Costly Advertising and Information Congestion

Ryoji Jinushi*†

November 11, 2022

Abstract

As consumers have limited capacity to process information, advertisers must compete for attention. This creates information congestion which produces social loss like unread advertisements. We apply population games and best response dynamics to analyze information congestion. Multiple equilibria impair traditional policies, and thus, non-traditional policies are examined to lead the system to a Pareto efficient equilibrium. We achieve this, for example, by changing the cost per message multiple times during the evolutionary process. In this process, policymakers gradually investigate externalities. However, these policies are costly, which confirms the inefficiency of advertising structures where advertisers send messages regardless of consumer interests.

JEL Classification: D62; D82; D83; M37; M38

*University of California, Riverside. 900 University Avenue, Riverside, California, 92507(e-mail:rjinu001@ucr.edu). Acknowledgements: Discussions with Richard Arnott and Takamasa Iryo have contributed to this work. I gratefully appreciate their support. I am grateful as well for the very useful comments of Minoru Osawa, Daisuke Fukuda, Guillaume Rocheteau, Siyang Xiong, Victor Ortego-Marti, Yang Xie, Simon Anderson, Ichiro Obara and Barton Lipman. For helpful comments, I thank the following seminar audiences: the 20th International Industrial Organization Conference 2021, the 48th European Association for Research in Industrial Economics 2021, the 16th Economics Graduate Students' Conference 2021, the Royal Economic Society Symposium of Junior Researchers 2022 and the Summer Workshop on Economic Theory 2022.

†Declarations of interest: none.

Keywords: Advertising; Information Congestion; Limited Attention; Evolutionary Implementation; Evolutionary Dynamics; Evolutionary Stability

1 Introduction

Due to the limited capacity of consumers to process information, advertisers have to compete across industries to gain consumer attention. This creates what is called *information congestion*[1]. Information congestion produces a certain loss of social welfare. For example, the advertiser and consumer both lose out when a useful message is missed because of too many other messages[2]. In a primitive advertising structure where consumer information is not available, advertisers cannot avoid sending unsolicited messages to masses of consumers, including non-potential customers. Spam mail is a typical example of such unsolicited messages[3]. It is a large part of all email traffic, approximately 55% (Symantec 2019) to 85% (Cisco 2019), and impairs the efficiency of our economy because of the negligible societal benefit (Rao and Reiley 2012). In a primitive structure in which many advertisers send unsolicited messages, information congestion occurs.

In this study, we theoretically analyze an implementation problem with both hidden actions and information and identify the primitive structure's limitations. This can be used as a benchmark to assess the efficiency of advanced modern advertising structures. In particular, we answer the following question: How can we achieve an efficient allocation of limited consumer attention across many advertisers and industries, when advertisers send messages regardless of consumer interests? Throughout this study, we support the claim that traditional interventions[4] cannot always solve information congestion, and costly interventions are required to guarantee efficient allocation.

In our information congestion model, a continuum of message-senders (i.e., advertisers) send a discrete number of messages to obtain receivers' limited attention capacity, as in the

[1]Early literature, such as Hirshleifer (1973), points to the externality among firms wanting to get consumer attention as a congestion effect.

[2]This is often called *information overload*. Malhotra (1984) discusses earlier contributions regarding information overload in consumer decision making, using TV advertising as an example. Eppler and Mengis (2004) review the early literature about information overload across disciplines.

[3]As Anderson and de Palma (2009) indicate, conventional advertising (such as billboards, radio/television and telemarketing) has a similar structure, where advertisers create and deliver messages regardless of consumer interest or attention.

[4]For example, increasing costs of actions with negative externalities is an example of traditional intervention.

study by Anderson and de Palma (2013). When receivers deal with a message at least once, the sender obtains benefits which depend on the sender's exogenous type. The unit cost of sending a message is exogenous and identical. When the total number of messages is larger than the capacity, receivers allocate their attention to each message at random. Therefore, each message has a negative externality for the other active senders; and there are sometimes Pareto-ranked multiple equilibria for a given unit cost.

To assess whether the system converges to an efficient allocation after a policy intervention, we apply the *population game* framework (Sandholm 2010) along with evolutionary dynamics; in particular, we apply best response dynamics (Gilboa and Matsui 1991) to information congestion (Anderson and de Palma 2013). Our following discussion reflects the difficulty in controlling this primitive structure where senders send messages regardless of receivers' prior interests or actions.

Although a typical solution for congestion problems is to increase the cost of the action associated with the negative externality, an additional cost per message can have unintended consequences. This additional cost sometimes forces small-benefit-type senders to withdraw from the competition. Consequently, the probability of repetition increases if large-benefit-type senders continue to send multiple messages. Therefore, an additional cost sometimes hampers the information flow "efficiency"[5]. In addition, once the advertising competition intensifies, sometimes, any amount of additional cost cannot make the system converge to an efficient equilibrium (see Section 5.1). This negative result provides a new insight, unlike the discussion in previous studies[6], which supports placing an additional cost on sending messages when information congestion occurs. In other words, traditional static pricing, as discussed in the previous literature, sometimes does not work well and generates additional inefficiency, which we explain in the next paragraph.

To capture differences among the Pareto efficient equilibria, we assess system efficiency based on three types of social losses. First, if a message is not dealt with by receivers, the production cost for sending the message is wasted. Second, receivers' attention is wasted when no message is received; namely, their unutilized attention is wasted. Third, the repetition of messages is wasted for not only senders but also receivers. Receivers sometimes receive the same messages multiple times, and this repetition generates a negative externality for other senders. We aim to achieve a Pareto efficient equilibrium while minimizing

[5]This study focuses on the benefit of information transmission and ignores the benefit of repetition for senders (and receivers).

[6]See Van Zandt (2004), and Anderson and de Palma (2009, 2012, 2013).

these three losses.

Since the traditional one-shot intervention does not necessarily work well, we consider an idealized setting in which a policy-maker can put in place two new types of policies: a two-step cost change and a full-aid for first messages. We can always achieve a Pareto efficient equilibrium with almost minimum social loss by following two steps: (1) raising the senders' cost per message until sending no messages becomes the dominant strategy for all senders (and waiting for the senders' adjustment); (2) gradually reducing the senders' cost until congestion appears (or all senders send a message). Although the first policy works regardless of the initial condition and distribution of the senders' benefit, it requires a complicated process that may take time and financial resources. Full-aid for first messages, under which all senders send the first message without paying the cost, leads the system to a stable Pareto efficient equilibrium if there are enough potential senders[7]. Although this second policy is simpler than the first, and may be more realistic, it incurs a social loss of unreceived messages and always consumes financial resources. Although each policy is desirable in certain situations, their requirements limit their feasibility. A simpler policy may exist, but as far as we can deduce, in the primitive structure, costly policies are required to achieve the best allocation of receivers' (consumer) attention.

Both policies change the cost of actions regardless of the players' type. Additionally, under the best response dynamics, global convergence is guaranteed. These policies do not require policymakers to monitor private information and actions. These are the typical characteristics of *evolutionary implementation* for situations where players' actions are anonymous (Sandholm 2002).

However, the logic behind the two-step cost change is distinct from that of a traditional evolutionary implementation. In the traditional evolutionary implementation, thanks to exogenous information, before implementation begins, a policy-maker can fully understand the externality at each aggregate state[8], namely, the number of players who select each action. Thus, in the beginning, the policy-maker can create a complete path-independent policy for each aggregate state.

We do not require this complete knowledge. In the two-step cost change, senders' aggregate reactions gradually reveal the distribution of the benefit. This is because senders will continue to send messages only if the cost of sending the message is less than the

[7]If enough advertisers send a message, the pressure from the messages significantly reduces the marginal benefit of a second message. Accordingly, in the equilibrium, repetition disappears and Pareto efficient equilibrium appears.

[8]In the framework of the population game, we call a strategy distribution a (population) state.

exogenous benefit from the received message. Eventually, policymakers will understand the externality among players and will decide the timing of each step in the policy. Although the two-step cost change depends on the unique characteristics of information congestion, the underlying logic is more general; evolutionary implementation, in its process, can reveal extensive information, which can be used by policymakers to design a policy that achieves an efficient state[9]. We contribute to the field of evolutionary implementation by highlighting this novel advantage.

In this study, we utilize a regular Taylor evolutionarily stable state as a solution concept. This concept is stronger than the Nash equilibrium and has asymptotically local stability under many popular dynamics, including the best response and impartial pairwise comparison dynamics (Sandholm, 2010). The field of information congestion focuses on competition across industries, and thus, many advertisers play the game. As it is difficult for all advertisers to always select the optimal choice that keeps equilibrium, we focus on locally stable equilibria.

The rest of the study is organized as follows. In Section 2, we briefly review the associated literature and our contributions in the field. Section 3 formulates information congestion in the framework of the population game. Section 4 defines the criteria of local stability, regular Taylor evolutionarily stable state, and presents a simple example of (Pareto) suboptimal equilibrium with local stability. This example demonstrates the structural inefficiency and the importance of interventions. In Section 5, we propose the additional measure of efficiency and discuss a case where an additional cost impairs the efficiency. Therefore, proving that, naíve interventions do not always work. Section 6 proposes and analyzes the two policies achieving a Pareto efficient equilibrium with stability and discusses the cost of policies, such as time and budget. In Section 7, we compare our results with those of Anderson and de Palma (2013). Section 8 summarizes our results and implications.

