Discussion Paper Series – CRC TR 224

Discussion Paper No. 219
Project B 04

Dynamic Competition for Attention

Jan Knoepfle [1]

October 2020

[1] University of Bonn

Funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through CRC TR 224 is gratefully acknowledged.

Dynamic Competition for Attention

Jan Knoepfle*

September 2020

Abstract

This paper studies information transmission in situations in which multiple senders compete for the attention of a decision maker. Senders are partially informed about a state and choose how to reveal information over time to attract maximal attention. A decision maker wants to learn about the state but faces attention costs. I characterise an equilibrium with simple strategies that lead to full information transmission in minimal time. The attention each sender receives is proportional to the residual value of her information. With endogenous information acquisition, increasing initial public information may decrease the aggregate information in society.
Keywords: Attention, Dynamic Information Provision, Media Competition
JEL Codes: D43, D83, L86

*University of Bonn. I am indebted to Marina Halac, Johannes Hörner, Stephan Lauermann and Sven Rady for invaluable feedback and support. For insightful conversations and suggestions, I thank Ian Ball, Dirk Bergemann, Francesc Dilmé, Ryota Iijima, Patrick Lahr, Niccolò Lomys, Aniko Öry, Anna Sanktjohanser, Eran Shmaya, Andre Speit, Philipp Strack, Allen Vong and Weijie Zhong. I am grateful to the Yale Economics Department for their hospitality during my visit in the 2018/2019 academic year, where this project was initiated. This work was supported by the German Research Foundation (DFG) through CRC TR 224 (Project B04).

1 Introduction

Information providers compete for attention. Most online content, such as news, professional product reviews, or weather forecasts, is offered free of charge, and the websites exploit the attracted attention to create revenue, primarily through advertisements. Without monetary prices, information providers compete through two key factors that determine their profits. First, they have to decide how much and what type of information to acquire. Several papers studying this question highlight the importance of attention as the currency in media markets.[1] Second, providers have to decide how to reveal their information over time. This aspect – how to optimally disseminate information when competing for attention – is the focus of this paper.

Attention is collected from a decision maker who is interested in the information held by the providers. As attention requires time and effort, the decision maker decides sequentially which providers to visit and when to stop, depending on the information previously observed. Recent work has studied the design of optimal dynamic information policies from the perspective of the decision maker.[2] Yet, in many situations, the power to design how information is revealed over time lies with providers.

How much information can be transmitted from the providers to the decision maker and what type of information processes arise when providers design offers to attract attention?

To answer these questions, I build a dynamic model in which information providers – the senders – compete for the attention of a decision maker – the receiver. The receiver has to take an action and wants to learn about an unknown state to maximise his utility. Senders are interested in maximising the number of visits and do not care about the receiver's action. At the beginning of the game, each sender is endowed with imperfect information over the state through a signal. Subsequently, there are multiple rounds of communication in which senders compete for a visit by the receiver. At the beginning of each round, senders offer experiments over their signals, that is, each sender commits to a distribution over messages, conditional on the realisation of her signal. Senders cannot commit across rounds. The receiver observes all offers. He either pays an attention cost to visit one sender and continue to the next round, or he stops learning and takes the optimal action with the current level of information. The model captures broad information and preference specifications with the condition that attention can be split finely enough and each sender's signal is informative enough so that it is worth at least one unit of attention, independent of the information previously delivered by her competitors.

The main result characterises an equilibrium in which all information is transmitted from the senders to the receiver. Each sender attracts attention proportional to the expected residual value of her information. This is a lower bound on attention for each sender. Therefore, this equilibrium is receiver-preferred, information is transmitted in minimal time. Offers made in equilibrium are of a simple class: each sender posts a probability with which the experi-

[1] See Galperti and Trevino (2018), Perego and Yuksel (2018), and Pant and Trombetta (2019).

[2] Most notably, Zhong (2019) characterises the optimal process designed by the decision maker with full flexibility. See discussion below.

ment reveals her initial signal fully. With the remaining probability, the experiment delivers no information. I refer to this class as *All-or-Nothing* (AoN) offers.

The market for information considered in this paper features intertemporal externalities. The information observed at any sender changes the receiver's valuation for future information as well as his probability assessment of other senders' signals. Furthermore, the design of an offer and its cost (in terms of attention) are closely intertwined. With externalities and in the absence of prices or general contracts, the existence of an efficient equilibrium is not a foregone conclusion.[3] To construct the equilibrium mentioned above, this paper introduces a substitute condition on the senders' signals that requires that any sender's information is more valuable when her competitors have revealed less.

To gain intuition on the equilibrium and the class of AoN offers, consider the case of a single sender. What is the maximal amount of attention a monopolist can extract from the receiver? The receiver is willing to pay a total attention cost equal to the difference in expected utility from taking the action with or without the sender's information. Due to the lack of intertemporal commitment, the sender cannot simply require the receiver to visit her for a fixed number of rounds and then reveal all her information at the last visit. In general, a non-committed monopolist cannot give out more information than necessary to make the receiver indifferent between spending another round of attention and taking the action at the current information. A simple way to keep the receiver indifferent is to make an AoN offer as introduced above. The sender chooses the AoN probability that all information is revealed as low as possible so that the receiver accepts. AoN offers imply that no information is revealed until a geometrically distributed arrival time, at which time all information is revealed.

When there are multiple senders, they design experiments facing Bertrand competition in every round. Each sender offers an experiment that makes her indifferent between being accepted and the lower bound of attention she can attract if she is not visited. This lower bound consists of waiting until all competitors have revealed their information and, subsequently, playing the strategy of the monopolist. At this point, the receiver's information includes all signals of her competitors, and the value of the lower bound depends on the realisations of these signals. In equilibrium, every sender offers the AoN probability such that the expected attention is equal to the current expectation of her lower bound. This expectation and the senders' offers change over time. Once only one sender is left, the receiver is indifferent between stopping and accepting this last sender's offer. As signals are substitutes, the receiver strictly prefers to accept an offer when there are still multiple senders whose information he has not yet observed. Given that senders require attention proportional to the residual value of their signal, concentrating a fixed amount of information on fewer senders hurts the receiver. While all information is still transmitted, the total required attention increases.

I provide examples of information and preference specifications that are captured by the model together with applications that the literature has studied with these specifications. Among

[3] As shown by the example in Section 6, externalities may impede information transmission entirely.

these is the application of the Gaussian-information, quadratic-loss specification to examine competition in news markets. For this setup, I extend the game to consider optimal information acquisition by two competing news outlets that face a tradeoff between checking further sources more carefully and breaking the news as early as possible. This *investigation race* always leads to specialisation into a less informed outlet that offers a more superficial report early and a more informed outlet that investigates as long as possible to deliver high precision. If news outlets have different efficiency levels ex ante, i.e. different rates at which their precision increases over time, the more efficient newspaper is the one that investigates longer, thereby exacerbating its initial advantage. Perhaps surprisingly, increasing the precision of initially available public information may decrease the final precision at which the action is taken. The adverse effect on the incentives to investigate can outweigh the direct increase in precision. If the government considers increasing information on an issue through a campaign and ignores the incentives of other providers informing on the same issue, such campaigns may have the opposite effect and decrease information to the public.

After discussing the related literature below, Section 2 presents the model. Section 3 sets the stage for the analysis, examining the value of information and introducing useful notation. The equilibrium characterisation in Section 4 starts with the monopoly benchmark before deriving the results for multiple senders. The news application is considered in Section 5. Section 6 discusses modelling choices and the relation to Zhong (2019) in more detail. Concluding remarks are presented in Section 7. Proofs not included in the main text can be found in the Appendix.

Related Literature. This paper contributes to the literature on optimal dynamic information acquisition by a decision maker, firstly, by endogenising the information processes chosen by senders, and secondly, by considering attention maximisation. The tractable dynamic model with multiple senders who are partially informed presents a technical contribution to the dynamic information design literature. With the application to news markets, the analysis sheds light on the tradeoff between publishing news earlier or gathering more precise information. The relation to these three strands of literature, among others, is discussed in detail below.

Optimal **dynamic information acquisition** by a decision maker has been introduced to the economics literature by Wald (1947), where the decision maker decides when to stop observing an exogenous information process and take an action. Several papers enrich the decision maker's problem by allowing him to adjust the information intensity or to choose among exogenous processes, see Moscarini and Smith (2001), Mayskaya (2017), Che and Mierendorff (2019), Liang and Mu (2020), Liang et al. (2019), and others.[4] The decision maker in my paper faces a related acquisition problem but chooses among experiments that are offered endogenously by the senders. More recently, Zhong (2019) characterises the optimal inform-

[4]See also Morris and Strack (2017) and Fudenberg et al. (2018) for recent developments on the Wald problem in different directions.

ation process designed by a decision maker with full flexibility facing a precision cost. He shows that the optimal policy consists of a Poisson process that leads to immediate action after arrival.[5] The current paper contributes to the literature on optimal dynamic design by considering a related question from the opposite perspective. Senders choose flexibly how to provide their information over time to maximise the attention they attract from the receiver. The offer strategies presented in the current paper[6] imply a geometrically distributed arrival of all information from one sender. This is akin to a Poisson process in continuous time with fully revealing news and where the absence of arrival allows no inference (there is no belief drift). Section 6.2 discusses the connection between Zhong (2019) and the single sender case in the current paper. I identify the lack of intertemporal commitment as an additional motive for Poisson processes.

My paper is related to the literature on **Bayesian persuasion**, based on Kamenica and Gentzkow (2011), in which a sender designs information to influence a receiver's behaviour. Senders in my model maximise attention and have no persuasion motive, that is, the action eventually taken by the receiver does not affect the senders' utilities. Most contributions model information design using belief-based techniques. I use an experiment-based approach. Section 6.1 discusses the benefits of doing so for a setting with multiple senders who design how to reveal partial information.

Dynamic information design has been studied in Au (2015), Che and Hörner (2017), Ely (2017), Renault et al. (2017), Smolin (2017), Board and Lu (2018), Ball (2019), Che et al. (2020), Ely and Szydlowski (2020), Guo and Shmaya (2019), Orlov et al. (2020), and others. The main contrasts to the current paper are the persuasion motive mentioned above and the focus on single-sender[7] environments. Ely and Szydlowski (2020) study the problem of a sender with intertemporal commitment who wants to persuade the receiver to execute an option as early as possible or as late as possible. The latter case may be interpreted as paying attention for as long as possible before stopping, which relates to the single sender case in the current paper. The optimal information processes in both papers share similar features: the receiver's belief is kept constant, and he is indifferent between stopping and continuing until the information is fully revealed and he stops. As in Au (2015) and Che et al. (2020), senders in my paper commit to the information offered within a period but cannot commit across periods. The senders' focus on the receiver's costly attention connects the current paper and Che et al. (2020), who examine optimal dynamic persuasion when the receiver has to pay an attention cost.

Section 5 considers a concrete specification with Gaussian information and quadratic loss for the receiver. This tractable setting is widely used in the literature on **media competition**.

[5] This gives a theoretical justification for the common use of Poisson processes to model information in dynamic environments, partly due to its tractability laid out in Keller et al. (2005).

[6] Which attain the unique equilibrium payoff in the monopoly case and the receiver-preferred equilibrium with competition.

[7] With the exception of Board and Lu (2018) who consider multiple sellers which are randomly matched with potential buyers in a search market and want to induce buyers to buy from them rather than continuing to search. For a cheap talk model that features multiple senders in a dynamic environment, see Margaria and Smolin (2018).

This application to news markets is related to Mullainathan and Shleifer (2005), Besley and Prat (2006), Gentzkow and Shapiro (2006), Galperti and Trevino (2018), Perego and Yuksel (2018), and Pant and Trombetta (2019). These papers highlight the importance of capturing an audience or maximising attention for media companies. They study aspects from information acquisition to optimal provision. Mullainathan and Shleifer (2005), Gentzkow and Shapiro (2006), and Pant and Trombetta (2019) consider optimal provision, which is the main focus of the current paper. Yet, within the more tractable Gaussian setting, I extend the analysis and consider information acquisition where precision levels arise from a market entry game. News outlets face a tradeoff between investigating longer or breaking a story before the competition. Galperti and Trevino (2018) and Perego and Yuksel (2018) study optimal static acquisition decisions determined by competition for attention. Senders in Galperti and Trevino (2018) choose the accuracy and the clarity of their news pieces at a cost, while receivers have a coordination motive. Senders in Perego and Yuksel (2018) choose to report on more polarised issues due to competition. The market entry game in the current paper focuses on accuracy levels only, and I find that the investigation race to publish first leads to polarisation in accuracy, where the less productive newspaper reports early and the more productive newspaper deepens her advantage by investigating as long as possible.

One interpretation of the setting in the current paper is to view senders as selling information to a receiver requiring a price in units of attention time. For a survey on **markets for information**, see Bergemann and Bonatti (2019). Bergemann et al. (2018) study the optimal design and pricing of a menu of experiments to screen the receiver according to his willingness to pay. In my paper, the receiver's willingness to pay is known. The monopolist can require attention proportional to the price charged by the monopolist in Bergemann et al. (2018), who would only offer the fully revealing experiment if he knew the receiver's willingness to pay. The dynamic information market studied in the current paper shares two important features with **dynamic price competition** for standard goods. First, the product 'information' has important externalities. Bergemann and Välimäki (2006) study repeated Bertrand price competition with externalities – the surplus of each purchase depends on the history of previous purchases. For the special case without inter-group externalities – where the surplus generated by a trade with seller i depends only on the number previous trades with i – they show that a marginal contribution equilibrium exists and leads to efficiency. Information generally features inter-group externalities. Nevertheless, the receiver-preferred equilibrium presented in the current paper is also constructed by considering each sender's expected marginal contribution. The second noteworthy feature of this market is the presence of capacity constraints: Each sender is initially endowed with information. This fixed endowment is a capacity constraint, relating the current paper to Dudey (1992), Martínez-de Albéniz and Talluri (2011), and Anton et al. (2014). As information is assumed to be always worth one visit in the current paper, the capacity constraint is binding. Paralleling results in the above papers, this implies that each sender can extract positive surplus despite the competition.

Gossner et al. (2019) study attention with a different focus. They show that drawing attention to one of several considered options unequivocally increases the likelihood of this option being chosen by an agent who uses a fixed threshold rule. This can be seen as an additional rationale to compete for attention.

2 Model

2.1 Environment

One receiver and a finite number of senders $i \in \{1, \cdots, I\}$ interact in discrete rounds of communication, $k = 0, 1, \ldots$. There is a state of the world ω from the Polish[8] space Ω that remains constant and is not observed by any player. However, each sender i is endowed with partial information over the state, represented by a signal x_i from the Polish space X_i. Let $X = \bigtimes_i X_i$. The state and signals are jointly distributed according to the commonly known prior $\tilde{\mu} \in \Delta(\Omega \times X)$.

The receiver has to take an irrevocable action a from the closed space A to maximise his utility $u(a, \omega)$, where $u : A \times \Omega \to \mathbb{R}$ is continuous and bounded. For this, the receiver relies on information from the senders. In each round, k, he can either pay the attention cost $c > 0$ and visit one sender, or take an action with the information gathered so far. After the action is taken, the game ends.

Senders offer *experiments* over their own signal. To avoid signalling, I assume that senders do not observe their signal prior to revealing it through experiments. An experiment is a conditional distribution over messages m from the Polish space M. The message space M is equal for all senders and rich enough to contain all information about $\boldsymbol{x} = (x_1, \cdots, x_I)$, i.e. $X \subset M$. At the beginning of round k, each sender i simultaneously announces $\lambda_{i,k} : X_i \times \mathcal{B}(M) \to [0, 1]$, a (regular) conditional probability such that $\lambda_{i,k}(\cdot, W)$ is measurable for all $W \in \mathcal{B}(M)$, and $\lambda_{i,k}(x_i, \cdot)$ is a probability measure given any signal $x_i \in X_i$. The set of possible experiments for sender i is denoted by Λ_i. Senders compete for attention. In each round that sender i is visited, she receives utility normalised to one.

