs for each exposure decision is as follows:

$$\begin{aligned}
E[u_i | a_{ij} = 1] &= -(1-\lambda_H)[(1-q)(\rho + (1-\rho)(1-q)) + (1-\rho)q(1-q)] \\
E[u_i | a_{ij} = 0] &= -(1-\lambda_H)(1-q)
\end{aligned}$$

Comparison shows that no exposure strictly dominates the other, hence $a^*_{ij} = 0$. This further implies that $b^*_i = 0$ if $k_j = 1$.

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

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

Collecting the terms, this type's expected payoffs for each exposure decision is as follows:

$$\begin{aligned}
E[u_i | a_{ij} = 1] &= -(1-\lambda_H)[(1-\rho)((1-q)^2 + q^2) \cdot \frac{(1-q)^2}{(1-q)^2 + q^2)} + (\rho + (1-\rho)\{2q(1-q)\})] \\
E[u_i | a_{ij} = 0] &= -(1-\lambda_H)(1-q)
\end{aligned}$$

Comparison shows that no exposure strictly dominates the other, hence $a^*_{ij} = 0$. This further implies that $b^*_i = 0$ if $k_j = 0$.

(3) $(k_i, s_i, m_i) = (1, 0, m_H)$, $k_j = 1$:

$$\Pr[\pi_j = 1|k_j = 1] = \rho + (1-\rho)\{2q(1-q)\}$$
$$\Pr[\pi_j = 0|k_j = 1] = (1-\rho)(q^2 + (1-q)^2)$$
$$\mu^1_{(\pi_j=1,w=0)} = \frac{(q)(\rho + (1-\rho)(1-q))}{(q)(\rho + (1-\rho)(1-q)) + (1-q)(\rho + (1-\rho)q)}$$
$$\mu^1_{(\pi_j=0,w=1)} = \frac{(1-q)^2}{(1-q)^2 + q^2}$$

Then, this type's expected payoffs for each exposure decision is as follows:

$$E[u_i|a_{ij} = 1, k_j = 1] = -(1-\lambda_H)[\rho + (1-\rho)\{2q(1-q)\} \cdot \frac{(q)(\rho + (1-\rho)(1-q))}{(q)(\rho + (1-\rho)(1-q)) + (1-q)(\rho + (1-\rho)q)}$$
$$+ (1-\rho)(q^2 + (1-q)^2) \cdot \frac{(1-q)^2}{(1-q)^2 + q^2}]$$
$$E[u_i|a_{ij} = 0, k_j = 1] = -(1-\lambda_H)q$$

Comparison shows that exposure strictly dominates the other, hence $a^*_{ij} = 1$. Now we consider the belief update decision of this type for each possible message from the sender:

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

Comparison to the baseline (i.e., $b_i = 0$) shows that if $\pi_j = 1$, $b_i = 1$ strictly dominates no updating, and if $\pi_j = 0$, $b^*_i = 1$ if and only if

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

(4) $(k_i, s_i, m_i) = (1, 0, m_H)$, $k_j = 0$:

$$\Pr[\pi_j = 1|k_j = 0] = (1-\rho)(2q(1-q))$$
$$\Pr[\pi_j = 0|k_j = 0] = \rho + (1-\rho)(q^2 + (1-q)^2)$$
$$\mu^1_{(\pi_j=1,w=0)} = \frac{1}{2}$$
$$\mu^1_{(\pi_j=0,w=1)} = \frac{(1-q)(\rho + (1-\rho)(1-q))}{(1-q)(\rho + (1-\rho)(1-q)) + (q)(\rho + (1-\rho)q)}$$

Then, this type's expected payoffs for each exposure decision is as follows:

$$E[u_i|a_{ij}=1,k_j=0] = -\lambda_H - (1-\lambda_H)[(1-\rho)(2q(1-q))\cdot\frac{1}{2}$$
$$+ (\rho+(1-\rho)(q^2+(1-q)^2))\cdot\frac{(1-q)(\rho+(1-\rho)(1-q))}{(1-q)(\rho+(1-\rho)(1-q))+(q)(\rho+(1-\rho)q)}]$$
$$E[u_i|a_{ij}=0,k_j=0] = -(1-\lambda_H)q$$

Comparison shows that $a_{ij}=1$ is strictly dominated, hence $a_{ij}^*=0$. This further implies that $b_i^*=0$ if $k_j=0$.

(5) $(k_i,s_i,m_i)=(1,1,m_L)$, $k_j=1$:
Since this is a confirmed low-type, the expected payoffs of the exposure decision remains identical to case (1) above, but with different λ. Computation shows that the same result holds; $a_{ij}^*=0$.

(6) $(k_i,s_i,m_i)=(1,1,m_L)$, $k_j=0$:
This case is also identical to the case (2) above; $a_{ij}^*=0$.

(7) $(k_i,s_i,m_i)=(1,0,m_L)$, $k_j=1$:

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

Comparison shows that $a_{ij}=1$ strictly dominates the other if and only if

$$\lambda_L > \frac{\rho(2q-1)}{\rho(2q-1)+1} \equiv \underline{\lambda}_{L,1}$$

Now assuming $\lambda_L > \underline{\lambda}_{L,1}$, consider her belief update decision:

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

Comparison to the baseline $b_i=0$ shows that if $\pi_j=1$, $b_i=1$ if and only if

$$\lambda_L > \frac{-q+3q^2-2q^3-\rho+3q\rho-3q^2\rho+2q^3\rho}{q+q^2-2q^3+q\rho+q^2\rho+2q^3\rho} \equiv \underline{\lambda}_{L,2}$$

Conversely, if $\pi_j=0$, $b_i=1$ strictly dominates the other, so $b_i^*=1$.