2 Literature

In this section, we situate this paper within the economic analysis of advertising and related fields. Further, we explain our contribution to the field of information congestion.The

[9]Callander (2011) also discusses a similar situation in which the outcomes of each policy are uncertain, the set of feasible policies are uncountable and policymakers gradually learn the outcomes through trial and error. However, uncertainty follows a purely exogenous random motion (Brownian motion). In our discussion, policymakers gradually learn the externality in the game by observing aggregate-level endogenous reactions to the policy.

difference between information congestion and rational inattention is also explained.

In economic theory, advertising informs consumers *about the availability of new products* at a minimum (Renault 2015) and complements consumer search[10]. Despite information from other consumers, consumers sometimes fail to find their best choice without the aid of advertising[11]. This makes advertising an important element in the study of market competition.

The economic analysis of advertising usually focuses on the effect of advertising on market competition. In contrast, our study focuses on the process of advertising. Our study contributes to the field by supplying a useful tool for extending the mainstream discussions on the economic analysis of advertising. Bagwell (2007) offers an overview of advertising research in industrial organization. Bagwell proposes three advertising categories, namely, the persuasive, the informative, and the complementary perspectives of advertising. The differences in the categories are derived from the impact of advertising on consumer preferences (for the advertised goods). Most advertising literature analyzes the impact of advertising on market competition and can be classified using Bagwell's three categories. However, our study does not fall within these categories as it does not assume receivers' preferences. Therefore, we can put additional constraints on receivers' attention in those mainstream models without conflicting assumptions. For example, Van Zandt (2004) inspired Anderson and de Palma (2012), who analyzed the impact of consumers' limited attention on market competition.

Our discussion complements previous literature, focusing on advertising and consumers' limited attention, by analyzing an implementation problem with information congestion. Papers in the field put weight on either interesting equilibrium with free entry[12] or static pricing/policies[13]. However, as we see in Section 5.1, sometimes, any static pricing for sending a message cannot lead to an efficient allocation. To understand the performance of the primitive structure, we have to discuss implementation with dynamic interventions.

The field of rational inattention (RI) also examines cognitive limitations but focuses on different research targets. For example, Sims (2003) analyzes a decision maker who decides how much information to collect. There are two types of shocks: temporal mea-

[10]For example, see Stigler (1961) and Butters (1977). Renault (2015) reviews the literature relating to advertising and consumer search.

[11]For example, see Ali (2018) and Niehaus (2011).

[12]See Falkinger (2007, 2008) and Hefti (2018).

[13]See Van Zandt (2004), Anderson and de Palma (2009, 2012, 2013) and Anderson and Peitz (2022).

surement errors and long-term fundamental shocks. The decision maker can only partially distinguish between these two shocks because of capacity limitations on the accuracy of information. Therefore, the decision maker gradually adjusts the understanding of economic fundamentals, instead of accepting the latest information. There are two differences between information congestion and RI. First, in RI, decision makers are active and rational. They have a basic understanding of the economy and the distribution of errors. They completely control the messages they receive. In information congestion, receivers may not understand anything about the economy. Second, RI discusses long-term optimization. Information congestion examines the more myopic players. This difference is rooted in their respective research targets. RI aims to analyze serious problems for decision makers. Information congestion explains the situation in which the expected benefits for each player may be low.

3 Formulation and Equilibrium

This section introduces the information congestion model. In 3.1, we introduce the optimization problems of heterogeneous senders. In 3.2, the optimization problems are applied to the population game framework. In 3.3, we define the Nash equilibrium.

3.1 Formulation (Optimization Problem of Senders)

Information congestion models include message-senders and message-receivers. In this study, we consider a situation in which senders cannot distinguish among receivers and focus on a mass of homogeneous receivers. The receivers' volume is normalized to one. We consider a situation in which each receiver cannot adjust the capacity well. For example, if many flyers/spam mails are present in a mailbox, each receiver may unintentionally recognize the top one. The capacity for receivers to deal with messages is $\phi > 0$. If the total number of messages by senders, denoted by N, is more than the capacity (ϕ), messages are randomly dealt with by receivers. We apply this constant-capacity assumption to simplify the receivers' side and focus on senders' behaviors[14]. The presented basic model with dis-

[14]These assumptions are modestly consistent with some empirical literature in advertising research. For example, in Riebe and Dawes (2006), respondents recall 1.1 ads per three ads in low-clutter groups and 1.4 ads per nine ads in high-clutter groups. Hammer, Riebe, and Kennedy (2009) analyze four types of advertising data and conclude that the probability of recall for each is low/high in high/low-clutter situations. As long as an alternative function of ϕ/N is differentiable and de-

crete players is essentially identical to the basic model with two senders in Anderson and de Palma (2013).

First, we explain the basic model with discrete numbers of heterogeneous senders before introducing the population game with a continuous mass of players. Each sender i gains benefit π^i when all receivers receive at least one message from the sender. For example, if half of the receivers receive a message from sender i, the sender's benefit is $\pi^i/2$. We are interested in information transmission, and if a receiver deals with two or more messages from a sender, it is redundant in our model[15]. The cost of sending each message is γ. Senders decide their number of messages l by considering the tradeoff between the probability of getting attention and the cost of sending a message. Each sender i can select only discrete numbers $l \in \{0, 1, 2, ..., l_{max}\}$ as the number of messages, where l_{max} denotes the exogenous maximum number of messages for senders.

Suppose that there exist senders including a sender i and additional Q number of senders who send a single message. The optimization problem for sender i is

$$\max_{l^i \in \{0,1,2,...,l_{max}\}} U^i = \begin{cases} \pi^i(1-(1-\phi^j/(Q+l^i))^{l^i}) - \gamma l^i & \text{if } \phi^j < Q + l^i \\ \pi^i \mathbb{1}(l^i > 0) - \gamma l^i & \text{if } \phi^j \geqq Q + l^i \end{cases} \tag{1}$$

where $\mathbb{1}$ is the indicator function. When $\phi^j < Q + l^i$, by sending l messages, sender i gets the expected benefit of the message being received ($\pi^i(1-(1-\phi^j/(Q+l^i))^{l^i})$) and pays the cost of sending l messages (γl). When $\phi^j \geqq Q + l^i$, by sending at least one message ($l > 0$), sender i gets π and pays the cost of sending l messages (γl).

$(1-(1-\phi^j/(Q+l^i))^{l^i})$ in (1) is the probability that at least one message from sender i is received by each receiver. When we apply the population game framework, we assume that each of senders is small, so that they do not consider the impact of their message on the total number of messages, like $(1-(1-\phi^j/(Q))^{l^i})$ in this example. We consider advertising competition among many firms across industries, and this assumption would be reasonable. Anderson and de Palma (2013) explain that this assumption is similar to a standard monopolistic competition assumption where firms do not count the impact of their

creasing in N when $N > \phi$, we conjecture that our main results hold because the utility of each sender depends on ϕ/N and does not directly depend on ϕ.

[15] As Renault (2015) states, the minimum consensus in the economic analysis of advertising is that advertising communicates to consumers the availability of new products. We focus on this aspect of advertising.

actions on other firms.

As Anderson and de Palma (2013) point out that, when Q is large enough, the breakthrough probability $(1-(1-\phi^j/(Q+l^i))^{l^i})$ is decreasing in l^i; this characteristic is consistent with the shape of advertising response functions typically confirmed in empirical advertising research. For example, Taylor, Kennedy and Sharp (2009) confirm the decreasing marginal sales of advertising by analyzing data from four non-durable-goods categories [16]. Vakratsas et al. (2004) show the concave advertising response function with the probabilistic thresholds of advertising expenditure for attracting additional market share in durable goods markets. Our discussion supplies a theoretical explanation of these results.

3.2 Formulation (Population Game Framework)

$R=\{1,...,m\}$ is an exogenous set for heterogeneous senders. Each sender belongs to a single type $r\in R$, and senders with the same type r get the same benefit $\pi(r)>\gamma$ if their message is dealt with. The minimum benefit difference among senders is $MD=\min_{r,r'\in R}|\pi(r)-\pi(r')|$ and the senders' maximum benefit is $\pi_{max}=\max_{r\in R}\pi(r)$. The minimum benefit is $\pi_{min}=\min_{r\in R}\pi(r)$. The population of type $r\in R$ is denoted by $d(r)$. We assume that there exists a minimum basic unit $\omega>0$, and for any $r\in R$ there exists $j\in\mathbb{N}\cup\{0\}$ s.t. $d(r)=j\omega$[17]. The total population of senders is $D=\sum_{r\in R}d(r)$. x_l^r denotes the population of type r who selects strategy l. The set of available numbers of sending messages for type $r\in R$ is denoted by $L=\{0,1,2,...,l_{max}\}$.

n^r is the total number of pure strategies for type r and the total number of pure strategies for all types is $n=\sum_{r\in R}n^r$. The set of all possible states (strategy distributions) for type r is $X^r=\{x^r\in\mathbb{R}_+^{n^r}:\sum_{l\in L}x_l^r=d(r)\}$. The set of all possible states in the system (state space) is $X=\prod_{r\in R}X^r=\{\boldsymbol{x}=(x^0,...,x^r)\in\mathbb{R}_+^n:x^r\in X^r\}$. TX is the tangent space of X s.t. $TX=\prod_{r\in R}TX^r$ where $TX^r=\{z^r\in\mathbb{R}^{n^r}:\sum_{l\in L}z_l^r=0\}$. The total number of messages by all senders is $N(\boldsymbol{x})=\sum_{r\in R}\sum_{l\in L}lx_l^r$. Define a bijection $RS(r,l)$ s.t. $RS(r,l)\in\{1,...,n\}$ for any $r\in R$ and $l\in L$, $0<RS(r_1,l_0)<RS(r_2,l_0)\leqq n$ and $RS(r_1,l_1)<RS(r_1,l_2)$ if $r_1,r_2\in R$, $r_2>r_1$, $l_0,l_1,l_2\in L$ and $l_2>l_1$. For convenience, x_i represents a population with a tuple of type r and strategy l s.t. $i=RS(r,l)$ ($x_i=x_l^r$ for any $i\in\{1,2,...,n\}$, any type $r\in R$ and any strategy $l\in L$ if $i=RS(r,l)$).