2.2 Strategies, Payoffs, Equilibrium

First, nature draws the state ω and the signals $\boldsymbol{x} = (x_1, \cdots, x_I)$. At the beginning of each round $k \geq 0$, if the receiver has not taken an action previously, all senders simultaneously offer experiments $\lambda_{i,k}$.

The receiver observes the offers and chooses $d_k \in \{0, 1, \cdots, I\}$, where $d_k = 0$ encodes that he stops and $d_k = i \in \{1, \cdots, I\}$ means that he pays cost $c > 0$ and visits sender i. When the receiver stops, he takes an action $a \in A$, the game ends, and payoffs realise.[9] Visiting sender

[8] A Polish space is a separable and completely metrisable space. This ensures the existence of the conditional probability measures used below.

[9] For completeness, assume that $d_k = 0 \Rightarrow d_{k+1} = 0$.

i implies that he observes $m_{i,k} \in M$ drawn from the distribution $\lambda_{i,k}$ and the game continues to the next round.

A public **history** of the game is

$$h^k = \left(\left((\lambda_{i,1})_{i=1}^I, d_1, m_{d_1,1} \right), \cdots, \left((\lambda_{i,k})_{i=1}^I, d_k, m_{d_k,k} \right) \right) \qquad \text{for } k \geq 0,$$

with initial history $h^{-1} = \emptyset$. All players observe all past offers, the receiver's choices and the message of the chosen sender. Hence, all senders observe the information revealed by their competitors.[10] Denote by $\mathcal{H}^k$ the set of round-k histories.

A pure **strategy** for the receiver is a collection of maps $\left(\sigma_k^R\right)_{k \geq 0}$ with

$$\sigma_k^R : \mathcal{H}^{k-1} \times \left(\times_i \Lambda_i \right) \to \{0, 1, \cdots, I\}.$$

Likewise, for each sender $i \in \{1, \cdots, I\}$, a pure strategy is a collection $\left(\sigma_k^i\right)_{k \geq 0}$ with

$$\sigma_k^i : \mathcal{H}^{k-1} \to \Lambda_i.$$

The receiver's final **payoff** has two components. First, he gets utility $u(a, \omega)$ when stopping with action a if the state is ω. Second, there are attention costs that depend on how many times the receiver visits a sender before taking action. Each visit costs $c > 0$, so that the receiver's final payoff will be

$$u(a, \omega) - c \cdot \sum_{k \geq 0} \mathbb{1}_{\{d_k \neq 0\}}.$$

Senders maximise the attention they attract. Each visit gives utility normalised to 1. Sender i's final payoff is then

$$\sum_{k \geq 0} \mathbb{1}_{\{d_k = i\}}.$$

The **solution concept** is a Perfect Bayesian Equilibrium, with the additional requirement that beliefs are only updated with Bayes' rule according to the chosen experiment. This requirement ensures the 'no-signalling-what-you-don't-know' property (see Fudenberg and Tirole, 1991), whereby offers do not reveal information the senders do not hold. It also implies that experiments from off-path offers would be interpreted correctly if they were accepted.

[10] It is plausible that the senders can also visit their competitors, given that experiments are offered publicly. With Gaussian information considered in Section 5, I show how this assumption can be relaxed.

2.3 Examples

Before moving to the analysis, I give two specific setups encompassed by the model presented above. Applications for which these setups or a close variant have been used in the literature are mentioned in square brackets. The News Markets application is considered more carefully in Section 5.

Example 1: Gaussian Information and Quadratic Loss. [Global games: Morris and Shin (2002), Bergemann and Morris (2013), Angeletos and Pavan (2007). Social learning: Vives (1996). News markets: Chen and Suen (2019), Galperti and Trevino (2018)]

The receiver wants to learn the state of the world $\omega \sim \mathcal{N}(0, 1/p_0)$ as precisely as possible. His utility from the action is $u(a, \omega) = -(\omega - a)^2$, so that the expected stopping utility is minus the conditional variance $Var(\omega|\xi)$ given the current information.[11] Each sender $i \in \{1, \cdots, I\}$ is endowed with a conditionally independent signal $x_i \sim \mathcal{N}(\omega, 1/p_i)$, where $p_i > 0$ is sender i's *precision level*.

Example 2: Additive Attributes. [Consumer search: Wolinsky (1986), Choi et al. (2018), Ke and Lin (2020). Advertising: Anderson and Renault (2009), Sun (2011)]

Let the receiver be a consumer who considers buying one of two objects, $A = \{1, 2\}$. The (net) utility of the two objects is $\omega = (\omega_1, \omega_2) \in \mathbb{R}^2$, where each object's utility is determined by a common and an idiosyncratic attribute as follows:

$$\omega_i = Y + \gamma_i \qquad\qquad \text{for } i \in \{1, 2\},$$

where Y is the common component distributed according to F on $[\underline{Y}, \bar{Y}]$ and γ_i are distributed independently according to G_i on $[\underline{\gamma}, \bar{\gamma}]$. Each sender holds information about one of the options in the form of a noisy signal of the total utility but is unable to distinguish between the common and the idiosyncratic component: $x_1 = \omega_1 + \epsilon_1$ and $x_2 = \omega_2 + \epsilon_2$ with $\epsilon_i \overset{\text{iid}}{\sim} \mathcal{N}(0, 1)$. Senders in my model are indifferent about the receiver's actions. This is the case if the sender does not sell the product by herself, as with car or technology magazines on- and offline.

3 Updating and the Value of Information

This section introduces notation that will be used extensively in the remainder of the analysis. To determine the optimal action and compute the expected utility from stopping, the receiver has to form a belief about ω. The messages deliver information about signals $\boldsymbol{x} = (x_1, \cdots, x_I)$. Recall that the joint distribution over the state and signals is $\tilde{\mu} \in \Delta(\Omega \times X)$. I denote its marginal

[11] The fact that u is not bounded from below does not create problems here since the stopping utility at the prior is equal to $-\frac{1}{p_0} > -\infty$.

with respect to $\boldsymbol{x}$, that is, the unconditional prior distribution of signals, by $\xi^0 \in \Delta(X)$. It will be convenient to work with the posterior signal-belief ξ in the rest of the paper.

The receiver's *stopping utility* with belief ξ, i.e. the expected utility from the optimal action, given that he currently holds posterior ξ, is

$$U(\xi) \equiv \max_{a \in A} \mathbb{E}_{\omega \sim \mu(\xi)} [\ u(a, \omega)\ ],$$

where $\mu(\xi)$ denotes the belief about the state given that the belief about the signals is ξ. The belief μ is not used in the analysis. The formula is given in Appendix A together with further details on this section.

At the end of round k, the belief ξ^k is updated to ξ^{k+1} by Bayes' rule after observing message m_k resulting from the selected experiment λ_k. Denote the *updating rule* by ξ' such that

$$\xi^{k+1} = \xi'(\xi^k, m).$$

Note that the notation suppresses the chosen experiment in the updating rule. Appendix *A* contains the updating rule in detail and shows that it is well defined. I denote by $\xi'(\xi^0, \boldsymbol{x}_{-i})$ the belief that results if the signals of all senders different from i are known and nothing has been revealed about x_i.

With this, define the *value* of offer λ_i at current belief ξ as

$$v(\lambda_i|\xi) \equiv \mathbb{E}_{x_i \sim \xi} \left[\mathbb{E}_{m \sim \lambda_i(x_i, \cdot)} \left[U\left(\xi'(\xi, m) \right) \right] \right] - U(\xi).$$

The value is defined as the expected difference between the stopping utilities with and without the additional information from λ_i. Note that $v \geq 0$ always.

For the special case in which sender i's experiment reveals her exact signal, let

$$\bar{v}_i(\xi) \equiv v\left(\delta_{\{x_i\}}|\xi\right)$$

denote the value of all her information given belief ξ. Here, the experiment that reveals i's signal precisely is denoted by the Dirac measure $\lambda_i(x_i, \cdot) = \delta_{\{x_i\}}(\cdot)$.

I assume that attention can be split finely enough to make every sender's information worth one unit of attention, independent of the realisation of her opponents' signals.

Assumption 1. *For all $i \in \{1, \cdots, I\}$ and for all $\boldsymbol{x}_{-i}$ in the support of $\xi^0(X_i, \cdot)$:*

$$\bar{v}_i\left(\xi'(\xi^0, \boldsymbol{x}_{-i})\right) > c. \tag{A1}$$

This condition ensures that for any realisation of the other senders' signals, even if the receiver knows these exactly, sender i still has enough information to attract at least one visit. In particular, condition (A1) implies that no sender has perfect information. Given the Bertrand competition, if at least two senders had perfect information, all senders would offer all information in the first round, and the receiver would become perfectly informed after one visit.[12]

4 Equilibrium

This section identifies a simple class of information-transmission processes that is sufficient to achieve the unique equilibrium payoffs in the case of a single sender and the receiver-preferred equilibrium payoff with multiple senders, in which all information is transmitted in the shortest amount of time possible.

4.1 Single Sender

Consider the case of a single sender, $I = 1$. What is the maximal expected attention cost the receiver is willing to pay for the sender's information? It is equal to the difference between the stopping utility with no information and the expected stopping utility with all information. This is precisely $\bar{v}_1(\xi^0)$. As each visit requires a cost of c, the maximal expected number of visits the sender can attract is

$$\frac{\bar{v}_1\left(\xi^0\right)}{c}.$$

If the sender could commit across rounds, the simplest strategy to implement this outcome would require $\frac{\bar{v}_1(\xi^0)}{c} - 1$ visits from the receiver at which no information is revealed, and then all information would be revealed at the last visit.[13] However, the sender lacks the intertemporal commitment to credibly promise all information in the last round. She may, for example, repeat the round-0 strategy.

A simple sender strategy to overcome the non-commitment issue and to deal with potential integer problems is to offer revealing x_1 with probability $\lambda^* \in [0, 1]$ and revealing no information with probability $1 - \lambda^*$.[14] Offers in this class are denoted as *All-or-Nothing* (AoN) offers. To give a simple example of an AoN offer, assume that the sender's signal is the result of a coin flip, $x_1 \in \{0, 1\}$, with $\xi^0 = 0.6$ prior probability that $x_1 = 1$. Let the receiver's utility be 1 if he guesses the signal correctly and 0 otherwise. That is, $\omega = x_1 \in \{0, 1\}$, $A = \{0, 1\}$, and $u(a, \omega) = \mathbb{1}_{\{a=\omega\}}$.[15]

[12] See Section 5 for the equivalent of (A1) for Example 1. For Example 2, one sufficient condition for (A1) is that the noise term of each sender has sufficient variance.

[13] This intuitive argument neglects non-divisibilities that make this potential strategy suboptimal as it could only achieve integer amounts of visits.

[14] I abuse the notation by letting λ^* denote a probability in [0,1], while λ generally denotes distributions over messages. Formally, the AoN experiment is represented by the conditional distribution $\lambda(x_1, \cdot) = \lambda^* \delta_{\{x_1\}} + (1 - \lambda^*)\delta_{\{m\}}$ for an arbitrary message $m \in M \backslash X_i$ that conveys no information.

[15] Here, state and signal are identical. For this section, the general model could equivalently be specified with $\omega = \boldsymbol{x}$. However, in Section 5, signals are chosen endogenously by the senders, so that modelling the payoff-relevant state separately allows keeping the endogenous signals and the exogenous state distribution apart.

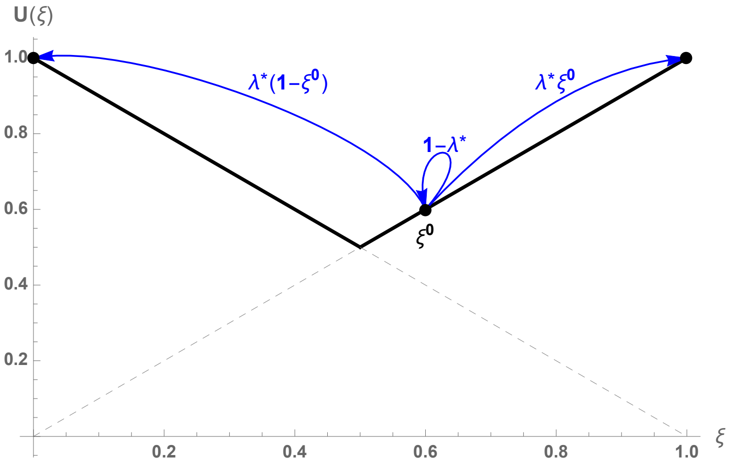


Figure 1: AoN experiment with binary signal

Figure 1 shows the receiver's expected utility as a function of the belief as a solid black line. The three arrows indicate the possible jumps in belief induced by the AoN experiment. As the middle arrow shows, with probability $1 - \lambda^*$, the experiment reveals no information and the belief remains unchanged. With probability λ^*, the sender's signal is revealed perfectly so that with probability $\lambda^*(1 - \xi^0)$, the belief jumps left to 0, and with probability $\lambda^* \xi^0$, the belief jumps right to 1.

The following result shows that an equilibrium in AoN strategies generally exists in the monopoly game. Equilibrium payoffs are unique.

Lemma 1. *Let* $I = 1$. *There is an AoN equilibrium in which, in each round, the sender offers AoN probability*

$$\lambda_1^*(\xi^0) = \frac{c}{\bar{v}_1(\xi^0)}.$$

The receiver accepts every round until x_1 *is revealed. In any equilibrium of the monopoly game the expected payoffs are* $\frac{\bar{v}_1(\xi^0)}{c}$ *for the sender and* $U\left(\xi^0\right)$ *for the receiver.*

Note that assumption (A1) ensures that $\lambda_1^*(\xi^0) < 1$. If the sender's strategy prescribes AoN offers until all information is transmitted, the receiver's continuation payoff in the event of no revelation (which happens with probability $1 - \lambda_1^*(\xi^0)$) remains at his initial payoff. Hence, the AoN probability $\lambda_1^*(\xi^0)$ that makes the receiver indifferent between taking action immediately and accepting the offer, has to satisfy

$$U\left(\xi^0\right) = -c + \lambda_1^*(\xi^0)\, \mathbb{E}_{x_1 \sim \xi^0}\left[U\left(\xi'(\xi^0, x_1)\right)\right] + (1 - \lambda_1^*(\xi^0))U\left(\xi^0\right).$$

Accepting the offer creates attention cost c. With probability $\lambda_1^*(\xi^0)$, the receiver learns x_1 and stops with utility $U(\xi'(\xi^0, x_1)))$. With probability $1 - \lambda_1^*(\xi^0)$, the receiver learns no information, which gives utility $U(\xi^0)$ as the sender will keep him indifferent in the following round again.

The offer $\lambda_1^*(\xi^0)$ is accepted by the receiver in every round until the information is eventually revealed. The number of rounds until revelation follows a geometric distribution with parameter $\lambda_1^*(\xi^0)$, so that the expected number of rounds is $\frac{1}{\lambda_1^*(\xi^0)}$. As the receiver is indifferent between accepting and stopping in every round, it should not be a surprise that solving the above indifference condition for λ^* gives

$$\lambda_1^*(\xi^0) = \frac{c}{\bar{v}_1(\xi^0)}.$$

The expected attention is precisely the upper bound the receiver is willing to spend.

Depending on the information structure, there may be other strategies that resolve the sender's non-commitment and attract the maximal amount of attention.[16] The attractiveness of the AoN strategy lies in the fact that it works for general information structures and in its simplicity. It leads to a stationary information-arrival process. Furthermore, the lack of intertemporal commitment requires that any experiment delivers, in expectation, a strictly positive increase in the stopping utility. In Section 6.2, I discuss in detail that this strict increase requires beliefs to jump with positive probability whenever the action set is finite and how – even in the limit as the period length shrinks – the information process necessarily features a Poisson-jump component. Thereby, this identifies an additional driver of Poisson information, stemming from the lack of commitment, rather than risk preferences induced by discounting (see Zhong, 2019). The above results are robust to discounting; the analysis remains almost unchanged when the receiver and the sender share a common discount factor.