(8) $(k_i,s_i,m_i)=(1,0,m_L)$, $k_j=0$:

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

Comparison shows that $a_{ij}=1$ is strictly dominated, hence $a_{ij}^*=0$, which further implies $b_i^*=0$.

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 different "conflicted" types who are more likely to deviate: $(k_j,s_j,m_j)=(1,0,m_H)$ and $(1,0,m_L)$.

(1) $(k_j,s_j,m_j)=(1,0,m_H)$, conjectured to set $\pi_j=1$:

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

By the initial assumption that $\lambda_H>\bar{\lambda}^0$, the expected payoff from setting $\pi_j=0$ is strictly dominated, hence $\pi_j^*=1$.
(2) $(k_j,s_j,m_j)=(1,0,m_L)$, conjectured to set $\pi_j=0$:

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

Conversely, by the initial assumption that $\lambda_L<\bar{\lambda}^0$, the expected payoff from setting $\pi_j=1$ is strictly dominated, hence $\pi_j^*=0$. Then, assuming the additional boundary conditions identified for certain types of receivers hold, the equilibrium is sustainable as characterized. □

B.2 Formal Characterization of Proposition 3 (Unknown Info Source)

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

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

(k_i,s_i)	$(1,1)$	$(1,0)$	$(0,0)$	$(0,1)$
	0	1	1	0

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

3. Receiver posterior beliefs:

   - Posterior belief on sender type: $\tilde{k}_j^*=\frac{1}{2}$.

- 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}):

(k_i, s_i)	$(1,1)$	$(1,0)$	$(0,0)$	$(0,1)$
	0	0	0	0

2. Sender equilibrium strategy: $\pi^*_j = k_j$ for all types.
3. Receiver posterior beliefs:
   - Posterior belief on sender type: $\tilde{k}^*_j = \frac{1}{2}$.
   - Posterior belief on SOW (based on π_j): $\mu^{1*}_{\pi_j,w=1} = \mu^0$ for all π_j.

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

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

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

   where $1^* = 1$ if and only if $\lambda_H < \bar{\lambda}_H \equiv \frac{2-4q-\rho+2q\rho}{1-4q-\rho+2q\rho}$; $1^+ = 1$ if and only if $\lambda_L > \underline{\lambda}_L \equiv \frac{\rho(2q-1)}{1+\rho(2q-1)}$

2. Sender equilibrium strategy:

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

3. Receiver posterior beliefs on the SOW (based on π_j):
   - $(s_i, \pi_j) = (1, 1)$:
     $\mu^{1*}_{1,w=0} = \frac{(1-q)[\frac{1}{2}\{\rho+(1-\rho)(1-q)\}\frac{1}{2}\{(1-\rho)(1-q)\}]}{(1-q)[\frac{1}{2}\{\rho+(1-\rho)(1-q)\}\frac{1}{2}\{(1-\rho)(1-q)\}]+(q)[\frac{1}{2}\{\rho+(1-\rho)(q)\}\frac{1}{2}\{(1-\rho)(q)\}]}$
   - $(s_i, \pi_j) = (1, 0)$:
     $\mu^{1*}_{0,w=1} = \frac{(q)[\frac{1}{2}\{(1-\rho)(1-q)\}+\frac{1}{2}\{(\rho)+(1-\rho)(1-q)\}]}{(q)[\frac{1}{2}\{(1-\rho)(1-q)\}+\frac{1}{2}\{(\rho)+(1-\rho)(1-q)\}]+(1-q)[\frac{1}{2}\{(1-\rho)q\}+\frac{1}{2}\{\rho+(1-\rho)(q)\}]}$
   - $(s_i, \pi_j) = (0, 1)$ and $(s_i, \pi_j) = (0, 0)$: replace the initial posterior in the corresponding expressions above.

B.3 Formal Characterization of Proposition 4 (Heterogeneous q)

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

   - Exposure decision (a_{ij}^*):

Perception of q_j	Uniformly Credible				Opposite Types Not Credible			
$k_j \setminus (k_i, s_i)$	$(1,1)$	$(1,0)$	$(0,1)$	$(0,0)$	$(1,1)$	$(1,0)$	$(0,1)$	$(0,0)$
1	1	1	1	1^*	1	1	0	0
0	1^*	1	1	1	0	0	1	1

   where $1^* = 1$ if $q_j > \underline{q}_{j,1} \equiv 3q_i - 1$ and $q_i < \frac{2}{3}$.

   - Belief update decision (b_i^*):

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

   where $1^+ = 1$ if $q_j > \underline{q}_{j,2} \equiv \frac{-q_i+3q_i^2}{2-6q_i+6q_i^2}$ and $q_i < \frac{2}{3}$. $1^{++} = 1$ if $q_j > \underline{q}_{j,3} \equiv \frac{q_i^2}{1-2q_i+2q_i^2}$.

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

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

   - 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 $q_{j,b}$.