[16]Taylor, Kennedy and Sharp (2009) report an exception, but the conclusion about the decreasing marginal sales of advertising typically holds.

[17]We use this assumption when we decide the amount of (finite) cost changes (later defined by γ_z') in policies.

For the following reasons, the population game framework matches with information congestion. First, because the timing of decision making among players does not have a solid structure, we can avoid a specific assumption about it. In population games, players sporadically and myopically[18] select their strategies. Second, some players may gradually learn the optimal level of advertising in each situation. Information congestion occurs in advertising competition across industries, and each sender may not understand the other players well. Thus, myopic decision making based on realized payoffs is reasonable for analyzing such a situation with a large number of players. Population games allow for various types of dynamics, and we find robust implications among these dynamics. Third, because the population game framework is used for equilibrium selection, we can use insights from the literature to find the important equilibria among multiple equilibria.

3.3 Definition of Equilibrium

Definition 1. *State $\boldsymbol{x} \in X$ is a Nash equilibrium in this information congestion game if and only if*

$$
\begin{aligned}
U(r,l,\boldsymbol{x}) &\leqq \bar{U}(r,\boldsymbol{x}) \\
x_l^r(\bar{U}(r,\boldsymbol{x}) - U(r,l,\boldsymbol{x})) &= 0 \qquad\qquad \forall r \in R \text{ and } \forall l \in L \\
x_l^r &\geqq 0
\end{aligned}
\tag{2}
$$

where

$$U(r,l,\boldsymbol{x}) = \pi(r)\min\Big(1, (1-(1-\phi/N(\boldsymbol{x}))^l)\Big) - \gamma l \tag{3}$$

$$\bar{U}(r,\boldsymbol{x}) = \frac{1}{d(r)} \sum_{l \in L} x_l^r U(r,l,\boldsymbol{x}). \tag{4}$$

Equation (2) indicates that, in equilibrium, all players select their best strategies to maximize their utility and each population cannot be negative. Equation (3) shows that, the utility depends on the benefits, the costs, and the probability that at least one message is dealt with by receivers. Since U is continuous in X, from Theorem 2.1.1. in Sandholm (2010), this *population game admits at least one Nash equilibrium.*

[18] Here, "sporadically" suggests that some portion of players can simultaneously change their strategies. "Myopically" suggests that all players make their decision by assuming that other players retain their strategies.

4 Stability of Equilibria

We use the Regular Taylor Evolutionarily Stable State (RTESS) as a criterion of stability. Sandholm (2010) shows that an equilibrium is locally asymptotically stable under many reasonable dynamics, such as the best response and any impartial pairwise comparison dynamics, if the equilibrium is an RTESS. Whether an equilibrium is an RTESS only depends on the payoff function of its neighborhood. Therefore, this concept enables us to discuss the stability of equilibria without assuming a specific dynamic.

From Sandholm(2010), the definition of an RTESS is as follows:

Definition 2. *State $\boldsymbol{x} \in X$ is a Regular Taylor ESS if the utility function U is Lipschitz continuous in X and the following two conditions are satisfied.*

$$\textit{For any } r \in R \textit{ and } i, j \in L,\ if\ x_i^r > 0 \textit{ and } x_j^r = 0,\ U(r, i, \boldsymbol{x}) = \bar{U}(r, \boldsymbol{x}) > U(r, j, \boldsymbol{x}). \tag{5}$$

$$\textit{For any } \boldsymbol{y} \in X - \{\boldsymbol{x}\},\ \ (\boldsymbol{y} - \boldsymbol{x})' DU(\boldsymbol{x})(\boldsymbol{y} - \boldsymbol{x}) < 0\ \ if\ (\boldsymbol{y} - \boldsymbol{x})' U(\boldsymbol{x}) = 0. \tag{6}$$

Here, $(\boldsymbol{y} - \boldsymbol{x})'$ is the transpose of the matrix $\boldsymbol{y} - \boldsymbol{x}$. DU is the derivative of the linear map such that $U(\boldsymbol{y}) = U(\boldsymbol{x}) + DU(\boldsymbol{x})(\boldsymbol{y} - \boldsymbol{x}) + o(\boldsymbol{y} - \boldsymbol{x})$ where $o(\boldsymbol{z})$ represents a function $h : TX \to \mathbb{R}^n$ s.t. $\lim_{\boldsymbol{z} \to \boldsymbol{0}} h(\boldsymbol{z})/|\boldsymbol{z}| = \boldsymbol{0}$,

$$DU(\boldsymbol{x}) = \begin{pmatrix} \frac{\partial U_1(\boldsymbol{x})}{\partial x_1} & \cdots & \frac{\partial U_1(\boldsymbol{x})}{\partial x_n} \\ \vdots & \vdots & \vdots \\ \frac{\partial U_n(\boldsymbol{x})}{\partial x_1} & \cdots & \frac{\partial U_n(\boldsymbol{x})}{\partial x_n} \end{pmatrix} \tag{7}$$

where U_i for $i \in \{1, ..., n\}$ represents the utility function of players with a tuple of type $r \in R$ and strategy $l \in L$ s.t. $i = RS(r, l)$. Thus, $U_i(\boldsymbol{x}) = U(r, l, \boldsymbol{x})$. $\frac{\partial U_i(\boldsymbol{x})}{\partial x_j}$ is the partial derivative of U_i with respect to x_j.

When $\boldsymbol{x}$ satisfies Equation (5), all optimal strategies are selected, and non-optimal strategies are not selected in $\boldsymbol{x}$. Equation (6) implies that, in the neighborhood of $\boldsymbol{x}$, a small deviation among optimal strategies will be penalized with utility loss. In other words, if a small number of players begin to follow another strategy profile $\boldsymbol{y}$, they are defeated by the majority who follow the strategy profile $\boldsymbol{x}$. These conditions are satisfied, for example, if each type of player has a unique type-specific optimal strategy. When a single type has multiple best replies in the equilibrium and all other types have a unique type-specific optimal strategy, and if each of the multiple best replies is selected and has a self-defeating externality, these conditions are satisfied.

4.1 Evolutionary Stability and Dynamics

Is the RTESS a good criterion for the stability of information congestion? The answer depends on whether dynamics covered by the RTESS are fit for competition among advertisers. As typical examples, an RTESS is locally asymptotically stable under any impartial pairwise comparison and best response dynamics (Theorem 8.4.7 in Sandholm (2010)).

Let us consider the case in which senders follow an impartial pairwise comparison dynamic. This dynamic indicates that, each of senders decides the number of messages while comparing the payoff from their current strategy and one with an alternative. If the payoff of the alternative is better than the current strategy, the player changes the current strategy to the alternative in positive probability. A dynamic following this rule is called pairwise comparison. For each type of player, if a unique function of the payoff difference decides the change to each strategy, this dynamic is called *impartial pairwise comparison*. If senders can perform a quantitative adjustment, this dynamic seems to be justified. In the context of advertising, advertisers can exercise quantitative adjustments at least through trial and error.

In addition, advertisers can estimate or observe the intensity of congestion (ϕ/N) in advertising. Thus, they can directly select the best strategy for the situation. This is called the *best response dynamic*. Therefore, we consider that both of these dynamics are valid, in the context of advertising.

Hereafter, we assume the best response dynamics because this dynamic allows a simple interpretation. Particularly, in Section 6, when we discuss the policy that utilizes the history of players' reactions as additional information for policymakers, our logic depends on the characteristic of this dynamic.

$t \in [0, \infty)$ denotes continuous time in the economy. At $t = 0$, an exogenous initial condition $\boldsymbol{x}(0) \in X$ exists, and the system follows the best response dynamics. $\boldsymbol{x}(t)$ denotes one of the possible states at t under the best response dynamics for a given initial condition $\boldsymbol{x}(0)$.

The best response dynamic for type r is given by[19]

Definition 3. *Best Response Dynamic:* $\dot{x}^r \in d(r)M^r(U(r, \boldsymbol{x})) - x^r$

where

$$M^r(U(r, \boldsymbol{x})) = \underset{y^r \in \Delta^r}{\operatorname{argmax}} (y^r)' U(r, \boldsymbol{x}) \tag{8}$$

is the maximizer correspondence for type r and

$$\Delta^r = \left\{ y^r \in \mathbb{R}_+^{n^r} : \sum_{i \in L} y_i^r = 1 \right\}. \tag{9}$$

is the set of mixed strategies for type r. When we state that the system follows the best response dynamics, each $r \in R$ follows the best response dynamic given above.

Consider $x_l^r(t)$ (the population of type-r senders who select strategy l at t) where $r \in R$ and $l \in L$. $x_l^r(\lim)$ denotes the limit of $x_l^r(t)$ ($x_l^r(t) \rightarrow x_l^r(\lim)$ as $t \rightarrow \infty$). When we say $x_l^r(t)$ is in ϵ-neighborhood of the limit $x_l^r(\lim)$, $|x_l^r(\lim) - x_l^r(t)| < \epsilon$. For this study, when $x_l^r(t)$ is in ϵ-neighborhood of the limit, and if all possible $x_l'^r(t')$ conditional on $x_l'^r(t) = x_l^r(t)$ are ϵ-neighborhood of the limit for any $t' > t$, we state that ϵ-convergence happens at t[20].