4.2 Multiple Senders

In the general case with $I \geq 2$ senders, equilibrium payoffs are no longer unique. The subsequent analysis focuses on receiver-preferred equilibria. This selection best captures the tradeoff between the amount and the speed of information transmission since information has no instrumental value for the senders. The welfare-maximising equilibrium crucially depends on the normalisation of the value the senders derive from each visit. In particular, if $c = 1$, maximising welfare is equivalent to maximising the amount of information transmitted as visits from the receiver to any sender have no impact on welfare. Whenever $c \geq 1$, the receiver-preferred equilibrium is also welfare-maximising.

With multiple senders, there are informational externalities that may impede information transmission. For illustration, consider the following example with the detailed argument presented in Section 6.3. Suppose there are two senders. Each sender's signal is an independent, fair coin flip. The receiver has to guess whether the two coins match or not. For

[16] The sender could reveal some information every round, successively increasing the receiver's stopping utility to commit herself to offer even more information in the following round. Appendix A.3 includes an example of such a process when information is normally distributed. The special feature of the Gaussian distributions allows the sender to achieve information transmission in a deterministic number of visits (modulo integer problems).

this decision problem, the signals form complements. Each signal is valuable only in conjunction with the other. Since senders cannot commit across rounds, complements cause a hold-up problem: after one sender has revealed her information, the following sender would require maximal attention, keeping the receiver at the current stopping utility with one signal only. Anticipating this, the receiver is not willing to spend any attention for the first signal given that it delivers no value on its own.

To rule out this class of problems and ensure information transmission, I introduce the following condition.

Definition 1. *The senders' signals are substitutes if, for all i and for all beliefs ξ with $supp(\xi) \subseteq supp(\xi^0)$:*

$$\bar{v}_i(\xi) \geq \mathbb{E}_{\boldsymbol{x}_{-i}\sim\xi}\left[\bar{v}_i\left(\xi'(\xi, \boldsymbol{x}_{-i})\right)\right]. \tag{SU}$$

Signals are substitutes if the current value of x_i at belief ξ is greater than the expected value after knowing all other senders' signals.[17] That sender i's information is more valuable the less is known from her competitors is consistent with many applications. This is especially the case when senders report on a single issue or, as in Example 2, when the signals allow inference about a common component that affects all options. In both examples above, signals are substitutes.

In equilibrium, competing senders make offers that make them indifferent between being accepted or rejected. Constructing an AoN equilibrium requires determining the maximal AoN probability a competing sender is willing to offer. Consider the situation in which all senders but i have revealed their signals, and sender i has revealed no information at all. That is, $\boldsymbol{x}_{-i}$ is known and the belief is $\xi'(\xi^0, \boldsymbol{x}_{-i})$. Sender i can extract maximal attention from the receiver. The receiver is willing to visit her

$$\frac{\bar{v}_i\left(\xi'(\xi^0, \boldsymbol{x}_{-i})\right)}{c}$$

rounds to learn x_i.

As will be shown below, a lower bound on the attention sender i can extract is given by waiting until all competitors have revealed their information and offering AoN probability

$$\lambda_i^*\left(\xi'(\xi^0, \boldsymbol{x}_{-i})\right) = \frac{c}{\bar{v}_i\left(\xi'(\xi^0, \boldsymbol{x}_{-i})\right)}.$$

[17] Börgers et al. (2013) introduce notions of substitutes and complements for a pair of signals. Viewing x_i and $\boldsymbol{x}_{-i}$ as two signals, (SU) corresponds to the notion of substitutability in Börgers et al. (2013) for given ξ and restricted to the specific decision problem considered here. Their requirement is independent of the decision problem and therefore stronger.

However, this value depends on the realisations of $\boldsymbol{x}_{-i}$, so that its expectation – taken over all competitors' signals given the current information – changes over time. Suppose the AoN probability offered by each sender is such that her expected payoff is precisely the expectation of the outside option mentioned above, assuming that this offer was repeatedly accepted until revelation. The following result shows that, if signals are substitutes, these strategies form an equilibrium. In addition, this equilibrium attains full information transmission in the shortest possible time among all equilibria, making it receiver-preferred.

Theorem 1. *If senders' signals are substitutes, there is an equilibrium with the following strategies. At belief ξ, senders whose information has not been revealed make AoN offers with probability*

$$\lambda_i^*(\xi) = \frac{c}{\mathbb{E}_{\boldsymbol{x}_{-i}\sim\xi}\left[\bar{v}_i\left(\xi'(\xi^0, \boldsymbol{x}_{-i})\right)\right]}. \tag{1}$$

The receiver is indifferent between visiting any of the senders whose information has not been revealed and visits them in arbitrary order until all information is transmitted. This equilibrium is receiver-preferred.

Proof. Note that the result characterises a class of equilibria rather than a single equilibrium as the receiver's behaviour is not fixed. By (A1), we have that $\lambda_i^*(\xi) < 1$ for all i whose information has not been revealed. The proof is organised in three claims:

Claim 1. *Fix any strategies by senders $\neq i$ and assume the receiver is playing a best response. Let the current belief be ξ and assume sender i has not revealed any information. Then, playing the AoN strategy from the theorem secures sender i an expected payoff of $\frac{1}{\lambda_i^*(\xi)}$.*

Proof of Claim 1. First, we show that the AoN strategy ensures that the receiver will not stop without observing sender i's information. Formally, for all beliefs ξ,

$$-c + \lambda_i^*(\xi) E_{x_i\sim\xi}\left[V_R\left(\xi'(\xi, x_i)\right)\right] + (1 - \lambda_i^*(\xi)) V_R(\xi) \geq U(\xi).$$

Here, V_R denotes the receiver's continuation value (suppressing history and strategy). Clearly, $V_R(\xi) \geq U(\xi)$, as the receiver always has the option to stop. For the above inequality to hold, substituting and rearranging gives that it is sufficient to show that

$$E_{x_i\sim\xi}\left[U\left(\xi'(\xi, x_i)\right)\right] - U(\xi) \geq \frac{c}{\lambda_i^*(\xi)}.$$

The left-hand side of this inequality is the definition of $\bar{v}_i(\xi)$. Replacing λ_i^* on the right-hand side with (1) shows that this inequality is equivalent to the definition of substitutes in (SU).

Second, we show that the expected payoff for sender i from using the AoN strategy is exactly $\frac{1}{\lambda_i^*(\xi)}$. To illustrate this concisely, the remainder of the argument for Claim 1 considers Markov strategies, so that the belief ξ determines the senders' payoffs. This restriction is not necessary for the result and a detailed argument without it is included in Appendix A. Observe that i's valuation satisfies:

$$V_i(\xi) = \begin{cases} 1 + \lambda_i^*(\xi)0 + (1 - \lambda_i^*(\xi))V_i(\xi) & \text{if } i \text{ is chosen} \\ 0 + \mathbb{E}_{x_j \sim \xi}\left[\mathbb{E}_{m_j \sim \lambda_j(x_j,\cdot)}\left[V_i(\xi'(\xi, m_j))\right]\right] & \text{if } j \neq i \text{ is chosen.} \end{cases}$$

In the first line, i is visited and her continuation value is 0 if her information is revealed and remains unchanged if no information is given out. The value $\frac{1}{\lambda_i^*(\xi)}$ follows immediately from re-arranging. In the second line, depending on the realisation of m_j and the receiver's choice in the following round, the value $V_i(\xi'(\xi, m_j))$ is either $\frac{1}{\lambda_i^*(\xi'(\xi,m_j))}$ if sender i is chosen in that round, or

$$V_i(\xi'(\xi, m_j)) = \mathbb{E}_{x_\ell \sim \xi'(\xi,m_j)}\left[\mathbb{E}_{m_\ell \sim \lambda_\ell(x_\ell,\cdot)}\left[V_i(\xi'(\xi'(\xi, m_j), m_\ell))\right]\right],$$

if a sender $\ell \neq i$ is chosen. As $c > 0$, there can be at most finitely many rounds and realisations before sender i is chosen so that, eventually, we arrive at realisations with belief $\hat{\xi}$ and $V_i(\hat{\xi}) = \frac{1}{\lambda_i^*(\hat{\xi})}$.

Since, by definition,

$$\frac{1}{\lambda_i^*(\hat{\xi})} = \mathbb{E}_{\boldsymbol{x}_{-i} \sim \hat{\xi}}\left[\bar{v}_i\left(\xi'(\xi^0, \boldsymbol{x}_{-i})\right)\right]\frac{1}{c},$$

we have that

$$\begin{aligned} &\mathbb{E}_{x_j \sim \xi}\left[\mathbb{E}_{m_j \sim \lambda_j(x_j,\cdot)}\left[\frac{1}{\lambda_i^*(\xi'(\xi, m_j))}\right]\right] \\ &= \mathbb{E}_{x_j \sim \xi}\left[\mathbb{E}_{m_j \sim \lambda_j(x_j,\cdot)}\left[\mathbb{E}_{\boldsymbol{x}_{-i} \sim \xi'(\xi,m_j)}\left[\bar{v}_i\left(\xi'(\xi^0, \boldsymbol{x}_{-i})\right)\right]\frac{1}{c}\right]\right] \\ &= \mathbb{E}_{\boldsymbol{x}_{-i} \sim \xi}\left[\bar{v}_i\left(\xi'(\xi^0, \boldsymbol{x}_{-i})\right)\frac{1}{c}\right] = \frac{1}{\lambda_i^*(\xi)}. \end{aligned}$$

Therefore, by taking expectations as many times as necessary from the last realisation to the current stage with belief ξ, we get the claimed payoff.

Claim 2. *Let all senders play the AoN strategies from the theorem and assume the receiver is playing a best response. Then, no sender has a profitable deviation.*

Proof of Claim 2. The receiver visits all senders on the equilibrium path by Claim 1. Suppose the on-path belief is ξ and that only information from senders $1, \cdots, j-1$ has been observed. Then, the equilibrium continuation utility of the receiver can be expressed as

$$V_R(\xi) = \mathbb{E}_{\boldsymbol{x} \sim \xi}\left[U\left(\xi'(\xi^0, \boldsymbol{x})\right)\right] - c\sum_{i=j}^{I} V_i(\xi), \tag{2}$$

where the proof of the last claim showed that $V_i(\xi) = \frac{1}{\lambda_i^*(\xi)}$ for all i.

Suppose now that one sender i' deviates to an offer that gives an expected payoff higher than $\frac{1}{\lambda^*_{i'}(\xi)}$ if accepted. By visiting the remaining, non-deviating senders, the receiver achieves an expected payoff of

$$\mathbb{E}_{\boldsymbol{x}_{-i'}\sim\xi}\left[U\left(\xi'(\xi^0,\boldsymbol{x}_{-i'})\right)\right]-c\sum_{\substack{i=j\\ i\neq i'}}^{I}\frac{1}{\lambda^*_i(\xi)}. \tag{3}$$

The λ^*_i are chosen such that, at the last sender, the receiver is indifferent between stopping without her information or paying the corresponding attention cost to obtain her information. Hence, the values (2) and (3) are equal. Even if the alternative strategy of sender i' would lead to all her information being revealed, the receiver still prefers to reject any offer yielding i' an expected payoff higher than $\frac{1}{\lambda^*_{i'}(\xi)}$.

Claim 3. *The AoN equilibrium achieves the maximal payoff for the receiver among all equilibria.*

Proof of Claim 3. The action is always taken with all information. As the value of information is always positive and, by claim 1, no sender can receive less attention in expectation, the AoN equilibrium is receiver-preferred.

□

Theorem 1 yields a simple computation of the equilibrium payoffs of the receiver and the senders. It suffices to compute the expected residual value of each sender's signal. After the following remark, the next subsection makes use of this to show that, if signals are substitutes, concentrating information on fewer senders slows down transmission.

Remark 1. *Note that the on-path strategies in this equilibrium are Markov. Senders' actions are fully determined by the state, ξ. The receiver's actions are fully determined by the state and the offers made in the current round. The state ξ is not strictly payoff-relevant in that two distinct posterior beliefs over signals may lead to the same belief μ over states and, therefore, to the same stopping utility. One might argue that μ is a more appropriate state variable. However, it is easy to verify that the belief μ captures too little information to determine optimal behaviour in the continuation game: consider a case with two senders who have symmetrically distributed signals. One of the two has revealed this signal to the receiver but the other has not. The same belief μ may be derived from sender 1's signal or sender 2's signal being known, but the value of future offers from senders 1 and 2 depends crucially on this distinction.*

4.3 Concentration of Information

As the action is taken after all information is transmitted, the receiver's expected total payoff is

$$\mathbb{E}_{\boldsymbol{x}\sim\xi^0}\left[U\left(\xi'(\xi^0,\boldsymbol{x})\right)\right] - c\sum_{i=1}^{I}\frac{1}{\lambda_i^*(\xi^0)}.$$

Each sender's attention is proportional to the residual value of her information so that the receiver's total payoff is equal to

$$\mathbb{E}_{\boldsymbol{x}\sim\xi^0}\left[U\left(\xi'(\xi^0,\boldsymbol{x})\right)\right] - \sum_{i=1}^{I}\mathbb{E}_{\boldsymbol{x}_{-i}\sim\xi^0}\left[\mathbb{E}_{x_i\sim\xi'(\xi^0,\boldsymbol{x}_{-i})}\left[U\left(\xi'(\xi^0,\boldsymbol{x})\right)\right] - U\left(\xi'(\xi^0,\boldsymbol{x}_{-i})\right)\right].$$

Consider increasing the concentration of information by merging the signals of senders $i = 1$ and $i = 2$ into a single signal $x_{1,2} = (x_1, x_2)$ held by sender 2. Sender 1 has no information and is excluded from the game.

This decreases the speed of information transmission. To see this, consider the receiver's utility after the concentration. The first term remains the same as the overall information has not changed. The attention required by senders $i \in \{3, \cdots, I\}$ also remains unaffected. The change comes from the attention required for signal $x_{1,2}$, which is now equal to

$$\mathbb{E}_{\boldsymbol{x}\sim\xi^0}\left[U\left(\xi'(\xi^0,\boldsymbol{x})\right)\right] - \mathbb{E}_{\boldsymbol{x}_{-\{1,2\}}\sim\xi^0}\left[U\left(\xi'(\xi^0,\boldsymbol{x}_{-\{1,2\}})\right)\right]. \tag{4}$$

Before the concentration, observing x_1 and x_2 required the attention of

$$\begin{aligned}&\mathbb{E}_{\boldsymbol{x}\sim\xi^0}\left[U\left(\xi'(\xi^0,\boldsymbol{x})\right)\right] - \mathbb{E}_{\boldsymbol{x}_{-1}\sim\xi^0}\left[U\left(\xi'(\xi^0,\boldsymbol{x}_{-1})\right)\right] + \mathbb{E}_{\boldsymbol{x}\sim\xi^0}\left[U\left(\xi'(\xi^0,\boldsymbol{x})\right)\right]\\ &- \mathbb{E}_{\boldsymbol{x}_{-2}\sim\xi^0}\left[U\left(\xi'(\xi^0,\boldsymbol{x}_{-2})\right)\right].\end{aligned} \tag{5}$$

To see that the cost after concentrating the information in (4) is greater than the cost before in (5), consider the difference:

$$\begin{aligned}\mathbb{E}_{\boldsymbol{x}_{-1}\sim\xi^0}\left[U\left(\xi'(\xi^0,\boldsymbol{x}_{-1})\right)\right] + \mathbb{E}_{\boldsymbol{x}_{-2}\sim\xi^0}\left[U\left(\xi'(\xi^0,\boldsymbol{x}_{-2})\right)\right]\\ -\mathbb{E}_{\boldsymbol{x}_{-\{1,2\}}\sim\xi^0}\left[U\left(\xi'(\xi^0,\boldsymbol{x}_{-\{1,2\}})\right)\right] - \mathbb{E}_{\boldsymbol{x}\sim\xi^0}\left[U\left(\xi'(\xi^0,\boldsymbol{x})\right)\right].\end{aligned}$$

This can be rearranged to

$$\mathbb{E}_{\boldsymbol{x}_{-\{1,2\}}\sim\xi^0}\left[\bar{v}_1(\xi'(\xi^0,\boldsymbol{x}_{-\{1,2\}}))\right] - \mathbb{E}_{\boldsymbol{x}_{-1}\sim\xi^0}\left[\bar{v}_1(\xi'(\xi^0,\boldsymbol{x}_{-1}))\right],$$

which is positive since signals are substitutes. Hence, concentrating the same amount of information on fewer senders slows down information transmission and hurts the receiver.