Our discussion in Section 6 frequently utilizes the following basic characteristic of the best response dynamics.

Lemma 1. *Consider $t2 > t1 \geqq 0$ and $\boldsymbol{x}(0) \in X$ and suppose that $\boldsymbol{x}(t)$ follows the best response dynamics. For each $t \in [t1, t2]$, if $\exists l' \in L$ s.t. $U(r, l, \boldsymbol{x}(t)) < U(r, l', \boldsymbol{x}(t))$, then the population of type r that selects the suboptimal strategy l is upper-bound ($x_l^r(t) \leqq De^{-t+t1}$) at t.*

Proof: When $l \in L$ is always strictly suboptimal at $t \in [t1, t2]$, $\dot{x}_l^r = -x_l^r \Leftrightarrow x_l^r(t) = x_l^r(t1)e^{-t+t1}$. Because D is the total number of senders, $x_l^r(t1) \leqq D$, and thus $x_l^r(t) \leqq De^{-t+t1}$.

[19]We follow the notation in Sandholm (2010). Gilboa and Matsui (1991) propose the best response dynamics and analyze the cyclically stable sets of strategy profiles ($\boldsymbol{x}$ in this project) that have a trajectory based on the best response dynamics from any other strategy profile in the set.

[20]Gilboa and Matsui (1991) introduce ϵ-accessibility with which the system with the best response dynamics can move to a strategy profile (state $\boldsymbol{x}$ in this study) from ϵ-neighborhood of another strategy profile. We discuss implementations and later define ϵ-convergence after which each possible trajectory $x_l^r(t')$ after $t' > t$ has a limit (and stays in its ϵ-neighborhood of the limit after t). ϵ-accessibility captures general movements like a circular movement.

This implies that, if one strategy is strictly better than any other strategies for a long time, the population that selects the suboptimal strategy eventually converges to 0. In addition, for each suboptimal strategy, we have a simple upper-bound $x_l^r(t) \leqq De^{-t+t1}$, which depends on the total population D and $t - t1$ (the pass time after the intervention begins). For general evolutionary dynamics, the existence of such an ϵ-convergence for a strictly best strategy in a certain period is not guaranteed[21].

However, if we assume a stronger requirement as specified below[22], a similar convergence occurs for the following family of dynamics.

Definition 4. *Impartial Pairwise Comparison Dynamic:*

$$\dot{x}_i^r = \sum_{l \in L} x_l^r \zeta_i^r (U_i^r(\boldsymbol{x}) - U_l^r(\boldsymbol{x})) - x_i^r \sum_{l \in L} \zeta_l^r (U_l^r(\boldsymbol{x}) - U_i^r(\boldsymbol{x})) \tag{10}$$

where $\zeta_l^r : \mathbb{R} \to \mathbb{R}_+$ is a Lipschitz continuous function s.t.

$$\mathrm{sgn}(\zeta_l^r(U_i^r(\boldsymbol{x}) - U_l^r(\boldsymbol{x}))) = \mathrm{sgn}(\max(U_i^r(\boldsymbol{x}) - U_l^r(\boldsymbol{x}), 0)), \tag{11}$$

and sgn(z) $= 1$ if $z > 0$, sgn(z) $= 0$ if $z = 0$(, and sgn(z) $= -1$ if $z < 0$).

Suppose that there exists a closed subset $X' \subseteq X$ s.t. $\boldsymbol{x}(t) \in X'$ for all $t \geqq t1$, and a strictly dominant strategy conditional on X' ($l \in L$) exists for type $r \in R$ over time after $t1$. Accordingly, under any impartial pairwise comparison dynamics, when such a conditional dominant strategy exists, the population with the dominant strategy strictly increases and the suboptimal strategies eventually vanish over time (and ϵ-convergence happens in finite time.). We assumes that $\boldsymbol{x}(t)$ follows impartial pairwise comparison dynamics with a specific ζ, and this assumption is valid until the end of this paragraph. As we assume above, l is a (conditional) strictly dominant strategy ($\exists q0 > 0$ s.t. $U(r, l, \boldsymbol{x}(t)) - q0 - U(r, i, \boldsymbol{x}(t)) > 0$ for any $t \in [t1, \infty)$ and for any $i \in L - \{l\}$). Since $(U(r, l, \boldsymbol{x}) - U(r, i, \boldsymbol{x}))$ is continuous in X, ζ_l^r is continuous in X. In addition, ζ_l^r preserves the sign so that $\zeta_l^r(U(r, l, \boldsymbol{x}(t)) - U(r, i, \boldsymbol{x}(t))) > 0$. Because l is strictly optimal for any $t \to \infty$, $x_l^r(t)$ is monoton-

[21] As a more general discussion, Bandhu and Lahkar (2022) discuss a family of evolutionary dynamics which satisfy the convergence for a strictly dominant strategy to analyze an evolutionary foundation of dominant strategy implementation.

[22] The requirement is as follows: there exists a closed subset $X' \subseteq X$ s.t. $\boldsymbol{x}(t) \in X'$ for all $t \geqq t1$ where $t1$ indicates the point at which a policy begins. Further, if a policy-maker makes a strategy strictly dominant for type r conditional on X' at $t1$, type-r senders eventually select the strategy after $t1$. In information congestion, because $U(r, l, \boldsymbol{x})$ depends on r, l and $N(x)$, it is not difficult to find such a closed subset.

ically increasing and bounded ($x_l^r \leqq d(r)$), which is why $x_l^r(t)$ converges to a point denoted by $x_l^r(\lim)$. Since X' is compact, we can find a minimum value $q > 0$ s.t. $\zeta_l^r(U(r,l,\boldsymbol{x}) - U(r,i,\boldsymbol{x})) \geqq q$ for all $\boldsymbol{x} \in X'$. Further, there exists the minimum positive portion (represented by q) that senders move from each strategy i to l over $t \in [t1, \infty)$. Therefore, if we keep the (conditional) strict optimality in $[t1, \infty)$, all senders with type r select the conditional strict dominant strategy l at the limit ($x_l^r(\lim) = d(r)$). (However, the speed of the convergence depends on γ and ζ.)

Our policy discussion in Section 5 focuses on the best response dynamics and does not discuss policies for more general dynamics. There are two reasons for this approach. First, our goal is to demonstrate the difficulty of controlling information congestion under evolutionary dynamics. Thus, it is sufficient to show that the control is difficult under the best response dynamics. Second, under impartial pairwise comparison dynamics and more general evolutionary dynamics, it is relatively difficult to find a simple and general interpretation/convergence on the senders' reactions to the policies. Our following results show that, at least under the specific dynamic, we can control information congestion if we spend time and other resources. There would exist more robust policies for general dynamics, but we save these for future research.

When we consider a sequence of numbers $z(t)$ that converges to a number z as $t \to \infty$, for any $\epsilon > 0$, we find $\bar{t} > 0$ s.t., if $t > \bar{t}$, $|z - z(t)| < \epsilon$. In other words, when policymakers keep a strategy strictly optimal for each type r, and if the dynamic makes all type-r players eventually select a strictly optimal strategy, policymakers are certain of the ϵ-convergence where almost all (at least $d(r) - \epsilon$) players select the best strategies in finite time. Since R is a finite set, we can find the maximum of required time for each type. $t(\epsilon)$ denotes the maximum time for such an ϵ-convergence. In the following discussion, we assume the best response dynamics, which is why policymakers know $t(\epsilon)$ for ϵ-convergence when there exists a strictly best strategy.

4.2 A Simple Example of Suboptimal Stable Equilibrium

Even in a simple setting, our model has multiple equilibria[23], including a suboptimal and locally stable equilibrium. For simplicity, we focus on the case of homogeneous senders ($R = \{1\}$) with the parameters s.t. $0.2 = \gamma/\pi$. This specific setting is sufficient to show the suboptimal stable equilibrium.

[23]This characteristic is consistent with Anderson and de Palma (2013).

Figure 1 shows the best responses of senders for each $\phi/N(\boldsymbol{x})$, which represents the probability that a unit message is received, with the restriction that the maximum number of sending messages is 5. (see the Appendix for details.) The set of optimal strategies for senders at given $\boldsymbol{x}$ is either a single strategy or the pair of two consecutive numbers[24] (see the Appendix for the proof.). Therefore, when the ratio of capacity to senders' population (ϕ/D) is given, we can calculate the possible range of ϕ/N for each pair of strategies. We combine the ranges and Figure 1 to derive Figure 2.

The curve in Figure 2 shows the equilibria for each γ/π. The orange rectangle represents $0.2 = \gamma/\pi$. Each point (A and C) is an RTESS[25] (see the Appendix for the proof.). The arrows in Figure 2 explain the direction of the dynamics at ϕ/N when we apply the best response dynamics to the specific state $\boldsymbol{x}$ in which $N = N(\boldsymbol{x})$ and all senders select the strategies that use two consecutive numbers like sending a message once or twice. When we consider a point above the curve, the dynamics gradually reduce N. When we consider a point below the curve, the dynamics gradually increase N. The Appendix provides the details about the stability and the derivation of Figure 2.

To define Pareto efficiency for a continuum of players, we introduce the following distribution of senders' utility in $\boldsymbol{x} \in X$. For any $r \in R$ and any $j \in \mathbb{R}$, we define

$$F^r(j, \boldsymbol{x}) = \sum_{l \in L} x_l^r \mathbb{1}(U(r, l, \boldsymbol{x}) \leqq j). \tag{12}$$

$F^r(j, \boldsymbol{x})$ is the number of type r senders whose utility does not exceed j in $\boldsymbol{x}$. Since each player in each type is anonymous, for $\boldsymbol{x}, \boldsymbol{x}' \in X$, when $F^r(j, \boldsymbol{x}) \geqq F^r(j, \boldsymbol{x}')$ for all $j \in \mathbb{R}$, and if there exists $j' \in \mathbb{R}$ s.t. $F^r(j', \boldsymbol{x}) > F^r(j', \boldsymbol{x}')$, we consider that $\boldsymbol{x}$ is inferior to $\boldsymbol{x}'$ for senders in type r.