5 News Markets

This section applies the main results to the specification in Example 1 to derive and interpret further comparative statics and extend the model by endogenous information acquisition. Variants of this Gaussian setting have been applied to study various aspects of media markets in Galperti and Trevino (2018), Chen and Suen (2019), and others.

The receiver wants to be informed about the state of the world $\omega \sim \mathcal{N}(0, 1/p_0)$. He wants to match the state with his action a and gets utility $u(a, \omega) = -(a-\omega)^2$. Each newspaper i holds some information about the state represented by a signal that is independent conditional on the state: $x_i \sim \mathcal{N}(\omega, 1/p_i)$ where $p_i > 0$ is called i's *precision level*. The receiver's optimal action at belief ξ is $a^*(\xi) = \mathbb{E}_{\omega\sim\mu(\xi)}[\omega]$, and the expected utility from stopping with belief ξ is $-\mathbb{E}_{\omega\sim\mu(\xi)}\left[(a^*(\xi)-\omega)^2\right] = -Var(\omega \mid \xi)$. Hence, the receiver's stopping utility at prior information is $-\frac{1}{p_0}$. The reduction in variance caused by any sender's signal is independent of the realisation of her own or her opponents' signal. In particular, if the receiver knows the signals of all senders, his stopping utility is $-\frac{1}{p_0+p_1+\ldots+p_I}$. Precision increases linearly.

5.1 Exogenous Precision

To characterise the AoN equilibrium analogous to Theorem 1, define by

$$P \equiv p_0 + \sum_{i=1}^{I} p_i,$$

the precision level of all senders plus p_0, the precision of the state distribution. Let $P_{-i} = P - p_i$ denote the total precision without sender i. In this case, assumption (A1) boils down to the requirement that, for all i:

$$\frac{1}{P_{-i}} - \frac{1}{P_{-i}+p_i} > c,$$

so that the residual value of sender i's information exceeds the attention cost c.

Corollary 1. *There is an AoN equilibrium analogous to Theorem 1 in which each sender i offers AoN probability*

$$\lambda_i^* = c\frac{P_{-i}(P_{-i}+p_i)}{p_i}$$

in every round until her signal is revealed.

In this AoN equilibrium, sender i expects to attract the total attention of

$$\frac{1}{\lambda_i^*} = \frac{p_i}{cP_{-i}(P_{-i}+p_i)}.$$

With higher precision, she can attract more attention. A higher precision of her competitors' signals or of the initial distribution both lead to a higher P_{-i} and decrease sender i's expected attention.

The receiver's final payoff is

$$-\frac{1}{P} - c\sum_{i=1}^{I} \frac{p_i}{cP_{-i}(P_{-i} + p_i)}.$$

Hence, fixing the total precision level P, the reader is better off, the smaller is

$$\sum_{i=1}^{I} \frac{p_i}{P - p_i} = \sum_{i=1}^{I} \left(\frac{P}{P - p_i} - 1 \right).$$

The highest utility the receiver can get is trivially achieved at maximal prior precision with $p_0 = P$ and $p_i = 0$ for all $i \geq 1$. If we fix P and p_0, does the receiver prefer the remaining precision to be distributed evenly among all newspapers or to be skewed with some papers holding a lot and others holding very little information? The fraction on the right side of the equality is convex in p_i. The receiver prefers a uniform distribution of precision levels over senders, that is, $p_i = \frac{P-p_0}{I}$ for all $i \in \{1, \cdots, I\}$.

5.2 Information Acquisition

In practice, the precision of a newspaper's information is endogenously determined by its investigation process and editorial policies. One crucial factor that affects precision is the time at which a story is reported. Investigating a newsworthy issue features a natural tradeoff between checking further sources more carefully and running a story as early as possible.[18]

This subsection considers an *investigation race* between two newspapers to examine how this time tradeoff affects precision levels. Each paper's precision is determined by the time elapsed until it starts reporting the story. The following results show that the investigation race leads to specialisation of the two papers into an early reporter with lower precision and a late reporter with higher precision. This is the case even if their productivity levels, the increase in precision per investigated time, are identical. More than that, when the precision levels are unequal, the investigation race exacerbates the inequality: the more productive newspaper will deepen its advantage by investigating longer than the less productive competitor. Further comparative statics offered below show that increasing initial public precision may lead to a decrease of total final precision.

To allow for cleaner exposition, the following results are presented in terms of a continuous-time game in which newspaper i's precision level is $k\rho_i$ after market entry at time k. The increase in precision per instant, ρ_i, can be interpreted as the investigation productivity of

[18] See the paper 'The thirst to be first' by Lewis and Cushion (2009) for a discussion of the importance of breaking news earlier than competitors.

newspaper i. That is, investigating from time 0 until some time $k \geq 0$ results in a signal $x_i \sim \mathcal{N}(\omega, \frac{1}{k\rho_i})$.[19] Appendix A.3 presents the discrete-time game underlying this subsection. The outcomes presented here are to be interpreted as equilibrium results in the discrete-time game, considering arbitrarily short periods. I assume that the receiver incurs attention costs only after the first sender entered the market. One interpretation is that the issue at hand only becomes eminent for the receiver after the first piece of news is offered. Furthermore, I assume senders are productive enough so that their investigation is initially worthwhile from the receiver's perspective: the marginal increase in utility, $\frac{\partial}{\partial k}\frac{-1}{p_0+\rho_i k}$, at $k = 0$ is higher than the marginal cost, or, equivalently:

Assumption 2. *For both newspapers $i = 1, 2$:*

$$\rho_i > p_0^2 c. \tag{A2}$$

I focus on pure-strategy equilibria and restrict attention to equilibria in AoN strategies such that, after both newspapers entered the market, they play the AoN equilibrium from the previous subsection. In what follows, such equilibria are called *pure AoN equilibria*. This restriction rules out collusive equilibria in which the equilibrium selection after the second paper enters is used to punish or reward specific entry choices. See Appendix A.3 for a discussion of other equilibria.

The first result for the entry game states that there cannot be a pure AoN equilibrium in which both newspapers enter the market at the same time. Consequently, I will refer to the first paper to enter the market as the *leader* and to the second paper as the *follower*.

Lemma 2. *In any pure AoN equilibrium of the investigation race, senders enter at different times. Suppose the leader, $i = \ell$, enters at time k_ℓ. Then,*

i) the follower, $i = f$, enters only after the leader has revealed all information.

ii) the leader's expected payoff is

$$\mathbb{E}[k_f] - k_\ell = \frac{1}{c}\left(\frac{-1}{p_0 + k_\ell \rho_\ell} - \frac{-1}{p_0}\right).$$

iii) if the follower enters at k_f, her expected payoff is

$$\frac{1}{c}\left(\frac{-1}{p_0 + k_\ell \rho_\ell + k_f \rho_f} - \frac{-1}{p_0 + k_\ell \rho_\ell}\right).$$

Lemma 2 states further that (i) the follower will enter the market only once the leader has no private information. At this point, not entering would induce the receiver to stop. As

[19]This arises for example if we assume that paper i observes a Brownian motion with drift ω and instantaneous variance $\frac{1}{\rho_i}$.

long as the leader has enough private information to keep the receiver engaged, the follower prefers to increase her precision and enter later. More precision gives the follower a higher payoff in the AoN equilibrium after she enters. In turn, the leader will not risk the receiver stopping as long as she has enough private information. Item (*ii*) says that the leader keeps the receiver indifferent between stopping at prior information (with utility $\frac{-1}{p_0}$) and observing the leader's information.[20] Similarly for item (*iii*), the follower is a monopolist once she enters the market[21] and keeps the receiver indifferent between stopping at the current information (utility $\frac{-1}{p_0+k_\ell\rho_\ell}$) and stopping with both senders' information (utility $\frac{-1}{p_0+k_\ell\rho_\ell+k_f\rho_f}$).

The two papers separate into different editorial processes. The leader starts informing after checking fewer sources, and the follower investigates as long as possible to deliver more in-depth information.

The payoffs in Lemma 2 pin down the expected payoffs in the investigation race as a function of the leader's identity and her entry time k_ℓ. Let $L_i(k_\ell)$ be paper i's payoff if it enters as the leader ($\ell = i$) at time k_ℓ. Let $F_i(k_\ell)$ be paper i's payoff if it becomes the follower as paper $\ell \neq i$ enters at time k_ℓ. The following result collects the properties of the functions L_i and F_i that allow characterising the unique pure AoN equilibrium of the investigation race.

Theorem 2. *For both newspapers $i = 1, 2$;*

- *the leader's payoff $L_i(k_\ell)$ is strictly increasing for all $k_\ell \geq 0$.*
- *there is a time $k_i^* > 0$ with the property that, for all $k_\ell \leq k_i^*$, we have $F_i(k_\ell) \geq L_i(k_\ell)$, and for all $k_\ell > k_i^*$, we have $F_i(k_\ell) < L_i(k_\ell)$.*
- *$k_1^* < k_2^*$ if and only if $\rho_1 < \rho_2$.*

In the unique pure AoN equilibrium of the investigation race, the less productive paper starts reporting first at time $k^* = \max\{k_1^*, k_2^*\}$.

To gain intuition for this result, consider Figure 2, which depicts the case in which $\rho_2 > \rho_1$. After $k^* = \max\{k_1^*, k_2^*\} = k_2^*$, both papers strictly prefer to enter as the leader. By continuity, for any potential leader entry time later than k^*, the follower prefers to undercut slightly. For paper i, entering the market as the leader at any $k < k_i^*$ is dominated by entering at k_i^*: if the competitor does not enter before, this is due to the monotonicity of L_i, if the competitor does consider entry before, this is due to $F_i > L_i$. Paper 2 will not enter as the leader before k_2^*, the time at which she is indifferent between entering and becoming the leader or becoming the follower by 1's entry. If 2 does not stop at k_2^* (but at any time strictly later), the best response of paper 1, is to enter as the leader at k_2^*.

[20] By item (*i*), there are no competing offers from the follower.
[21] The leader's information was fully revealed before.

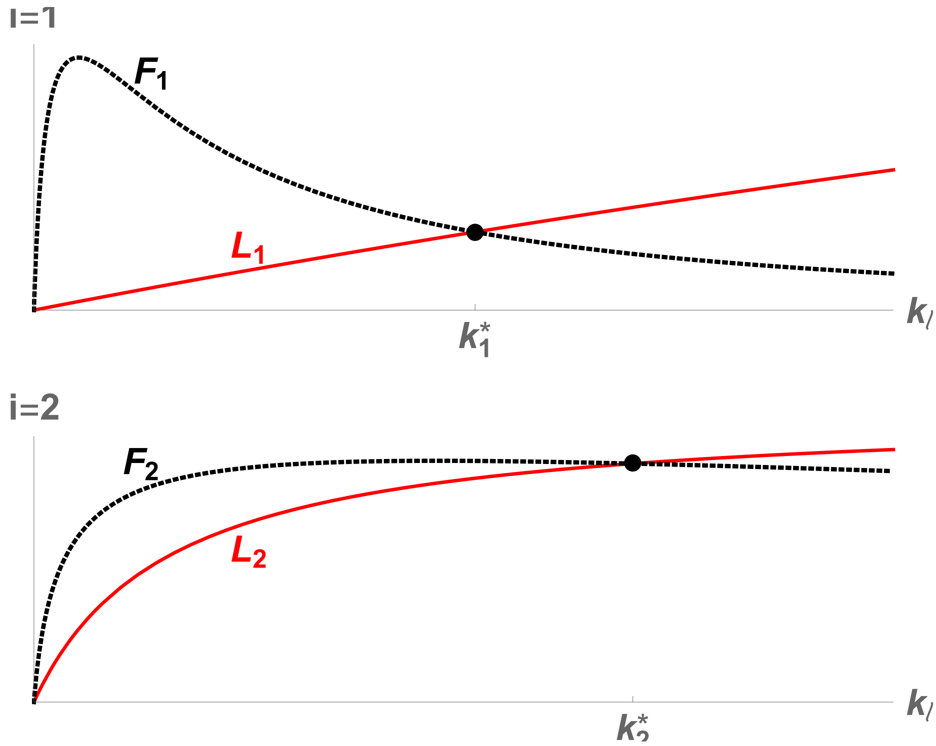


Figure 2: Leader and Follower Payoffs

Theorem 2 shows that the investigation race presented in this model exacerbates the inequality in precision levels. The more productive paper investigates longer. This resonates with a news cycle in which the 'yellow press' paper first runs a news story with less careful fact-checking, and a more investigative newspaper informs the receiver later but more precisely.

With Theorem 2, we can do comparative statics on the total information discovered in equilibrium. Assume from now on that $\rho_2 \geq \rho_1$, so that newspaper 1 is the first one to enter the market at $k^* = k_2^*$. Then, the expected utility of the receiver from the action is $\mathbb{E}_{k_2}\left[\frac{-1}{p_0+k^*\rho_1+k_2\rho_2}\right]$. This gives a measure for the total information obtained in this game. The following result considers how it changes in parameter values.

Lemma 3. *Holding all other parameters fixed, in the investigation-race equilibrium, the value* $\mathbb{E}_{k_2}\left[\frac{-1}{p_0+k^*\rho_1+k_2\rho_2}\right]$ *is*

i) decreasing in c, and

ii) decreasing in p_0 for all $p_0 \in [0, \underline{p}]$ with $\underline{p} > 0$.

Point *i*) states that lower attention costs lead to a higher level of knowledge reached in equilibrium. According to point *ii*), interestingly, the overall information may decrease if p_0 increases. Hence, if society considers a measure that delivers public information initially, the incentive effect on the papers that will investigate less as a response may outweigh the first-order effect and lead to less overall information.

The intuition for this last result is as follows. How long the follower investigates is determined by the time the leader can report on an issue with her own information. As the prior precision becomes very small, the leader can attract a lot of attention even with little information gathered previously. The follower can then investigate for a long time.

6 Discussion

6.1 Experiment-Based vs Belief-Based Modelling

I model information using a signal-/experiment-based approach instead of the commonly used belief-based approach of working directly in the space of distributions over posterior beliefs. See Kamenica and Gentzkow (2011) for static and Ely et al. (2015) for dynamic settings. The experiment-based approach is more convenient for games with multiple senders and dynamic games in particular.

Gentzkow and Kamenica (2016) study a static multiple-sender game with belief-based techniques. They introduce *Blackwell connectedness*, a condition on the information senders can offer. It ensures that each sender can unilaterally deviate to any feasible but more informative posterior distribution. In a simultaneousmove one-shot game, this condition allows them to consider Nash equilibria in which senders choose the same posterior distribution, and no sender has an incentive to deviate to a more informative posterior distribution. In the setting presented here, Blackwell connectedness holds if and only if all senders have one identical signal. Due to the competition, this case is rather uninteresting in my model. There is a competitive equilibrium in which all senders offer all their information in the first round. The receiver chooses randomly which sender to visit, after which all information is observed and the game ends.

Another reason for the experiment-based approach is that different beliefs over signal x_i may arise depending on the information observed previously. Identical offers would give different distributions over posterior beliefs depending on the current belief. Alonso and Camara (2016) identify a bijection between the posteriors that emerge when players with different priors update their beliefs through a commonly understood signal. Given that histories are public in my game, this connection would allow me to set up the model with the belief-based approach: letting the sender choose the posterior distribution for some baseline belief and keep track of the resulting posterior distributions corresponding to different beliefs. However, with a large set of possible beliefs that can emerge in any round, this is not tractable. Furthermore, the model with experiments can be easily extended to the case with multiple receivers who may have observed different realisations before choosing from the same set of signals, and the case in which senders cannot observe the realisations of their competitors.[22]

[22] Section 5 allows for the second possibility.

6.2 Lack of Commitment and Poisson Arrival

I relate my result of the single-sender case to Zhong (2019) and identify the lack of intertemporal commitment by the sender as an additional driver in favour of Poisson processes against Gaussian processes.

Zhong (2019) shows that Poisson learning is uniquely optimal for a decision maker who designs an optimal information process subject to costs proportional to the expected reduction in entropy. His paper shows that this is driven by discounting and the resulting risk preferences. With linear time-costs instead of discounting, Poisson and Gaussian information are both optimal for the decision maker. Featuring no discounting, my model identifies another channel that requires a jump component in the revelation of information – namely, the lack of intertemporal commitment by the sender. Together with the non-concavity in the value of information, this requires discrete jumps in the receiver's belief with positive probability, even in the limit as the length of a period goes to 0.

To illustrate this, consider the following example. The receiver has to guess the outcome of a coin flip, which is the information a single sender holds. That is, $\Omega = X_1 = A = \{0, 1\}$ with $x_1 = \omega$. Let the receiver get utility 1 whenever he guesses correctly and 0 otherwise. Then, for belief $\xi = Pr[\omega = 1]$, we have $U(\xi) = \max\{\xi, 1-\xi\}$. If the current belief is ξ, Lemma 1 implies that the monopolist gets $(1 - \max\{\xi, 1 - \xi\})\,\frac{1}{c}$ rounds of attention. This holds for all rounds and beliefs.[23] For the receiver to be willing to pay the attention cost c in the current round, any experiment has to satisfy

$$\mathbb{E}_m\left[\max\{1-\xi'(\xi,m), \xi'(\xi,m)\}\right] - c \geq \max\{1-\xi,\xi\}.$$

It follows that the message m resulting from the offered experiment has to change the receiver's optimal action with positive probability. If the chosen action stays constant,

$$\begin{aligned}\mathbb{E}_m\left[\max\{1-\xi'(\xi,m), \xi'(\xi,m)\}\right] &= \max\left\{\mathbb{E}_m\left[1-\xi'(\xi,m)\right], \mathbb{E}_m\left[\xi'(\xi,m)\right]\right\}\\ &= \max\{1-\xi,\xi\}\end{aligned}$$

and the experiment's value is 0. Whenever the current belief ξ is different from $\frac{1}{2}$, this implies that, with positive probability, the experiment has to induce a discrete jump in the belief.

[23] As long as $(1 - \max\{\xi, 1 - \xi\}) \geq c$. If this is not fulfilled, no further information transmission is possible.

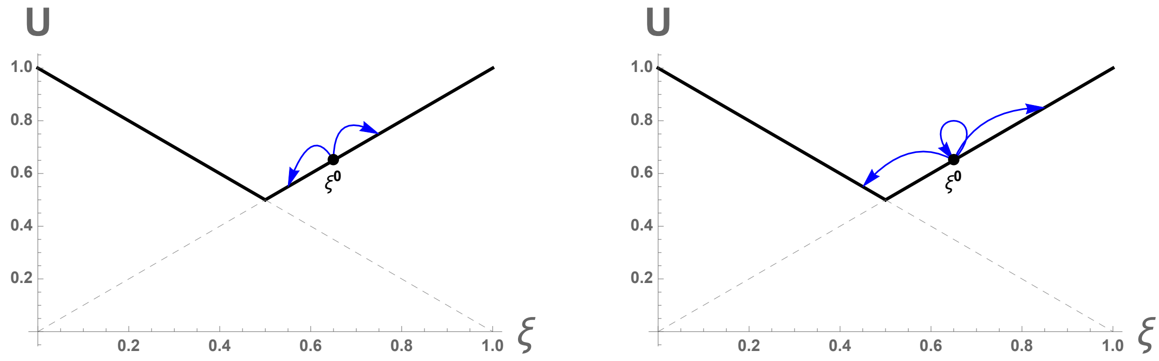


Figure 3: Experiments without and with action change

The left panel in Figure 3 shows an experiment with value 0. Since the action is unchanged at both possible posteriors, and U is linear in between, the expected stopping utility is unchanged. To deliver positive value, any experiment has to change the action with positive probability, which implies for the example in the graph that a posterior belief $< \frac{1}{2}$ has to be reached with positive probability, as shown in the right panel. This is certainly true for discrete rounds that require cost $c > 0$. Yet, jumps in the revelation remain necessary, even in the continuous-time limit with attention cost cdt per interval with length dt. Zhong (2019) shows that without loss, any posterior belief process can be decomposed into a Poisson component with jumps and a gradual Gaussian component. Letting the period length go to 0, the probability of a belief change induced by a Gaussian process vanishes exponentially. Together with the above observation, we can conclude that the information offered by the sender has to include at least some jump component, even as periods become arbitrarily small. Note that the reason for Poisson here is different from the risk preferences induced by discounting in Zhong (2019). In the current model without discounting, Poisson is required by lack of intertemporal commitment on the side of the sender.

6.3 Complementary Signals and Hold-Up Problem

To illustrate how complementarities in the senders' information hinder transmission in equilibrium, consider two senders. Each sender observes the outcome of an independent, fair coin flip. The receiver has to guess whether the two coins match or not. Let the receiver's utility again be 1 if he guesses correctly and 0 otherwise. In this case, the two signals (coin flips) are perfect complements. In particular, the value of observing one signal without any information about the other is 0.

Suppose that sender 1 has revealed the result of her coin flip. Sender 2 is a monopolist and requires $\frac{1}{c}\frac{1}{2}$ visits in expectation to reveal her information.[24] Anticipating this, the receiver's

[24] The value of sender 2's information after knowing x_1 is the difference between being able to guess correctly for sure or with probability $\frac{1}{2}$.

willingness to pay for sender 1's signal is 0. The receiver is not willing to invest a single visit, even if sender 1 offers to reveal her information for sure. The cost $c > 0$ is too high.

There can be no information transmission in equilibrium due to this hold-up problem that arises with one sender after having observed the other sender's information.

Going away from the case where the first sender offers to reveal her information perfectly, suppose that sender 1 revealed partial information, and for concreteness, let the current belief about x_1 be ξ_1 with $\frac{1}{2} < \xi_1 < 1$. Then, the value of sender 2's information is $\bar{v}_2(\xi_1) = \xi_1 - \frac{1}{2}$. After knowing the result of the second coin, the receiver guesses correctly whether they match or not with probability $\xi_1 > \frac{1}{2}$. Sender 2 can extract at least $\bar{v}_2(\xi_1)/c$ units of attention with the corresponding AoN strategy. Note that, as signals are complements, the value of her information will increase in expectation with further revelations about x_1.

7 Concluding Remarks

This paper presents a tractable model to study dynamic information provision by senders who are interested in maximising attention. A simple class of processes suffices to transmit all information from senders to the receiver with minimal attention. For the single sender case, I identify the lack of intertemporal commitment as a novel driver for Poisson information. With competition, I identify a condition on the informational externalities that ensures that all information can be transmitted. The concentration of information on fewer senders decreases the receiver-payoff in his preferred equilibrium. In the case of Gaussian information where each sender's informational endowment can be parametrised by a single number, equal precision levels among senders are preferable for the receiver.

If the senders' precision levels are determined in an investigation race, however, they are polarised. The more efficient newspaper exacerbates its informational advantage by investigating longer than the less efficient competitor. An exogenous increase of initially available public information may decrease the newspapers' incentives to investigate enough to decrease the final precision reached in society. Hence, measures that deliver public information on an issue may be counterproductive and lead to less total information on this issue.

The model lends itself to several extensions that are beyond the scope of this paper. While I assumed that all messages are publicly observable, modelling information as experiments can handle heterogeneous priors. Heterogeneous priors may arise if the senders do not observe the message of the visited competitor or if they do not observe the receiver's visit history. Incorporating such non-observabilities would allow a welfare comparison between the case in which information providers can track their users across sites and the case in which they are not permitted to do so.

Other interesting avenues for future research are different aspects of information acquisition, such as the choice of issues to report on or the decision between seeking more or less correlation with other newspapers. The tractable computation of equilibrium payoffs in this model can be used as a reduced-form of the payoffs and applied to those questions. Lastly, the introduction of prices in addition to attention allows for comparing membership-based business models to advertisement-based business models.

Appendix

A Additional Results and Omitted Proofs

A.1 Updating of Information

The belief about the state, $\mu(\xi)$, if the signal-belief is ξ, is given by

$$\mu(\xi)(\cdot) = \int_X \mu^0_{|\boldsymbol{x}}(\boldsymbol{x}, \cdot) d\xi(\boldsymbol{x}),$$

where $\mu^0_{|\boldsymbol{x}}$ is the conditional probability of the state ω, given the signals $\boldsymbol{x}$. Note that two different signal-posteriors $\xi \neq \hat{\xi}$ may induce the same state-belief $\mu(\xi)$. As discussed after the proof of Theorem 1, working with μ as the state variable would not, therefore, contain enough information. As the signal space X is complete and separable, a regular conditional probability exists. It has the properties that $\mu^0_{|\boldsymbol{x}}(\cdot, W)$ is measurable for all $W \in \mathcal{B}(\Omega)$ and $\mu^0_{|\boldsymbol{x}}(\boldsymbol{x}, \cdot)$ is a probability measure for all $\boldsymbol{x} \in X$.

At the end of round k, the receiver uses message m_k resulting from the selected experiment λ_k to update the belief from ξ^k to ξ^{k+1}. If, in the following expression, the denominator on the right-hand side is non-zero, the receiver forms ξ^{k+1} through Bayes' rule as follows:

$$\xi^{k+1}(\cdot) = \frac{\int_{\cdot} \lambda_k(x_{d_k}, m_k) d\xi^k(\boldsymbol{x})}{\int_X \lambda_k(x_{d_k}, m_k) d\xi^k(\boldsymbol{x})}.$$

In order to define the updating rule $\xi'(\xi^k, m)$ more generally, note that $L^k(\cdot) \equiv \int_X \lambda_k(x_{d_k}, \cdot) d\xi^k(\boldsymbol{x})$ constitutes a probability measure over M. The updating rule $\xi'(\xi^k, m)$ is the non-negative function that satisfies

$$\int_{M'} \xi'(\xi^k, m)(\cdot) dL^k(m) = \int_{\cdot} \lambda_k(x_{d_k} M') d\xi^k(\boldsymbol{x}), \qquad \text{for all } M' \in \mathcal{B}(M).$$

Such a function ξ' exists and is unique L^k-almost everywhere by the Radon-Nikodym Theorem, as for any $X' \in \mathcal{B}(X)$, the right-hand side, interpreted as a measure on M, is absolutely continuous with respect to L^k (see Billingsley, 1995, p. 422).

An experiment from sender i contains information only about x_i directly, i.e. $\lambda(x_i, \cdot)$ is independent of x_j for all $j \neq i$. However, the receiver's belief about x_j will still change through the correlation among signals. If $d_k = i$, the likelihood ratio for two distinct $\boldsymbol{x}_{-i}, \boldsymbol{x}'_{-i}$, given x_i will not be changed through the updating. That is, $\frac{\xi^{k+1}(x_i, \boldsymbol{x}_{-i})}{\xi^{k+1}(x_i, \boldsymbol{x}'_{-i})} = \frac{\xi^k(x_i, \boldsymbol{x}_{-i})}{\xi^k(x_i, \boldsymbol{x}'_{-i})}$ for all x_i.

Lastly, $v \geq 0$ always, since

$$\begin{aligned} U(\xi) &= \max_{a \in A} \mathbb{E}_{\omega \sim \mu(\xi)} [\, u(a, \omega) \,] \\ &= \max_{a \in A} \mathbb{E}_{m|\xi} \left[\mathbb{E}_{\omega \sim \mu(\xi'(\xi, m))} [\, u(a, \omega) \,] \right] \leq \mathbb{E}_{m|\xi} \left[\max_{a \in A} \mathbb{E}_{\omega \sim \mu(\xi'(\xi, m))} [\, u(a, \omega) \,] \right], \end{aligned}$$

where I shorten the notation from $\mathbb{E}_{x_i \sim \xi} \mathbb{E}_{m \sim \lambda(x_i, \cdot)}$ to $\mathbb{E}_{m|\xi}$. The second equality is due to the martingale property of beliefs, which ensures that $\mathbb{E}_{m|\xi} [\xi'(\xi, m)] = \xi$ for any experiment.

A.2 Proofs for General Model

Proof of Lemma 1

The AoN equilibrium follows from the text following Lemma 1. Further,

$$\frac{\bar{v}_1\left(\xi^0\right)}{c}$$

is clearly an upper bound for the expected rounds of attention. The AoN strategy ensures the sender this payoff so that she is not willing to deviate to any strategy with a strictly lower payoff. $\square$

Proof of Theorem 1 without Markov Restriction

This proof refers to the proof of Theorem 1, regarding the second step of Claim 1. Let sender i play the AoN strategy from the theorem. By the first step of the claim, the receiver does not stop before sender i's information is revealed. We can therefore determine the number of visits sender i attracts as

$$\mathbb{E}\left[\sum_{n\geq 0}\prod_{m=1}^{n}\left(1-\lambda_i^*(\xi^{k(m)})\right)\right], \tag{A.1}$$

where n counts the number of visits to sender i and $k(m)$ is the round in which sender i is visited the m'th time. The process $\left(\mathbb{E}\left[\bar{v}(\xi'(\xi^0,\boldsymbol{x}_{-i}))\;\middle|\;\mathcal{F}_k\right]\right)_{k\geq 0}$ is a martingale. I write the sigma algebra $\mathcal{F}_k$ explicitly instead of the belief ξ^k. The definition (1) shows that λ^* is a convex function of the above process, so that the process $\left(\lambda^*(\xi^k)\right)_{k\geq 0}$ is a submartingale. For any finite m, the stopping time $k(m)$ is finite almost surely. Further, $\lambda^*\in[0,1]$, so that the submartingale has bounded increments. This implies that we can apply the optional stopping theorem to derive that, for any $m'>m$:

$$1-\lambda_i^*(\xi^{k(m)})\geq\mathbb{E}\left[\left(1-\lambda_i^*(\xi^{k(m')})\right)\middle|\;\mathcal{F}_{k(m)}\right].$$

The following steps show that this permits deriving a lower bound for the number of visits in (A.1) given by

$$\sum_{n\geq 0}\left(1-\lambda_i^*(\xi^0)\right)^n=\frac{1}{\lambda_i^*(\xi^0)}. \tag{A.2}$$

To see how this is derived, consider, for illustration, the sum in (A.1) until $n=2$, which satisfies

$$\begin{aligned}
&\mathbb{E}\left[1+\left(1-\lambda_i^*(\xi^{k(1)})\right)\left(1+\left(1-\lambda_i^*(\xi^{k(2)})\right)\right)\middle|\;\mathcal{F}_0\right]\\
&\geq\mathbb{E}\left[1+\mathbb{E}\left[\left(1-\lambda_i^*(\xi^{k(2)})\right)\middle|\;\mathcal{F}_{k(1)}\right]\left(1+\left(1-\lambda_i^*(\xi^{k(2)})\right)\right)\middle|\;\mathcal{F}_0\right]\\
&=\mathbb{E}\left[1+\mathbb{E}\left[\left(1-\lambda_i^*(\xi^{k(2)})\right)\middle|\;\mathcal{F}_{k(1)}\right]\left(1+\mathbb{E}\left[\left(1-\lambda_i^*(\xi^{k(2)})\right)\middle|\;\mathcal{F}_{k(1)}\right]\right)\middle|\;\mathcal{F}_0\right].
\end{aligned}$$