Using F^r above, in this study, we define a Pareto efficient state as follows:

Definition 5. *$\boldsymbol{x} \in X$ is Pareto efficient if and only if there does not exist $\boldsymbol{x}' \in X$ s.t., for any $r \in R$ and for any $j \in \mathbb{R}$,*

$$F^r(j, \boldsymbol{x}) \geqq F^r(j, \boldsymbol{x'}) \tag{13}$$

[24] In this study, we define that consecutive numbers are the sequence of (non-negative) integers like 0,1,2,3... without intervals. When we say two consecutive numbers, we imply a pair of two integers such as (0,1), (1,2) and (3,4).

[25] The circle (B) represents an equilibrium without stability.

and there exists at least a pair $r' \in R$ and $j \in \mathbb{R}$ s.t.

$$F^{r'}(j, \boldsymbol{x}) > F^{r'}(j, \boldsymbol{x'}). \tag{14}$$

When $\boldsymbol{x} \in X$ satisfies these conditions in the definition, any $x' \in X$ cannot enable any portion of players to be better off, while maintaining the utility of other players. If $\boldsymbol{x} \in X$ does not satisfy these conditions, there exists $\boldsymbol{x'} \in X$ such that a portion of players who attain better utility, while maintaining the others' utility. Therefore, we define a Pareto efficient state as $\boldsymbol{x} \in X$ satisfying these conditions. This Pareto efficient state discusses efficiency conditional on the information-congestion games. Later, in Section 4.2, we relax this restriction on resource allocation and define another stronger concept (Pareto efficient resource allocation) than the Pareto efficient state.

In equilibrium C in the figure, all senders send their messages twice, and the probability that each receiver deals with these messages is $3/4$. This is Pareto inefficient compared with equilibrium A in which all senders send their messages only once, and all messages are dealt with by all receivers. Therefore, we derive the following proposition.

Proposition 1. *A (Pareto) suboptimal equilibrium is sometimes a regular Taylor evolutionarily stable state.*

This section demonstrates that the model has multiple equilibria, and the suboptimal equilibrium is sometimes locally stable. In other words, the primitive structure generates suboptimal outcomes. Therefore, interventions may improve "efficiency" for this economy. We have used Pareto efficiency as the measure of efficiency up to this point. However, other measures may fit this model. The next section proposes additional measures of efficiency inspired by transportation research and shows that the additional cost sometimes impairs the efficiency of the system.

5 Measure of Information Congestion

5.1 Social Loss in Information Congestion

Advertising has several similarities with transportation. The receivers' resources to process advertisements are limited and not storable, just as the road capacity for processing traffic demand is limited and not storable. Because both demands are derivative, extrinsic factors strongly influence their demand. In addition, there is a negative externality among users through congestion. Thus, it is difficult to achieve efficient allocation.

The traffic assignment in transportation networks with a given demand has been one of the most important topics in transportation research. As Wardrop (1952) selects, models of traffic networks often treat the total travel time of travelers as the measure of efficiency. Although this approach cannot maximize social welfare in general, this approach has an advantage. In this approach, phenomena can be analyzed without a demand model. Since the demand for traffic networks is usually derivative, the traffic demand model is sensitive to outside factors. The time evaluation approach avoids this sensitivity. In addition, it is difficult to measure utility from travel behaviors because of the derivative demand. Thus, in transportation research, instead of maximizing social welfare, many studies aim to minimize total travel time representing the social cost of transportation.

Instead of travel time, we focus on the three social losses in the model. First, when the total number of messages is insufficient compared with the limited attention ($\phi > N(\boldsymbol{x})$), a part of the receivers' attention is left unutilized. Second, if the total number of messages exceeds the receivers' limited attention ($\phi < N(\boldsymbol{x})$), a part of the messages are wasted. Third, if receivers receive multiple messages from the same sender, it harms social welfare because the other senders lose the opportunity of captivating attention, the production cost of sending/processing a message is typically positive, and the original sender does not receive any additional benefit from the duplication. Therefore, the social loss on senders' side is the number of wasted messages such that

$$\begin{aligned} Loss^S/\gamma =& (N(\boldsymbol{x}) - \phi)\mathbb{1}(N(\boldsymbol{x}) > \phi) \\ &+ \sum_{r\in R}\sum_{l\in L_{-1}} x_l^r \left(\sum_{h\in L_{-1} s.t.\ h\leqq l} \frac{l!(h-1)}{h!(l-h)!} \min\left(1, \left(\frac{\phi}{N(\boldsymbol{x})}\right)^h \left(1-\frac{\phi}{N(\boldsymbol{x})}\right)^{l-h}\right)\right). \end{aligned} \tag{15}$$

where $L_{-1} = \{l \in L | l > 1\}$ and, for simplicity, we assume that γ, the private cost of sending a unit amount of messages, is identical to the cost for the society.

This measure (15) counts both the excess number of messages, $N(\boldsymbol{x}) - \phi$, and the duplication of messages (the residual part) if $N(\boldsymbol{x}) > \phi$. Otherwise, the duplication of the messages is counted. The part in the large parentheses after x_l^r represents the expected number of duplicated messages by the type-r sender sending l messages. In particular, $(h-1)$ is the number of duplications if h messages are received and $\frac{l!}{h!(l-h)!}$ is the number of the combinations that h messages are received in l messages.

The social loss on the receivers' side is the amount of unutilized capacity such that

$$Loss^R/c = (\phi - N(\boldsymbol{x}))\mathbb{1}(N(\boldsymbol{x}) < \phi) + \sum_{r \in R} \sum_{l \in L_{-1}} x_l^r \left(\sum_{h \in L_{-1} s.t.\ h \leqq l} \frac{l!(h-1)}{h!(l-h)!} \min\left(1, \left(\frac{\phi}{N(\boldsymbol{x})}\right)^h \left(1 - \frac{\phi}{N(\boldsymbol{x})}\right)^{l-h}\right)\right). \tag{16}$$

where c is the social value of receivers' attention per unit. Similar to the social loss on senders' side, the duplication of messages is counted. In addition, if $N(\boldsymbol{x}) < \phi$, the excess capacity, $\phi - N(\boldsymbol{x})$, is added.

Thus, the total social loss is

$$Loss^T = Loss^S + Loss^R. \tag{17}$$

These measures are consistent with the discussions in previous literature. For instance, in Anderson and de Palma (2013), increasing the unit cost by a specific amount supports the existence of the efficient[26] equilibrium under certain conditions. In the efficient equilibrium, all active senders send a message, and the total number of messages is smaller than the capacity. Thus, the social loss on the senders' side is 0 in this efficient equilibrium.

Using these measures, we can show that an additional cost per message sometimes increases social loss, as in the following example. Consider two types of senders $R = \{r_1, r_2\}$, $l_{max} = 2$, $\gamma = 0.9$, $\phi = 1$, $d(r_1) = 0.6$, $d(r_2) = 0.4$, $\pi(r_1) = 100$ and $\pi(r_2) = 1.6$. If $x_2^{r_1} = d(r_1)$ and $x_1^{r_2} = d(r_2)$ at $t = 0$, $\frac{\phi}{N(\boldsymbol{x}(0))} = 5/8$ and $\boldsymbol{x}(0)$ is a RTESS. If we change the cost from $\gamma = 0.9$ to $\gamma' = 1.6$, and all players follow the best response dynamics, r_1 senders keep $l = 2$ but r_2 senders eventually select $l = 0$. At the limit, $x_2^{r_1} = d(r_1)$ and $x_0^{r_2} = d(r_2)$. In this change, the social loss on senders' side decreases, but the social loss on receivers' side increases from $d(r_1)25/64$ to $d(r_1)25/36$. Therefore, the effect on the total social loss depends on c and γ, and the total social loss can increase with the additional cost.

In addition, in this example, since the capacity is enough for all senders, there is a single Pareto efficient equilibrium where $x_1^{r_1} = d(r_1)$ and $x_1^{r_2} = d(r_2)$. However, any γ' (one-

[26]Here, "efficient" means that raising cost invites Pareto improvement without transfer and thus the new equilibrium is Pareto superior to the original equilibrium.

shot cost change) cannot achieve an evolutionary process which converges to the efficient equilibrium. (If $1.6 > \gamma' > 1.0$, a similar consequence happens. If $1.0 > \gamma'$, the congestion remains. If $\gamma' > 1.6$, r_2 senders completely exit the market.) This points to the limitation of the traditional one-shot intervention for information congestion.