The inequality uses (A.2) and the equality follows from the tower property of conditional expectations. This step can be reiterated. By Doob's martingale convergence theorem, the limit

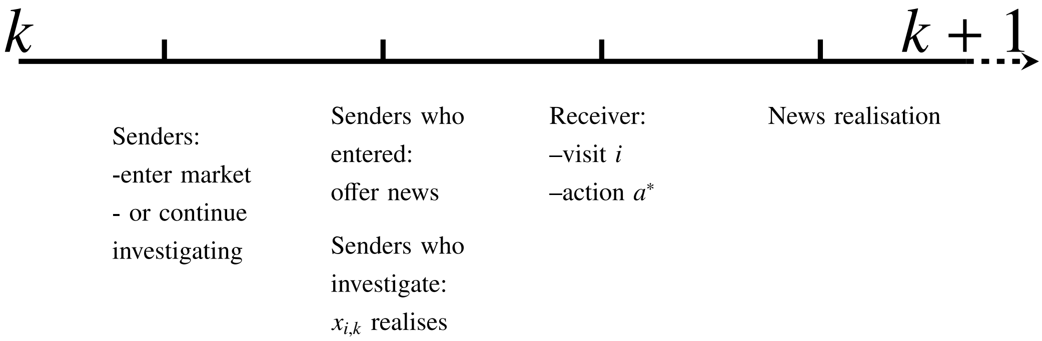
$$\lim_{k\to\infty}\mathbb{E}\left[\left(1-\lambda_i^*(\xi^k)\right)\middle|\;\mathcal{F}_{k(m)}\right]$$

exists and is smaller or equal to $\left(1-\lambda_i^*(\xi^0)\right)$. Applying these steps for all $n\in\mathbb{N}_0$, where the submartingale inequality and the tower property have to be used repeatedly for terms with $n>2$, gives the desired result. $\square$

A.3 Proofs for Investigation Race

This section presents the discrete time investigation race underlying Section 5. As before, the state distribution is $\mathcal{N}(0, 1/p_0)$. The length of each time period is $\Delta > 0$. Each newspaper's precision is determined endogenously in the following stopping game. To obtain information about the state, the papers can investigate before entering the market to disseminate news. Investigating in round n, i.e. from time $n\Delta$ until $(n+1)\Delta$, means that newspaper i is endowed with signal $x_{i,k} \sim \mathcal{N}(\omega, \frac{1}{\rho_i}\frac{1}{\Delta})$. Conditional on the state, signals are independent across senders and rounds. Entering the market allows the newspaper to offer news from that round onward. Note that the normal distribution implies that the signals gathered by sender i from round 0 up to market entry at round n are equivalent to observing a normal signal with precision level $n\rho_i\Delta$. The senders get payoff Δ per round. The receiver's cost is $c\Delta$ and, as mentioned above, he incurs costs only after the first newspaper entered.

Timing



The timing in each round is as follows. Senders decide whether to enter the market. This decision is publicly observed. Senders who entered in this round or before, offer news. For senders who continue investigating, signal $x_{i,k}$ realises. News offers become public, and the receiver decides whether to visit one of the senders who offers news or take the action.

In the main text, I introduced the assumption, $\rho_i > p_0^2 c$, to ensure that each newspaper is efficient enough so that investigation is efficient initially. For period length Δ, the corresponding assumption is that the first round of investigation be efficient:

$$-\frac{1}{p_0 + \rho_i\Delta} - c\Delta > -\frac{1}{p_0} \;\Leftrightarrow\; \rho_i(1 - p_0 c\Delta) > c p_0^2. \tag{A.3}$$

Note that this assumption implies that for any length Δ, $p_0 c\Delta < 1$. As Δ goes to zero, (A.3) reduces to the assumption in the main text.

As in the main text, I focus on pure strategy equilibria, and I rule out that different entry times are rewarded or punished through the equilibrium that is played after both senders enter the market by focusing on equilibria in which, after both senders are in the market, they play the AoN continuation equilibria corresponding to Theorem 1. However, different from the main text, this section also considers different equilibrium strategies by the leader while she is the only sender in the market.

Results

Lemma A.1. *Suppose the leader has entered the market in round n_ℓ and the follower has not entered. In any equilibrium of the continuation game, the follower does not enter the market before the leader has revealed all her information to the receiver.*

Proof. Playing the AoN equilibria after the follower's market entry implies that her payoff, as a function of entry rounds n_ℓ of the leader and n_f of the follower, is

$$\left(\frac{n_f \rho_f \Delta}{c(p_0 + n_\ell \rho_\ell \Delta)(p_0 + n_\ell \rho_\ell \Delta + n_f \rho_f \Delta)}\right) \mathbb{1}_{\{\frac{n_f \rho_f}{c(p_0+n_\ell\rho_\ell\Delta)(p_0+n_\ell\rho_\ell\Delta+n_f\rho_f\Delta)}>1\}}.$$

The indicator function whenever

$$\frac{-1}{p_0 + n_\ell \rho_\ell \Delta + n_f \rho_f \Delta} - \frac{-1}{p_0 + n_\ell \rho_\ell \Delta} > c\Delta.$$

This is condition (A1), ensuring that the information held by the follower is worth at least one round of attention. If the precision of the leader grows too large and her information is revealed too early, the follower cannot attract any attention.

This payoff is increasing in n_f (both the value and the likelihood that the indicator function is one), so that the follower will enter the market only if the receiver would otherwise stop in this round. In the case that the follower does not enter, what makes the receiver stop? If the leader has revealed too much information so that giving out her exact signal is worth less than $c\Delta$, the receiver stops. Waiting one more round and hoping that the follower will enter is not profitable as the follower will extract all the surplus from the receiver. For the leader, it is optimal to replace any such realisations that would lead the receiver to stop (absent entry of the follower) with full information revelation. To see why, consider the leader's payoff in such a round. That is, in a round n in which she offers an experiment λ, such that the set $M^{st} \equiv \{m : \frac{1}{p_0+\rho_\ell n_\ell \Delta(1-\frac{Var(\mathbb{E}[x_\ell|\xi'(\xi^n,m)])}{Var(x_\ell)})} - \frac{1}{p_0+\rho_\ell n_\ell \Delta} < c\Delta\}$ occurs with positive probability. M^{st} includes all messages that make the information about the leader's signal x_ℓ precise enough so that the receiver is not willing to spend a further $c\Delta$, even with the promise of getting all information.[25] Clearly, the leader cannot attract any further visit after a message in M^{st} has realised. The following change in the offered experiment increases the leader's expected utility and ensures that the receiver still accepts. Consider the overall probability of such a message

$$L^n(M^{st}) = \int_X \int_{M^{st}} \lambda(x_\ell, dm) \xi^n(d\boldsymbol{x})$$

and replace all messages in this set by revealing no information with probability $\alpha L^n(M^{st})$ and all information with probability $(1-\alpha)L^n(M^{st})$. To ensure the same continuation value for the receiver, α is chosen such that

$$\begin{aligned} &\alpha \frac{-1}{p_0 + n_\ell \rho_\ell \Delta(1 - \frac{Var(\mathbb{E}[x_\ell|\xi^n])}{Var(x_\ell)})} + (1-\alpha)\frac{-1}{p_0 + n_\ell \rho_\ell \Delta} \\ &= \frac{1}{L^n(M^{st})} \int_X \int_{M^{st}} \frac{-1}{p_0 + \rho_\ell n_\ell \Delta(1 - \frac{Var(\mathbb{E}[x_\ell|\xi'(\xi^n,m)])}{Var(x_\ell)})} \lambda(\tilde{x}_\ell, dm) \xi^n(d\tilde{\boldsymbol{x}}). \end{aligned}$$

The right-hand side is the expected value after a message of set M^{st} (note that even with the follower entering, the receiver will be left at her current stopping utility because of the follower's monopoly power). The left-hand side equals this expected utility, either giving no further information or all information held by the leader. Note that $\alpha \in [0,1]$ since for all m, $Var(\mathbb{E}[x_\ell|\xi^n]) \leq Var(\mathbb{E}[x_\ell|\xi'(\xi^n, m)]) \leq Var(x_\ell)$. The leader is better off, the probability with which this round is her last round of attention decreases, and with positive probability, she reached the next round with the receiver's belief remaining unchanged. □

Knowing that the follower keeps investigating instead of competing actively as long as the leader still holds private information gives rise to the following result, akin to Lemma 1 in the main text with the leader in the role of the monopolist. In every round after she enters, the leader will not offer more information than necessary to

[25] Note that the set M^{st} depends on the current belief and the experiment offered.

keep the receiver indifferent between stopping and visiting.

Lemma A.2. *Suppose the leader enters in round n_ℓ. Let $n_f - 1$ be the round in which all her information is revealed. Then, her expected payoff, $(\mathbb{E}[n_f] - n_\ell)\Delta$, is equal to*

$$(\mathbb{E}[n_f] - n_\ell)\Delta = \frac{1}{c}\left(\frac{1}{p_0} - \frac{1}{p_0 + n_\ell\rho_\ell\Delta}\right) = \frac{n_\ell\rho_\ell\Delta}{cp_0(p_0 + n_\ell\rho_\ell\Delta)}.$$

This pins down the follower's expected entry time n_f.[26] The realisation of n_f, however, depends on the leader's offer strategy applied from n_ℓ onward. There are several such strategies. The main text focused on the equilibrium in which the leader plays an AoN strategy from n_ℓ onward until her information finally realises. Therefore, I will first consider this equilibrium in what follows and provide the proofs for the main text. Subsequently, I consider different strategies to argue that the effects and results presented in the main text do not hinge on this equilibrium selection.

The AoN offer for the leader entering in round n_ℓ reveals her information in each period with probability

$$\lambda_\ell(n_\ell) = \frac{cp_0(p_0 + n_\ell\rho_\ell\Delta)}{n_\ell\rho_\ell}.$$

By the assumptions above, this is smaller than one for any $n_\ell \geq 1$.

Fixing the leader's strategy, the payoffs of leader and follower in the investigation race are determined by the identity of the leader ($\ell = 1$ or $\ell = 2$) and the entry time n_ℓ. For the leader, $i = \ell$:

$$L_i(n_\ell) = \frac{1}{c}\frac{n_\ell\rho_i\Delta}{p_0(p_0 + n_\ell\rho_i\Delta)} = \frac{1}{\lambda_i(n_\ell)}. \tag{A.4}$$

For the follower, $i = f$ and $j = \ell$:

$$F_i(n_\ell) = \sum_{n_f = n_\ell+1}^{\infty} \left(1 - \lambda_j(n_\ell)\right)^{n_f - (n_\ell+1)} \lambda_j(n_\ell)\frac{1}{c}\frac{n_f\rho_i\Delta\mathbb{1}_{\{\frac{n_f\rho_i}{c(p_0+n_\ell\rho_j\Delta)(p_0+n_\ell\rho_j\Delta+n_f\rho_i\Delta)}>1\}}}{(p_0 + n_\ell\rho_j\Delta)(p_0 + n_\ell\rho_j\Delta + n_f\rho_i\Delta)}. \tag{A.5}$$

If both newspapers enter in the same period:

$$\begin{aligned}
B_i(n_\ell) \qquad &= \mathbb{1}_{\{\frac{1}{c}\frac{n_\ell\rho_i}{(p_0+n_\ell\rho_j\Delta)(p_0+n_\ell(\rho_1+\rho_2)\Delta)}>1\}}\frac{1}{c}\frac{n_\ell\rho_i\Delta}{(p_0 + n_\ell\rho_j\Delta)(p_0 + n_\ell(\rho_1+\rho_2)\Delta)} \\
&+ (1 - \mathbb{1}_{\{\frac{1}{c}\frac{n_\ell\rho_i}{(p_0+n_\ell\rho_j\Delta)(p_0+n_\ell(\rho_1+\rho_2)\Delta)}>1\}})\max\{0, \frac{1}{c}\frac{n_\ell\rho_i\Delta}{(p_0 + n_\ell\rho_j\Delta)(p_0 + n_\ell(\rho_1+\rho_2)\Delta)} \\
&+ \Delta - \frac{1}{c}\frac{n_\ell\rho_j\Delta}{(p_0 + n_\ell\rho_i\Delta)(p_0 + n_\ell(\rho_1+\rho_2)\Delta)}\}.
\end{aligned}$$

For the follower i, each round $n_f > n_\ell$ is reached with probability $\left(1 - \lambda_j(n_\ell)\right)^{n_f-(n_\ell+1)}$. If the leader's information hits, which happens with probability $\lambda_j(n_\ell)$, the follower receives attention that makes the receiver indifferent between only the leader's or both the leader's and the follower's information. However, as mentioned above, this is only if the difference is worth at least one visit, as captured by the indicator function.

The next results show that if Δ is small enough, there cannot be an equilibrium in which both papers enter the market at the same time, unless both enter in the very first round.

Lemma A.3. *There exist $\epsilon > 0$ such that for any $\Delta \leq \epsilon$, for all $n > 1$, for $i = 1, 2$:*

$$F_i(n) > B_i(n).$$

[26] For completeness, I call n_f the follower's entry time, even if the indicator function above is 0 and she cannot attract any attention.

Proof.

$$
\begin{aligned}
&F_i(n) > B_i(n) \\
&\Longleftarrow \sum_{n_f=n+1}^{\infty} \left(1 - \lambda_j(n)\right)^{n_f-(n+1)} \lambda_j(n) \frac{n_f \rho_i \Delta}{(p_0 + n\rho_j \Delta)(p_0 + n\rho_j \Delta + n_f \rho_i \Delta)} \\
&\qquad > \frac{n\rho_i \Delta}{(p_0 + n\rho_j \Delta)(p_0 + n(\rho_1 + \rho_2)\Delta)}.
\end{aligned}
$$

Note that the second line inequality is sufficient because in the cases where the indicator function in any term in F is 0, it is also 0 for B. In the opposite case, one of the claimed inequalities is always fulfilled. Furthermore, the fraction is increasing in n_f, so that any term of the left-hand sum is greater than the right-hand fraction. It is multiplied with the probability function of a geometric distribution, which sums to one so LHS has to be greater than RHS. □

This shows that there cannot be a pure strategy equilibrium in which the senders enter the market in the same round. The next result shows that if Δ is small enough, the follower's payoff in early periods strictly exceeds the leader's payoff for both players. This, together with the fact that the leader's payoff is increasing in n_ℓ, implies that entering as the leader is strictly dominated by investigating in early periods.

Lemma A.4. *There exist $\epsilon > 0$ such that for any $\Delta \leq \epsilon$, for both papers, $i = 1$ and $i = 2$: there exists n_i^* >1, such that $F_i(n_\ell) > L_i(n_\ell)$ for all $n_\ell < n_i^*$.*

Proof. To take care of the indicator function in F_i, characterised in (A.5), note that the term on the LHS of the inequality is increasing in n_f. Therefore, we can define

$$
\underline{n}_j(n_\ell) \equiv \min \left\{ n \in \mathbb{N} \;\middle|\; n \geq n_\ell + 1 \;\cap\; \frac{n\rho_j}{c(p_0 + n_\ell \rho_i \Delta)(p_0 + n_\ell \rho_i \Delta + n\rho_j \Delta)} > 1 \right\}.
$$

We can write $L_i(n_\ell) < F_i(n_\ell)$ as

$$
\frac{n_\ell \rho_i \Delta}{p_0(p_0 + n_\ell \rho_i \Delta)} < \sum_{n_f=\underline{n}_i(n_\ell)}^{\infty} \left(1 - \lambda_j(n_\ell)\right)^{n_f-(n_\ell+1)} \lambda_j(n_\ell) \frac{n_f \rho_i \Delta}{(p_0 + n_\ell \rho_j \Delta)(p_0 + n_\ell \rho_j \Delta + n_f \rho_i \Delta)}.
$$

The last fraction in the sum satisfies

$$
\frac{n_f \rho_i \Delta}{(p_0 + n_\ell \rho_j \Delta)(p_0 + n_\ell \rho_j \Delta + n_f \rho_i \Delta)} = \frac{1}{p_0 + n_\ell \rho_j \Delta} \left(1 - \frac{p_0 + n_\ell \rho_j \Delta}{p_0 + n_\ell \rho_j \Delta + n_f \rho_i \Delta}\right).
$$

The entire sum is then equal to

$$
\begin{aligned}
&\sum_{n_f=\underline{n}_i(n_\ell)}^{\infty} \Big(1-\lambda_j(n_\ell)\Big)^{n_f-(n_\ell+1)} \lambda_j(n_\ell)\frac{1}{p_0+n_\ell\rho_j\Delta} \\
&- \sum_{n_f=\underline{n}_i(n_\ell)}^{\infty} \Big(1-\lambda_j(n_\ell)\Big)^{n_f-(n_\ell+1)} \lambda_j(n_\ell)\frac{1}{p_0+n_\ell\rho_j\Delta+n_f\rho_i\Delta} \\
&= \Big(1-\lambda_j(n_\ell)\Big)^{\underline{n}_i(n_\ell)-(n_\ell+1)} \lambda_j(n_\ell)\frac{1}{p_0+n_\ell\rho_j\Delta}\sum_{m=0}^{\infty}\Big(1-\lambda_j(n_\ell)\Big)^{m} \\
&- \sum_{n_f=\underline{n}_i(n_\ell)}^{\infty} \Big(1-\lambda_j(n_\ell)\Big)^{n_f-(n_\ell+1)} \lambda_j(n_\ell)\frac{1}{p_0+n_\ell\rho_j\Delta+n_f\rho_i\Delta} \\
&= \Big(1-\lambda_j(n_\ell)\Big)^{\underline{n}_i(n_\ell)-(n_\ell+1)} \\
&\left(\frac{1}{p_0+n_\ell\rho_j\Delta}-\lambda_j(n_\ell)\sum_{m=0}^{\infty}\Big(1-\lambda_j(n_\ell)\Big)^{m}\frac{1}{p_0+n_\ell\rho_j\Delta+(n_\ell+1)\rho_i\Delta+m\rho_i\Delta}\right).
\end{aligned}
$$

If Δ and n_ℓ are small enough, we get $\underline{n}_i(n_\ell)=n_\ell+1$ so that the above term is larger than

$$
\left(\frac{1}{p_0+n_\ell\rho_j\Delta}-\frac{1}{p_0+n_\ell\rho_j\Delta+(n_\ell+1)\rho_i\Delta}\right)=\frac{(n_\ell+1)\rho_i}{(p_0+n_\ell\rho_j\Delta)(p_0+n_\ell\rho_j\Delta+(n_\ell+1)\rho_i\Delta)}.
$$

With this, a sufficient condition for $L_i(n_\ell)<F_i(n_\ell)$ is

$$
\frac{n_\ell+1}{n_\ell}\frac{p_0}{p_0+n_\ell\rho_j\Delta}\frac{p_0+n_\ell\rho_i\Delta}{p_0+n_\ell\rho_j\Delta+(n_\ell+1)\rho_i\Delta}>1,
$$

which is satisfied for Δ small enough as the second and third fraction become arbitrarily close to 1 as Δ decreases to 0. For fixed Δ, if $\underline{n}_j(n_i)=n_i+1$, then $\underline{n}_j(n)=n+1$ for all $n\le n_i$. The existence of an n_i^* as in the lemma follows as the term above is decreasing in n_ℓ. □

Considering the limit of (A.4) and (A.5) as Δ goes to 0 and n_ℓ goes to ∞ fixing the time $k=n\Delta$, we get that

$$
L_i(k_\ell)=\frac{1}{\lambda_i(k_\ell)}=\frac{k_\ell\rho_i}{cp_0(p_0+k_\ell\rho_i)},
$$

and

$$
\begin{aligned}
F_i(k_\ell)&=\frac{1}{p_0+k_\ell\rho_j}-\int_0^\infty e^{-k\lambda_j(k_\ell)}\lambda_j(k_\ell)\frac{1}{p_0+k_\ell(\rho_j+\rho_i)+k\rho_i} \\
&=\frac{1}{p_0+k_\ell\rho_j}-\frac{\lambda_j(k_\ell)}{\rho_i}e^{\frac{\lambda_j(k_\ell)}{\rho_i}(p_0+k_\ell(\rho_i+\rho_j))}\int_{\frac{\lambda_j(k_\ell)}{\rho_i}(p_0+k_\ell(\rho_i+\rho_j))}^{\infty}\frac{e^{-s}}{s}ds.
\end{aligned}
$$

Proof of Theorem 2

The next results on F and L prove Theorem 2. With the above characterisation, we have for both i that

$$
\lim_{k_\ell\downarrow 0}L_i(k_\ell)=\lim_{k_\ell\downarrow 0}F_i(k_\ell)=0.
$$

Taking the derivative with respect to k_ℓ and considering the limit, gives:

$$\lim_{k\downarrow 0}\left(\frac{\partial L_i(k)}{\partial k}\right) = \frac{\rho_i}{cp_0^2} \qquad \text{and}$$
$$\lim_{k\downarrow 0}\left(\frac{\partial F_i(k)}{\partial k}\right) = \frac{\rho_i}{cp_0^2} + \frac{\rho_i\rho_j}{c^2p_0^4}.$$

Hence, the follower's payoff is higher initially. Furthermore,

$$\lim_{k_\ell\to\infty} L_i(k_\ell) = \frac{1}{cp_0} \quad \text{and}$$
$$\lim_{k_\ell\to\infty} F_i(k_\ell) = 0.$$

This shows that F_i crosses L_i from above at least once. To show that this happens at only one $k_\ell > 0$, I show that the derivative of $F_i(k_\ell) - L_i(k_\ell)$ crosses 0 at most twice. This is sufficient to rule out a second positive intersection point since we have established $\lim_{k\ell\downarrow 0} \frac{\partial}{\partial k_\ell}\left(F_i(k_\ell) - L_i(k\ell)\right) > 0$ and $\lim_{k\ell\to\infty}\left(F_i(k_\ell) - L_i(k\ell)\right) < 0$. At the first intersection, F_i crosses L_i from above. If there were a second intersection point, F_i would again lie above L_i. For $\lim_{k\ell\to\infty}\left(F_i(k_\ell) - L_i(k\ell)\right) < 0$ to hold, this would require a third intersection which, in turn, requires that the derivative be 0 at least three times. Define

$$\psi_i(k_\ell) \equiv \frac{cp_0(p_0 + k_\ell\rho_j)(p_0 + k_\ell(\rho_i + \rho_j))}{k_\ell\rho_i\rho_j},$$

and consider

$$\begin{aligned}
&F_i(k_\ell) - L_i(k_\ell)\\
&= \frac{1}{c(p_0 + k_\ell\rho_j)} - \frac{p_0(p_0 + k_\ell\rho_j)}{k_\ell\rho_i\rho_j} e^{\psi_i(k_\ell)} \int_{\psi_i(k_\ell)}^{\infty} \frac{e^{-s}}{s}\,ds - \frac{k_\ell\rho_i}{cp_0(p_0 + k_\ell\rho_i)}\\
&= \frac{p_0^2 - k_\ell^2\rho_i\rho_j}{cp_0(p_0 + k_\ell\rho_i)(p_0 + k_\ell\rho_j)} - \frac{p_0(p_0 + k_\ell\rho_j)}{k_\ell\rho_i\rho_j} e^{\psi_i(k_\ell)} \int_{\psi_i(k_\ell)}^{\infty} \frac{e^{-s}}{s}\,ds
\end{aligned}$$

Multiply the term with

$$\frac{k_\ell\rho_i\rho_j}{p_0(p_0 + k_\ell\rho_j)} e^{-\psi_i(k_\ell)} > 0.$$

and consider the derivative

$$\begin{aligned}
&\frac{\partial}{\partial k_\ell}\left(\frac{(p_0^2 - k_\ell^2\rho_i\rho_j)(p_0 + k_\ell(\rho_i + \rho_j))}{p_0(p_0 + k_\ell\rho_i)(p_0 + k_\ell\rho_j)} \frac{e^{-\psi_i(k_\ell)}}{\psi_i(k_\ell)} - \int_{\psi_i(k_\ell)}^{\infty} \frac{e^{-s}}{s}\,ds\right)\\
&= e^{-\psi_i(k_\ell)} \frac{c(p_0 + k_\ell\rho_i)(p_0 + k_\ell\rho_j)\left(p_0^2 - k_\ell^2\rho_i\rho_j\right)\left(p_0^2 - k_\ell^2\rho_j(\rho_i + \rho_j)\right)}{ck_\ell p_0(p_0 + k_\ell\rho_i)^2(p_0 + k_\ell\rho_j)^3}\\
&\quad - e^{-\psi_i(k_\ell)} \frac{k_\ell\rho_i\rho_j\left(k_\ell^3\rho_i\rho_j(2\rho_i + \rho_j) + 5k_\ell^2 p_0\rho_i\rho_j + k_\ell p_0^2\rho_j - p_0^3\right)}{ck_\ell p_0(p_0 + k_\ell\rho_i)^2(p_0 + k_\ell\rho_j)^3}\\
&\quad + e^{-\psi_i(k_\ell)} \frac{(p_0^2 - k_\ell^2\rho_j(\rho_i + \rho_j)}{k_\ell(p_0 + k_\ell\rho_j)(p_0 + k_\ell(\rho_i + \rho_j))}.
\end{aligned}$$

This is equal to 0 iff

$$\frac{c(p_0 + k_\ell \rho_i)(p_0 + k_\ell \rho_j)\left(p_0^2 - k_\ell^2 \rho_i \rho_j\right)\left(p_0^2 - k_\ell^2 \rho_j(\rho_i + \rho_j)\right)}{c p_0 (p_0 + k_\ell \rho_i)^2 (p_0 + k_\ell \rho_j)^2}$$
$$- \frac{k_\ell \rho_i \rho_j \left(k_\ell^3 \rho_i \rho_j (2\rho_i + \rho_j) + 5k_\ell^2 p_0 \rho_i \rho_j + k_\ell p_0^2 \rho_j - p_0^3\right)}{c p_0 (p_0 + k_\ell \rho_i)^2 (p_0 + k_\ell \rho_j)^2}$$
$$+ \frac{p_0^2 - k_\ell^2 \rho_j (\rho_i + \rho_j)}{(p_0 + k_\ell(\rho_i + \rho_j))} = 0.$$

Multiplying and rearranging yields

$$\left(p_0 + k_\ell(\rho_i + \rho_j)\right)\left(c(p_0 + k_\ell \rho_i)(p_0 + k_\ell \rho_j)\left(p_0^2 - k_\ell^2 \rho_i \rho_j\right)\left(p_0^2 - k_\ell^2 \rho_j(\rho_i + \rho_j)\right)\right)$$
$$- \left(p_0 + k_\ell(\rho_i + \rho_j)\right)\left(k_\ell \rho_i \rho_j \left(k_\ell^3 \rho_i \rho_j (2\rho_i + \rho_j) + 5k_\ell^2 p_0 \rho_i \rho_j + k_\ell p_0^2 \rho_j - p_0^3\right)\right)$$
$$+ c p_0 (p_0 + k_\ell \rho_i)^2 (p_0 + k_\ell \rho_j)^2 \left(p_0^2 - k_\ell^2 \rho_j (\rho_i + \rho_j)\right) = 0.$$

Collecting k_ℓ with equal exponents gives

$$+ck_\ell^6 \rho_i \rho_j^2 (\rho_i + \rho_j)^2 + k_\ell^5 \left(c p_0 \rho_j^2 (\rho_i + \rho_j)^2 + c p_0 \rho_i \rho_j (\rho_i + 3\rho_j)(\rho_i + \rho_j)\right)$$
$$+k_\ell^4 \left(c p_0^2 \rho_i \rho_j (\rho_i + \rho_j) + c p_0^2 \rho_j (\rho_i + \rho_j)(\rho_i + 3\rho_j) - \rho_i \rho_j (\rho_i + \rho_j)(2\rho_i + \rho_j)\right)$$
$$+k_\ell^3 \left(-c p_0^3 \rho_i^2 - 3c p_0^3 \rho_i \rho_j + c p_0^3 \rho_j (\rho_i + \rho_j) - p_0 \rho_i \rho_j (7\rho_i + 6\rho_j)\right)$$
$$+k_\ell^2 \left(-3c p_0^4 \rho_i - 3c p_0^4 \rho_j - p_0^2 \rho_j (6\rho_i + \rho_j)\right) + k_\ell \left(p_0^3 \rho_i - 2c p_0^5\right) + p_0^4 = 0.$$

The factor after k_ℓ^3 is negative since $\rho_i > c p_0^2$ for both i. Therefore, there are exactly two 'sign changes' in the sequence, and by the rule of signs, the number of positive roots is at most two. Next, I show that the intersection point k_i^* as defined in Theorem 2 is lower for the less efficient newspaper (with lower ρ_i). For this, consider again the equation $F_i - L_i = 0$. This is equivalent to

$$\frac{(p_0^2 - k_\ell^2 \rho_i \rho_j)(p_0 + k_\ell(\rho_i + \rho_j))}{p_0 (p_0 + k_\ell \rho_i)(p_0 + k_\ell \rho_j)} = \psi_i(k_\ell) e^{\psi_i(k_\ell)} \int_{\psi_i(k_\ell)}^{\infty} \frac{e^{-s}}{s} ds. \tag{A.6}$$

with

$$\psi_i(k_\ell) = \frac{c p_0 (p_0 + k_\ell \rho_j)(p_0 + k_\ell(\rho_i + \rho_j))}{k_\ell \rho_i \rho_j}.$$

Note that the left-hand side is symmetric, so that it is the same whether $i = 1$ or $i = 2$. Further, $\psi_i(k_\ell) > \psi_j(k_\ell)$ if and only if $\rho_i < \rho_j$. The function on the right-hand side, $xe^x \int_x^\infty \frac{e^{-s}}{s} ds$, is increasing in x, so that the term on the right crosses the term on the left (from below) at an earlier k_ℓ for higher ψ_i. Hence, the higher efficiency newspaper has a later intersection point. This concludes the proof of Theorem 2. □

Proof of Lemma 3

To establish Lemma 3, suppose wlog that $\rho_1 < \rho_2$, so that paper 1 is the leader and enters the market at $k^* = k_2^*$. To avoid cluttering the notation, I will drop the follower's index from the stopping time and auxiliary function. That is, $k^* = k_2^*$ and $\psi = \psi_2$ for the remainder of the paper. This allows me to write partial derivatives as, for example, ψ_c. Consider the expected precision with which the receiver stops. By Lemma 2, the two papers extract all surplus from the receiver, so that the expected utility from the action, $\mathbb{E}_{k_2}\left[\frac{-1}{p_0+\rho_1 k^*+k_2\rho_2}\right]$, is equal to

$$\frac{-1}{p_0} + c\left(L_1(k^*) + F_2(k^*)\right).$$

Using the formula for L_1 given above and the fact that k^* was defined such that $F_2(k^*) = L_1(k^*)$, this simplifies to

$$\frac{-1}{p_0} + \frac{k^*\rho_1}{p_0(p_0+k^*\rho_1)} + \frac{k^*\rho_2}{p_0(p_0+k^*\rho_2)} = -\frac{p_0^2-(k^*)^2\rho_1\rho_2}{p_0(p_0+k^*\rho_1)(p_0+k^*\rho_2)}$$

Part i). First, consider the change in precision caused by c:

$$\frac{d}{dc}\left(-\frac{p_0^2-k^*(c)^2\rho_1\rho_2}{p_0(p_0+k^*(c)\rho_1)(p_0+k^*(c)\rho_2)}\right) = k_c\frac{4kp_0\rho_1\rho_2+p_0^2(\rho_1+\rho_2)+k^2\rho_1\rho_2(\rho_1+\rho_2)}{(p_0+k\rho_1)^2(p_0+k\rho_2)^2}\Bigg|_{k=k^*}. \tag{A.7}$$

Hence, to show that the total precision is decreasing in c, it is sufficient to show that $k_c < 0$.