Proposition 2. *Additional cost per message sometimes increases social loss of receivers' attention. When the cost per attention is larger than the cost per message, the additional cost per message sometimes increases the total social loss. In addition, sometimes, any additional cost per message cannot lead the system to a Pareto efficient state.*

5.2 Pareto Efficient Allocation

The social loss on receivers' side represents a fundamental efficiency in the allocation of receivers' capacity ϕ among active senders. To consider a more flexible allocation of the receivers' capacity ϕ, this subsection considers a subgroup in each population who belongs to $r \in R$ and selects $l \in L$. Each allocation Φ has a finite index set of the senders' subgroup denoted by E_Φ. For any $\boldsymbol{x} \in X$ and for any $r \in R$ and $l \in L$, we require that the amount of each finite subgroup $x^r_{l,e}$ satisfies $\sum_{e \in E_\Phi} x^r_{l,e} = x^r_l$. For any $r \in R$, $l \in L$ and $e \in E_\Phi$, $\Phi(r, l, e) \in \mathbb{R}_+$ denotes the amount of ϕ assigned to each sender in $x^r_{l,e}$. An allocation Φ is feasible on $\boldsymbol{x}$ if and only if $\sum_{r \in R} \sum_{l \in L} \sum_{e \in E_\Phi} x^r_{l,e} \Phi(r, l, e) \leqq \phi$, $\Phi(r, l, e) = 0$ for any $r \in R$, $l \in L$ $e \in E_\Phi$ s.t. $x^r_{l,e} = 0$ and $\Phi(r, 0, e) = 0$ for all $r \in R$ and $e \in E_\Phi$. This Φ represents any feasible allocation of ϕ to active senders in $\boldsymbol{x}$. The set of all feasible pairs of Φ and $\boldsymbol{x} \in X$ represents the potential achievement for a policy-maker, if the policy-maker knows the benefit distribution and can freely control all resources (and actions) in this economy.

Instead of the random assignment in information congestion, if we consider any arbitrary feasible allocation Φ on $\boldsymbol{x}$, the utility of $r \in R$ senders with $l \in L$ in the subgroup $e \in E_\Phi$ is $V(r, l, e, \Phi) = \pi(r) \min(1, \Phi(r, l, e)) - \gamma l$. For any $j \in \mathbb{R}$, we define

$$G^r(j, \Phi, \boldsymbol{x}) = \sum_{l \in L} \sum_{e \in E_\Phi} x^r_l \mathbb{1}(V(r, l, e, \Phi) \leqq j). \tag{18}$$

Further, we define a Pareto efficient allocation as follows:

Definition 6. *A feasible allocation Φ on $\boldsymbol{x} \in X$ is a Pareto efficient allocation if and only if there does not exist any feasible allocation Φ' on $\boldsymbol{x}' \in X$ s.t., for any $r \in R$ and for any*

$j \in \mathbb{R}$,

$$G^r(j, \Phi, \boldsymbol{x}) \geqq G^r(j, \Phi', \boldsymbol{x'}) \tag{19}$$

and there exists at least a pair $r' \in R$ and $j \in \mathbb{R}$ s.t.

$$G^{r'}(j, \Phi, \boldsymbol{x}) > G^{r'}(j, \Phi', \boldsymbol{x'}). \tag{20}$$

Hereafter, when a feasible Φ' on $\boldsymbol{x}'$ satisfies both (19) and (20) for Φ on $\boldsymbol{x}$ in the definition, we say that Φ' on $\boldsymbol{x'}$ (Pareto) dominates Φ on $\boldsymbol{x}$. For a given $r \in R$, when both (19) and (20) are satisfied, we state that $G^r(\Phi', \boldsymbol{x'})$ dominates $G^r(\Phi, \boldsymbol{x})$.

Definition 7. *An allocation Φ is induced by any $\boldsymbol{x} \in X$ in information congestion if and only if, for any $r \in R$, for any $l \in L$ and $e \in E_\Phi$, $\Phi(r, l, e) = l$ when $x^r_{l,e} > 0$ and $\phi \geqq N(\boldsymbol{x})$, $\Phi(r, l, e) = l\frac{\phi}{N(\boldsymbol{x})}$[27] when $x^r_{l,e} > 0$ and $\phi < N(\boldsymbol{x})$, and $\Phi(r, l, e) = 0$ otherwise.*

The allocation Φ induced by any $\boldsymbol{x} \in X$ does not require any subgroup because information congestion uniformly treats all senders belonging to $r \in R$ and selecting $l \in L$. For simplicity, hereafter, we assume that $E = \{1\}$ for any induced Φ, and we skip the notation for the subgroup if the abbreviation is clear.

In addition, hereafter, when we state an allocation Φ induced by any $\boldsymbol{x} \in X$ in information congestion, we always mean that the allocation Φ is on $\boldsymbol{x}$ and skip the statement.

Lemma 2. *An allocation Φ induced by any $\boldsymbol{x} \in X$ in information congestion is feasible (on $\boldsymbol{x}$).*

Proof: When $\phi \geqq N(\boldsymbol{x})$, $\sum_{r\in R}\sum_{l\in L} x^r_l \Phi(r, l) = N(\boldsymbol{x}) \leqq \phi$. When $\phi < N(\boldsymbol{x})$, $\sum_{r\in R}\sum_{l\in L s.t. l>0} x^r_l \Phi(r, l) = \phi$. In addition, $\Phi(r, l) = 0$ for any $r \in R$, $l \in L$ s.t. $x^r_l = 0$ and $\Phi(r, 0) = 0$ for all $r \in R$. Therefore, Φ is feasible on $\boldsymbol{x}$.

In addition,

Lemma 3. *When an allocation Φ is induced by any $\boldsymbol{x} \in X$ in information congestion, $G^r(j, \Phi, \boldsymbol{x}) = F^r(j, \boldsymbol{x})$ for any $j \in \mathbb{R}$ and for any $r \in R$.*

Proof: For any $r \in R$, $l \in L$, and $e \in E_\Phi$ s.t. $x^r_{l,e} > 0$, $V(r, l, e, \Phi) = \pi(r)\Phi(r, e, l) - \gamma l = U(r, l, \boldsymbol{x})$. Therefore, $G^r(j, \Phi, \boldsymbol{x}) = F^r(j, \boldsymbol{x})$ for any $j \in \mathbb{R}$ and $r \in R$.

Therefore,

[27] Since the expected number of messages being received is an expected value of a binomial distribution, $\Phi(r, l) = \sum_{h\in L s.t. h\leqq l} h\frac{l!}{h!(l-h)!}\left(\frac{\phi}{N(\boldsymbol{x})}\right)^h\left(1 - \frac{\phi}{N(\boldsymbol{x})}\right)^{l-h} = l\frac{\phi}{N(\boldsymbol{x})}$

Lemma 4. *If an allocation Φ induced by any $\boldsymbol{x} \in X$ in information congestion is Pareto efficient, $\boldsymbol{x}$ is a Pareto efficient state.*

Proof: From the definition, if Φ derived from $\boldsymbol{x}$ in information congestion is Pareto efficient, any pair of $\boldsymbol{x}'' \in X$ and any feasible Φ'' on $\boldsymbol{x}''$ cannot dominate Φ on $\boldsymbol{x}$. The set of Φ'' on $\boldsymbol{x}''$ includes Φ'' induced by any $\boldsymbol{x}'' \in X$ in information congestion. Thus, Lemma 3 implies that $\boldsymbol{x}$ is not dominated by any $x'' \in X$.

From Lemma 4, the Pareto efficient allocation is a stronger concept than the Pareto efficient state. This may be trivial because the latter concept is efficient conditional on the random assignment in information congestion.

We want to mention several characteristics of the Pareto efficient allocation. We claim the following lemma:

Lemma 5. *Consider a feasible Φ' on $\boldsymbol{x} \in X$. Φ' on $\boldsymbol{x}$ is not Pareto efficient if both of the following conditions are satisfied:*

1. $\sum_{r \in R} \sum_{e \in E_{\Phi'}} \sum_{l \in L} x_l^r \Phi'(r, l, e) = \phi - v$ *where* $v > 0$.
2. *There exist* $r \in R$, $l \in L - \{0\}$ *and* $e \in E_{\Phi'}$ *s.t* $1 > \Phi'(r, l, e)$ *and* $x_{l,e}^r > 0$.

Proof: We can find a feasible allocation on $\boldsymbol{x}$ which dominates Φ' on $\boldsymbol{x}$. Consider an allocation Φ on $\boldsymbol{x}$ with $E_\Phi = E_{\Phi'}$ satisfying the following two conditions: 1. $\Phi(r, l, e) = \Phi'(r, l, e) + v/x_{l,e}^r$ and 2. $\Phi(r', l', e') = \Phi'(r', l', e')$ for all $r' \in R$, $l' \in L$ and $e' \in E_\Phi$ if $(r', l', e') \neq (r, l, e)$. This Φ is feasible on $\boldsymbol{x}$ and $G^r(j, \Phi', \boldsymbol{x}) - G^r(j, \Phi, \boldsymbol{x}) > 0$ when $V(r, l, e, \Phi') < j < V(r, l, e, \Phi)$. For any other parameters, the difference of G is 0. Therefore, Φ on $\boldsymbol{x}$ dominates Φ' on $\boldsymbol{x}$.

From Lemma 5, Pareto efficient allocation requires that the capacity ϕ is fully utilized conditional on $\boldsymbol{x}$. In addition,

Lemma 6. *Consider a feasible Φ' on $\tilde{\boldsymbol{x}} \in X$ s.t. $\Phi'(r'', l'', e'') > 0$ for a combination of $r'' \in R$, $e'' \in E_{\Phi'} = \{1, ..., k\}$ and $l'' \in L$ s.t. $l'' > 1$ where k is a positive integer. Then, Φ' on $\tilde{\boldsymbol{x}}$ is dominated by a feasible allocation Φ on $\boldsymbol{x} \in X$ and $E_\Phi = E_{\Phi'} + \{k+1\}$ s.t.*

1. $x_{l,e}^r = \tilde{x}_{l,e}^r$ *and* $\Phi(r, l, e) = \Phi'(r, l, e)$ *for any* $r \in R$, $l \in L$ *and* $e \in E_{\Phi'}$ *s.t.* $(r, l, e) \neq (r'', l'', e'')$
2. $x_{1,k+1}^{r''} = \tilde{x}_{l'',e''}^{r''}$ *and* $\Phi(r'', 1, k+1) = \Phi'(r'', l'', e'')$
3. $x_{l,e}^r = 0$ *and* $\Phi(r, l, e) = 0$ *otherwise.*

Proof: Φ is feasible on $\boldsymbol{x}$ because of condition 2. Since $V(r'', 1, k+1, \Phi) - V(r'', l'', e'', \Phi') = \gamma(l''-1)$, $G^{r''}(j, \Phi, \boldsymbol{x}) - G^{r''}(j, \Phi', \tilde{\boldsymbol{x}}) < 0$ when $j \in (V(r'', l'', e'', \Phi'), V(r'', 1, k+1, \Phi))$. For any other parameters, the difference of G is 0. Therefore, Φ on $\boldsymbol{x}$ dominates Φ' on $\tilde{\boldsymbol{x}}$.