By the implicit function theorem and the definition of k^* in (A.6), we can determine the partial derivative k_c as

$$k_c = \frac{-\frac{\partial}{\partial c}\left(\frac{(p_0^2-k^2\rho_1\rho_2)(p_0+k(\rho_1+\rho_2))}{p_0(p_0+k\rho_1)(p_0+k\rho_2)}\right)\psi + \left(\frac{(p_0^2-k^2\rho_1\rho_2)(p_0+k(\rho_1+\rho_2))}{p_0(p_0+k\rho_1)(p_0+k\rho_2)}\right)\psi_c(\psi+1) - \psi_c\psi}{\frac{\partial}{\partial k}\left(\frac{(p_0^2-k^2\rho_1\rho_2)(p_0+k(\rho_1+\rho_2))}{p_0(p_0+k\rho_1)(p_0+k\rho_2)}\right)\psi - \left(\frac{(p_0^2-k^2\rho_1\rho_2)(p_0+k(\rho_1+\rho_2))}{p_0(p_0+k\rho_1)(p_0+k\rho_2)}\right)\psi_k(\psi+1) + \psi_k\psi}\Bigg|_{k=k^*},$$

Recall that the condition used to determine k^* was $F_2 - L_2 = 0$. As established above, F_2 crosses L_2 from above, so that the denominator of the last expression is negative at $k = k^*$.

It follows that $k_c < 0$ if and only if

$$\begin{aligned}
&-\frac{\partial}{\partial c}\left(\frac{(p_0^2-k^2\rho_1\rho_2)(p_0+k(\rho_1+\rho_2))}{p_0(p_0+k\rho_1)(p_0+k\rho_2)}\right)\psi \\
&+\left(\frac{(p_0^2-k^2\rho_1\rho_2)(p_0+k(\rho_1+\rho_2))}{p_0(p_0+k\rho_1)(p_0+k\rho_2)}\right)\psi_c(\psi+1) - \psi_c\psi > 0 \\
&\iff \frac{(p_0^2-k^2\rho_1\rho_2)(p_0+k(\rho_1+\rho_2))}{p_0(p_0+k\rho_1)(p_0+k\rho_2)} > \frac{\psi}{\psi+1}
\end{aligned}$$

Applying again the definition of k^*, the fraction on the left is equal to $\psi e^{\psi}\int_{\psi}\frac{e^{-s}}{s}ds$. The exponential integral satisfies the equation $\psi e^{\psi}\int_{\psi}\frac{e^{-s}}{s}ds > \frac{\psi}{\psi+1}$. The comparative static on c follows.

Part ii). To determine the change in p_0, consider the definition of k^*:

$$k^* = \left\{k > 0 \; : \; \frac{(p_0^2-k^2\rho_1\rho_2)(p_0+k(\rho_1+\rho_2))}{p_0(p_0+k\rho_1)(p_0+k\rho_2)} = \psi e^{\psi}\int_{\psi}^{\infty}\frac{e^{-s}}{s}ds\right\}. \tag{A.8}$$

The expected final precision is given by

$$-\frac{(p_0^2-k^2\rho_1\rho_2)}{p_0(p_0+k\rho_1)(p_0+k\rho_2)}.$$

Given the definition of k^*, we have

$$-\frac{(p_0^2-k^2\rho_1\rho_2)}{p_0(p_0+k\rho_1)(p_0+k\rho_2)} = -\frac{\psi}{(p_0+k(\rho_1+\rho_2))}e^{\psi}\int_{\psi}^{\infty}\frac{e^{-s}}{s}ds.$$

Note that $\frac{\psi}{(p_0+k(\rho_1+\rho_2))} = \frac{cp_0(p_0+k\rho_1)}{k\rho_1\rho_2} = \frac{cp_0^2}{k\rho_1\rho_2} + \frac{cp_0}{\rho_2}$ goes to zero fast enough so that we must have

$$\lim_{p_0 \downarrow 0}\left(-\frac{(p_0^2 - k^2\rho_1\rho_2)}{p_0(p_0+k\rho_1)(p_0+k\rho_2)}\right) = 0.$$

Note that the disutility is weakly negative everywhere. For the term to remain at zero, we need $k = \frac{p}{\sqrt{\rho_1\rho_2}}$ everywhere. This, however, gives strictly positive values on the right-hand side of the definition of k^*.

By continuity, this implies that there must be a $\underline{p} > 0$, such that

$$\frac{d}{dp_0}\left(-\frac{p_0^2 - k(p_0)^2\rho_1\rho_2}{p_0(p_0+k(p_0)\rho_1)(p_0+k(p_0)\rho_2)}\right) < 0$$

for all $p_0 \leq \underline{p}$. □

Gradual Information Release by Leader

While the leader's payoff is fully determined by k_ℓ in any pure strategy equilibrium, the follower's payoff depends on the leader's revelation strategy. The analysis above considered the case where the leader makes AoN offers. However, the leader could choose different distributions, as long as $\mathbb{E}\left[k_f\right] - k_\ell = \frac{k_\ell\rho_\ell}{cp_0(p_0+k_\ell\rho_\ell)}$ and the receiver is willing to pay attention in each round. This subsection considers the equilibrium with maximal information precision. As the senders extract all surplus from the receiver and the leader's payoff is fixed for fixed k_ℓ, this is equivalent to maximising the follower's payoff.[27]

Fixing k_ℓ and the resulting $\mathbb{E}\left[k_f\right]$, the expected payoff of the follower is

$$\mathbb{E}\left[\frac{k_f\rho_f}{c(p_0+k_\ell\rho_\ell)(p_0+k_\ell\rho_\ell+k_f\rho_f)}\right].$$

This is maximised at minimal variance of k_f. With this, we can characterise the information-maximal equilibrium of the stopping game. The payoffs in the stopping game for the leader are

$$L_i(k_\ell) = \frac{1}{c}\frac{k_\ell\rho_i}{p_0(p_0+k_\ell\rho_i)},$$

and for the follower

$$F_i(k_\ell) = \mathbb{E}\left[\frac{1}{c}\frac{k_2(k_\ell)\rho_j}{(p_0+k_\ell\rho_i)(p_0+k_\ell\rho_i+k_2(k_\ell)\rho_j)}\right].$$

In the information-maximal equilibrium, k_2 takes values $k_\ell + \lfloor\frac{k_\ell\rho_1}{cp_0(p_0+k_\ell\rho_1)}\rfloor$ and $k_\ell + \lceil\frac{k_\ell\rho_1}{cp_0(p_0+k_\ell\rho_1)}\rceil$.

In the limit as time periods become small, we get

$$\begin{aligned} F_i(k) &= \frac{1}{c}\frac{(k_\ell + \frac{k_\ell\rho_1}{cp_0(p_0+k_\ell\rho_1)})\rho_j}{(p_0+k_\ell\rho_i)(p_0+k_\ell\rho_i+(k_\ell+\frac{k_\ell\rho_1}{cp_0(p_0+k_\ell\rho_1)})\rho_j)} \\ &= \frac{k_\ell\rho_2(cp_0(p_0+k_\ell\rho_1)+\rho_1)}{c(p_0+k_\ell\rho_1)(cp_0(p_0+k_\ell\rho_1)(p_0+k_\ell(\rho_1+\rho_2))+k_\ell\rho_1\rho_2)}. \end{aligned}$$

The leader cannot commit to giving out more information than necessary to attract the receiver's attention. To make sure all her information is revealed by $k_f = k_\ell + \frac{k_\ell\rho_\ell}{cp_0(p_0+k_\ell\rho_\ell)}$ with probability one, she therefore has to release news gradually, so that the receiver is indifferent between stopping and visiting the sender from k_ℓ until

[27] Note that maximising precision for a fixed stopping time of the leader does not directly imply that the resulting equilibrium of the investigation race has maximal precision. However, this will be the case here as increasing the follower's payoff across all k_ℓ leads to an increase in the entry equilibrium level, k_ℓ^*.

k_f. For $k \in [k_\ell, k_f)$, let $\tau(k)$ be the non-decreasing precision level transmitted from the leader to the receiver.The indifference condition prescribes that at all k:

$$-\frac{1}{p_0 + \tau[k]\rho_\ell} - c(k - k_\ell) = -\frac{1}{p_0}$$
$$\Leftrightarrow \ \tau[k] = \frac{cp_0^2(k - k_\ell)}{\rho_\ell(1 - cp_0(k - k_\ell))}.$$

The information that the leader gives out per instant is then $\tau'[k]$, which increases gradually as time passes from k_ℓ to k_f. The more informed the receiver is, the faster his precision has to increase to keep him from stopping.

References

Alonso, R. and O. Camara (2016). Bayesian Persuasion with Heterogeneous Priors. *Journal of Economic Theory 165*, 672–706.

Anderson, S. P. and R. Renault (2009). Comparative Advertising: Disclosing Horizontal Match Information. *RAND Journal of Economics 40*(3), 558–581.

Angeletos, G.-M. and A. Pavan (2007). Efficient Use of Information and Social Value of Information. *Econometrica 75*(4), 1103–1142.

Anton, J. J., G. Biglaiser, and N. Vettas (2014). Dynamic Price Competition with Capacity Constraints and a Strategic Buyer. *International Economic Review 55*(3), 943–958.

Au, P. H. (2015). Dynamic Information Disclosure. *RAND Journal of Economics 46*(4), 791–823.

Ball, I. (2019). Dynamic Information Provision: Rewarding the Past and Guiding the Future. Working Paper.

Bergemann, D. and A. Bonatti (2019). Markets for Information: An Introduction. *Annual Review of Economics 11*.

Bergemann, D., A. Bonatti, and A. Smolin (2018). The Design and Price of Information. *American Economic Review 108*(1), 1–48.

Bergemann, D. and S. Morris (2013). Robust Predictions in Games With Incomplete Information. *Econometrica 81*(4), 1251–1308.

Bergemann, D. and J. Välimäki (2006). Dynamic Price Competition. *Journal of Economic Theory 127*(1), 232–263.

Besley, T. and A. Prat (2006). Handcuffs for the Grabbing Hand? Media Capture and Government Accountability. *American Economic Review 96*(3), 720–736.

Billingsley, P. (1995). *Probability and Measure*. Wiley Series in Probability and Statistics. Wiley, 3rd ed.

Board, S. and J. Lu (2018). Competitive Information Disclosure in Search Markets. *Journal of Political Economy 126*(5), 1965–2010.

Börgers, T., A. Hernando-Veciana, and D. Krähmer (2013). When are signals complements or substitutes? *Journal of Economic Theory 148*(1), 165–195.

Che, Y.-K. and J. Hörner (2017, 12). Recommender Systems as Mechanisms for Social Learning. *Quarterly Journal of Economics 133*(2), 871–925.

Che, Y.-K., K. Kim, and K. Mierendorff (2020). Keeping the Listener Engaged: A Dynamic Model of Bayesian Persuasion. Working Paper, arXiv:2003.07338.

Che, Y.-K. and K. Mierendorff (2019). Optimal Dynamic Allocation of Attention. *American Economic Review 109*(8), 2993–3029.

Chen, H. and W. Suen (2019). Competition for Attention in the News Media Market. Working Paper.

Choi, M., A. Y. Dai, and K. Kim (2018). Consumer Search and Price Competition. *Econometrica 86*(4), 1257–1281.

Dudey, M. (1992). Dynamic Edgeworth-Bertrand Competition. *The Quarterly Journal of Economics 107*(4), 1461–1477.

Ely, J., A. Frankel, and E. Kamenica (2015). Suspense and Surprise. *Journal of Political Economy 123*(1), 215–260.

Ely, J. C. (2017). Beeps. *American Economic Review 107*(1), 31–53.

Ely, J. C. and M. Szydlowski (2020). Moving the Goalposts. *Journal of Political Economy 128*(2), 468–506.

Fudenberg, D., P. Strack, and T. Strzalecki (2018). Speed, Accuracy, and the Optimal Timing of Choices. *American Economic Review 108*(12), 3651–84.

Fudenberg, D. and J. Tirole (1991). Perfect Bayesian Equilibrium and Sequential Equilibrium. *journal of Economic Theory 53*(2), 236–260.

Galperti, S. and I. Trevino (2018). Shared Knowledge and Competition for Attention in Information Markets. Working Paper.

Gentzkow, M. and E. Kamenica (2016). Competition in Persuasion. *Review of Economic Studies 84*(1), 300–322.

Gentzkow, M. and J. M. Shapiro (2006). Media Bias and Reputation. *Journal of political Economy 114*(2), 280–316.

Gossner, O., J. Steiner, and C. Stewart (2019). Attention please! Working Paper, University of Zurich, Department of Economics.

Guo, Y. and E. Shmaya (2019). The Interval Structure of Optimal Disclosure. *Econometrica 87*(2), 653–675.

Kamenica, E. and M. Gentzkow (2011). Bayesian Persuasion. *American Economic Review 101*(6), 2590–2615.

Ke, T. T. and S. Lin (2020). Informational Complementarity. *Management Science 66*(8), 3699–3716.

Keller, G., S. Rady, and M. Cripps (2005). Strategic Experimentation with Exponential Bandits. *Econometrica 73*(1), 39–68.

Lewis, J. and S. Cushion (2009). The Thirst to be First. *Journalism Practice 3*(3), 304–318.

Liang, A. and X. Mu (2020). Complementary Information and Learning Traps. *The Quarterly Journal of Economics 135*(1), 389–448.

Liang, A., X. Mu, and V. Syrgkanis (2019). Dynamically Aggregating Diverse Information. Working Paper.

Margaria, C. and A. Smolin (2018). Dynamic communication with biased senders. *Games and Economic Behavior 110*, 330–339.

Martínez-de Albéniz, V. and K. Talluri (2011). Dynamic Price Competition with Fixed Capacities. *Management Science 57*(6), 1078–1093.

Mayskaya, T. (2017). Dynamic Choice of Information Sources. Working Paper, California Institute of Technology.

Morris, S. and H. S. Shin (2002). Social Value of Public Information. *American Economic Review 92*(5), 1521–1534.

Morris, S. and P. Strack (2017). The Wald Problem and the Equivalence of Sequential Sampling and Static Information Costs. Working Paper.

Moscarini, G. and L. Smith (2001). The Optimal Level of Experimentation. *Econometrica 69*(6), 1629–1644.

Mullainathan, S. and A. Shleifer (2005). The Market for News. *American Economic Review 95*(4), 1031–1053.

Orlov, D., A. Skrzypacz, and P. Zryumov (2020). Persuading the Principal to Wait. *Journal of Political Economy 128*(7), 000–000.

Pant, A. and F. Trombetta (2019). The Newsroom Dilemma. Working Paper, SSRN 3447908.

Perego, J. and S. Yuksel (2018). Media Competition and Social Disagreement. Working Paper.

Renault, J., E. Solan, and N. Vieille (2017). Optimal Dynamic Information Provision. *Games and Economic Behavior 104*, 329–349.

Smolin, A. (2017). Dynamic Evaluation Design. Working Paper.

Sun, M. (2011). Disclosing Multiple Product Attributes. *Journal of Economics & Management Strategy 20*(1), 195–224.

Vives, X. (1996). Social Learning and Rational Expectations. *European Economic Review 40*(3-5), 589–601.

Wald, A. (1947). Foundations of a General Theory of Sequential Decision Functions. *Econometrica 15*(4), 279.

Wolinsky, A. (1986). True Monopolistic Competition as a Result of Imperfect Information. *Quarterly Journal of Economics 101*(3), 493–511.

Zhong, W. (2019). Optimal Dynamic Information Acquisition. Working Paper.