From Lemma 6,

Lemma 7. *A feasible allocation Φ on $\boldsymbol{x} \in X$ is Pareto efficient only if $\Phi(r, l, e) = \boldsymbol{x}^r_{l,e} = 0$ for any $r \in R$, for any $e \in E_\Phi$ and for any $l \in L$ s.t. $l > 1$.*

In other words, any duplication of messages is not allowed in any Pareto efficient allocation. In addition,

Lemma 8. *Consider $r \in R$ and feasible allocations Φ on $\boldsymbol{x} \in X$ and Φ' on $\tilde{\boldsymbol{x}} \in X$ satisfying the following conditions:*

1. $\sum_{e \in E_\Phi} x^r_{1,e} \Phi(r, 1, e) \geqq \sum_{e' \in E_{\Phi'}} \tilde{x}^r_{1,e'} \Phi'(r, 1, e')$.
   For any $e \in E_\Phi$,
2. $1 \geqq \Phi(r, 1, e)$,
3. $V(r, 1, e, \Phi) > 0$ *when* $\Phi(r, 1, e) > 0$ *and*
4. $x^r_{l,e} = 0$ *for any* $l \in L$ *s.t.* $l > 1$.

Then, either $G^r(j, \Phi, \boldsymbol{x}) - G^r(j, \Phi', \tilde{\boldsymbol{x}}) \leqq 0$ for any $j \in \mathbb{R}$ or there exist $j', j'' \in \mathbb{R}$ s.t. $G^r(j', \Phi, \boldsymbol{x}) - G^r(j', \Phi', \tilde{\boldsymbol{x}}) > 0$ and $G^r(j'', \Phi, \boldsymbol{x}) - G^r(j'', \Phi', \tilde{\boldsymbol{x}}) < 0$.

Proof: For the sake of contradiction, we assume that $G^r(j, \Phi, \boldsymbol{x}) - G^r(j, \Phi', \tilde{\boldsymbol{x}}) \geqq 0$ for any $j \in \mathbb{R}$ and there exists $j' \in \mathbb{R}$ s.t. $G^r(j', \Phi, \boldsymbol{x}) - G^r(j', \Phi', \tilde{\boldsymbol{x}}) > 0$. Without loss of generality, we assume that $\tilde{x}^r_{l,e} = 0$[28] for any $e' \in E_{\Phi'}$ and $l \in L$ s.t. $l > 1$. Since $V(r, 0, e, \Phi) = V(r, 0, e', \Phi') = 0$ and condition 4, the initial assumption implies that $\sum_{e \in E_\Phi} x^r_{1,e} V(r, 1, e, \Phi) < \sum_{e' \in E_{\Phi'}} \tilde{x}^r_{1,e'} V(r, 1, e', \Phi')$. However, since $V(r, 1, e, \Phi) = \pi(r) \min(1, \Phi(r, 1, e)) - \gamma$ and $1 \geqq \Phi(r, 1, e)$, $V(r, 1, e, \Phi) = \pi(r)\Phi(r, 1, e) - \gamma$ and so $\sum_{l \in L} \sum_{e \in E_\Phi} x^r_{l,e} V(r, l, e, \Phi) = \pi(r) \sum_{e \in E_\Phi} x^r_{1,e} \Phi(r, 1, e) - \sum_{e \in E_\Phi} x^r_{1,e} \gamma$. Then, because of condition 1, $\pi(r) \sum_{e \in E_\Phi} x^r_{1,e} \Phi(r, 1, e) \geqq \pi(r) \sum_{e' \in E_{\Phi'}} \tilde{x}^r_{1,e'} \min(1, \Phi(r, 1, e'))$, and so $-\sum_{e \in E_\Phi} x^r_{1,e} \gamma < -\sum_{e' \in E'_\Phi} \tilde{x}^r_{1,e'} \gamma$, and thus $x^r_1 > \tilde{x}^r_1$. However, from condition 3 and the initial assumption in this proof, $x^r_0 \geqq \tilde{x}^r_0$, which is why $x^r_1 \leqq \tilde{x}^r_1$. We get the contradiction.

[28]Because of the transitivity of the dominance in G^r, there always exists an alternative pair Φ'' and $\boldsymbol{x}''$ which is Pareto superior to Φ' on $\tilde{\boldsymbol{x}} \in X$ if $\tilde{x}^r_{l,e} > 0$ for any $l > 1$.

Lemma 8 implies that, if a feasible allocation Φ on $\boldsymbol{x} \in X$ satisfies the four conditions, and if we cannot allocate additional ϕ to type r, then $G^r(\Phi, \boldsymbol{x})$ would not be dominated by any $G^r(\Phi', \boldsymbol{x}')$ where Φ' is any feasible allocation on any $\boldsymbol{x}' \in X$ and satisfies condition 1. By using this result, we obtain the following:

Lemma 9. *A feasible allocation Φ on $\boldsymbol{x}$ is Pareto efficient if all of the following conditions are satisfied.*

1. $\sum_{r\in R}\sum_{e\in E_\Phi} x_1^r \Phi(r,1,e) = \phi$.
   For any $r \in R$ and $e \in E_\Phi$,
2. $1 \geqq \Phi(r,1,e)$,
3. $V(r,1,e,\Phi) > 0$ *when* $\Phi(r,1,e) > 0$ *and*
4. $x_{l,e}^r = 0$ *for any $l \in L$ s.t. $l > 1$.*

Proof: The conditions 2,3, and 4 in Lemma 9 and Lemma 8 imply that, for any $r \in R$, without any additional ϕ to r, we cannot find any feasible allocation Φ' on a state $\boldsymbol{x}'$ s.t. $G^r(\Phi', \boldsymbol{x}')$ dominates $G^r(\Phi, \boldsymbol{x})$. However, there is no additional amount of ϕ, because of condition 1. Therefore, Φ on $\boldsymbol{x}$ is a Pareto efficient allocation.

Lemma 10. *A feasible allocation Φ on $\boldsymbol{x}$ is Pareto efficient conditional on $x_0^r = 0$ for any $r \in R$ if all of the following conditions are satisfied.*

1. $\sum_{r\in R}\sum_{e\in E_\Phi} x_1^r \Phi(r,1,e) = \phi$.
   For any $r \in R$ and $e \in E_\Phi$
2. $1 \geqq \Phi(r,1,e)$ *and*
3. $x_{l,e}^r = 0$ *for any $l \in L$ s.t. $l > 1$.*

Proof: Because of the transitivity, if there is no Pareto efficient allocation that dominates Φ on $\boldsymbol{x}$, the allocation Φ on $\boldsymbol{x}$ is Pareto efficient. Because of Lemma 7 and the constraint $x_0^r = 0$, any Pareto efficient allocation Φ' on $\tilde{\boldsymbol{x}}$ satisfies $\tilde{x}_1^r = d(r)$ for any $r \in R$. For the sake of contradiction, Φ' on $\tilde{\boldsymbol{x}}$ dominates Φ on $\boldsymbol{x}$. Thus, for any $r \in R$, $G^r(j, \Phi, \boldsymbol{x}) - G^r(j, \Phi', \tilde{\boldsymbol{x}}) \geqq 0$ for any $j \in \mathbb{R}$ and there exist $r' \in R$ and $j' \in \mathbb{R}$ s.t. $G^{r'}(j', \Phi, \boldsymbol{x}) - G^{r'}(j', \Phi', \tilde{\boldsymbol{x}}) > 0$. However, this implies that $\sum_{e\in E_\Phi} x_{1,e}^{r'} V(r',1,e,\Phi) < \sum_{e'\in E_{\Phi'}} x_{1,e'}^{r'} V(r',1,e',\Phi')$ and, because of condition 2, $\sum_{e\in E_\Phi} x_{1,e}^{r'} \Phi(r',1,e) < \sum_{e'\in E_{\Phi'}} \tilde{x}_{1,e'}^{r'} \min(1, \Phi'(r',1,e')$

For $r \in R-\{r'\}$, $\sum_{e \in E_\Phi} x^{r'}_{1,e}\Phi(r', 1, e) \leqq \sum_{e' \in E_{\Phi'}} \tilde{x}^{r'}_{1,e'} \min(1, \Phi'(r', 1, e'))$. Further, because of condition 1, $\sum_{r \in R} \sum_{e' \in E_{\Phi'}} x^r_1 \Phi'(r, 1, e') > \phi$. Therefore, Φ' on $\tilde{\boldsymbol{x}}$ is not feasible. We obtain the contradiction because our definition of Pareto efficient allocation requires feasibility.

In this section, we introduce measures of efficiency to capture the social loss in information congestion. Afterwards, we check how an additional cost sometimes impairs efficiency. We introduce and analyze a Pareto efficient allocation where the receivers' attention is efficiently allocated among active senders by a policy-maker. In Section 6, instead of an additional arbitrary cost, we discuss desirable policies by utilizing the basic characteristics of Pareto efficient allocation.

6 Policies: Cost Change and Aid

We want to achieve a stable Pareto efficient equilibrium while minimizing the social loss defined in the last section. We assume that the policy-maker knows the population of senders D and capacity ϕ, minimum benefit difference MD, upper-bound of the required time for ϵ-convergence $t(\epsilon)$, minimum basic unit of population ω, and the feasible set of senders' benefits including maximum/minimum benefits $\pi_{max},\ \pi_{min}$[29]. In addition, we assume that the policy-maker can observe $N(\boldsymbol{x}(t))$, and can change the cost of sending a message γ. We denote the zth changed unit cost as γ'_z where z is an element of a finite set $\{$1,2...$\}$.

When we want to minimize the social losses described in the last section, the duplication of messages matters. As seen in Figure 1, there exist two extreme situations ($\phi/N \to 1$ and $\phi/N \to 0$) where no senders have incentive to send multiple messages. Therefore, a policy which invites either of the cases is promising. Through this section, we discuss properties of two such policies, namely, *two-step cost change* and *full-aid for first messages*.

In both policies, we utilize several characteristics of the best response dynamics. First, players move from a current strategy to another strategy if the payoff from another strategy is strictly better than the one from the current strategy. Second, because of the first characteristic, if the strictly dominant strategy exists for a type of player, the strategy is eventually selected by all players within the type. Third, conditional on playing dominant strategies, if another type has a strategy that strictly dominates all other strategies, this conditional dominant strategy is eventually selected by all players within the type. Further, from any original

[29] Instead of $t(\epsilon)$, ω, MD, policymakers can use exaggerated numbers like $MD > MD'$, which is why we consider that this requirement is relatively weak.

state, we can move to the state in which those types select the strictly dominant strategy or the conditional dominant strategy. If this realized state is the desirable state, or if we can find the policy from the realized state to the desirable state using the first characteristic, this whole process achieves the desirable state from any original state. Even if we restrict that the policy-maker cannot wait for an infinite time for each step, we can approximately implement the process given above. At least under the best response dynamics, in finite time, both policies achieve ϵ-convergence, and their limits are desirable outcomes.

6.1 Two-Step Cost Change and Loss Minimizing Equilibrium

Before introducing the exact policy, we first define an outcome, *Loss Minimizing (LM) equilibrium* with an arbitrary small upper-bound of loss $1/2 > \bar{\sigma} > 0$ s.t. $\pi_{min}(1 - \bar{\sigma}) > \gamma$ as follows.

Definition 8 (Loss Minimizing Equilibrium with a small $\bar{\sigma}$)**.** *An equilibrium $\boldsymbol{x}$ is a loss minimizing equilibrium if and only if the following conditions are satisfied:*

1. *All active senders send a single message ($x_l^r = 0$ if $l > 1$ $\forall r \in R$)*
2. *The benefit is larger than the cost for the active senders ($\pi(r) > \gamma''$ for any $r \in R$ s.t. $x_1^r > 0$)*
3. *$N(\boldsymbol{x}) = \min\{D, \phi/(1 - \sigma)\}$ where $\bar{\sigma} > \sigma \geqq 0$*
4. *When there exists a marginal type $r \in R$ s.t. $\pi(r)\frac{\phi}{N(\boldsymbol{x})} = \gamma''$, $x_0^r > 0$ and $x_1^r > 0$.*

where γ'' is the cost considered by senders at the equilibrium. When we discuss the two-step cost change we define later, γ'' is γ_z' at the end of the two-step cost change.

We analyze the characteristics of LM equilibrium and later define the policy which achieves LM equilibrium. LM equilibrium satisfies the following good properties:

Proposition 3. *LM equilibrium $\boldsymbol{x} \in X$ with a small enough $\bar{\sigma}$ induces a Pareto efficient allocation.*

Proof: Because $\boldsymbol{x}$ is NE, $\pi(r)\min(1, \frac{\phi}{N(\boldsymbol{x})}) \geqq \gamma''$ for any $r \in R$ s.t. $x_1^r > 0$. Thus, the conditions 1,2 and 4 in Lemma 9 are satisfied by the resource allocation Φ induced by $\boldsymbol{x}$ in information congestion. In addition, when $\bar{\sigma}$ is small, for any type $r \in R$, $\pi(r)(1-\bar{\sigma}) > \gamma$. Therefore, the condition 3 in Lemma 9 is satisfied. Further, because of Lemma 9, Φ induced by $\boldsymbol{x}$ is a Pareto optimal resource allocation.

Proposition 4. *LM equilibrium with a small $\bar{\sigma}$ is a regular Taylor evolutionarily stable state.*

Proof: see the Appendix.

The intuition of Proposition 4 is as follows. We consider an LM equilibrium in which marginal senders exist and are indifferent between sending or not sending a message (because, otherwise an LM equilibrium is an RTESS since all types have a type-specific unique best strategy). If a small portion of active marginal senders stop sending a message, the total message N decreases, and sending a message becomes more attractive for the marginal type. Further, a small portion of non-active marginal senders begin to send a message, and the system returns to the original equilibrium. The opposite deviation invites a similar reaction. When a small fluctuation appears, the marginal senders move to the opposite. For other (non-marginal) types, a small fluctuation does not change their optimal strategy.

From Theorem 8.4.7 in Sandholm (2010), RTESS is locally asymptotically stable under many dynamics. Thus, we can state that LM equilibrium has local stability. In addition,

Proposition 5. *In LM equilibrium $\boldsymbol{x} \in X$ with a small $\bar{\sigma}$, the social losses on both senders' and receivers' sides are almost minimized.*

Proof: In LM equilibrium, no senders send multiple messages, which is why no duplicates occur. When $D \leqq \phi$, no loss derived from the scarce attention appears on the senders' side and the receivers' utilized capacity (D) is maximum for a given population of senders. When $D > \phi$, $N(\boldsymbol{x}) = \phi/(1-\sigma)$. This means that the loss on the senders' side is almost 0 ($c(\phi/(1-\sigma) - \phi)$) with a tiny $\bar{\sigma} > \sigma$, even in the worst case. In addition, the receivers' capacity is maximally utilized since $N(\boldsymbol{x}) > \phi$ without any duplicates.

6.1.1 Two-Step Cost Change

We define *two-step cost change* which makes $\boldsymbol{x}(t)$ ϵ-converge to a LM equilibrium (with $\bar{\sigma}$) in a finite time. In the two-step cost change, the policy-maker is required to allow a small congestion (represented by $\bar{\sigma}$) to make the equilibrium locally stable when $\phi < D$. Let $\frac{\phi}{\underline{N}} \in (\frac{1}{2}, 1)$ s.t. $\frac{\phi}{\underline{N}} > (1-\bar{\sigma})$, $\pi_{max}(1-\frac{\phi}{\underline{N}})\frac{\phi}{\underline{N}} < \pi_{min}$[30] and $\frac{1}{\frac{\phi}{\underline{N}}} - 1 < \frac{MD}{2\pi_{max}}$[31] denote

[30]If this condition is satisfied and if $\frac{\phi}{N(\boldsymbol{x}(t))} > \frac{\phi}{\underline{N}}$ for all $t \in [t_z, t_{z+1}]$, no senders change the strategy from 1 to $l > 0$ under the best response dynamics.

[31]If this condition is satisfied, and if $\frac{\phi}{N(\boldsymbol{x}(t))} > \frac{\phi}{\underline{N}}$ for all $t \in [t_z, t_{z+1}]$, for any $r \in R$, the range of possible expected gain from sending a message $[\pi(r)\frac{\phi}{\underline{N}} - \gamma_z, \pi(r) - \gamma_z]$ is narrow. Thus, there

the lower-bound of the target (inverse) congestion intensity after the policy. If $\frac{\phi}{\underline{N}}$ is close enough to 1, these conditions are satisfied.

There exists $\underline{\delta}$ s.t. $(1+\underline{\delta})\frac{\phi}{\underline{N}} < 1$ and $(1-(1+\underline{\delta})\frac{\phi}{\underline{N}})(\pi_{max}+MD/3) < \min(MD/2, \pi_{min}-\pi_{max}(1-\frac{\phi}{\underline{N}})\frac{\phi}{\underline{N}})$ because both conditions are satisfied when $(1+\underline{\delta})\frac{\phi}{\underline{N}}$ is closer to 1 and less than 1. Since $(1+\delta')\frac{\phi}{\underline{N}}$ is a linear function of $\delta' \in \mathbb{R}$, there exists $\bar{\delta} > \underline{\delta}$ s.t. $(1+\bar{\delta})\frac{\phi}{\underline{N}} < 1$. Thus, the second condition is also satisfied by $\bar{\delta}$. In addition, for all $\delta' \in [\underline{\delta}, \bar{\delta}]$, both conditions are satisfied.

To make the outcome locally stable, any optimal strategy in the equilibrium must be selected by a small portion of senders. To satisfy this condition while achieving ϵ-convergence in finite time, we utilize the minimum unit of population ω[32]. Define $\Pi = \{\pi(r)\min(1, \frac{\phi}{j\omega})|r \in R, j \in \{0, ..., D/\omega\}\}$.

Define $\gamma'_1 = \pi_{max} + MD/3$. Pick $\delta \in (\underline{\delta}, \bar{\delta})$ and define $\gamma'_z = (1+\delta)\frac{\phi}{\underline{N}}\gamma'_{z-1}$ for $z \in \{2, 3, ..., z_{max}\}$ where $\gamma'_z < \pi_{min}$ if and only if $z = z_{max}$, and $\gamma'_z \neq \pi$ for all $\pi \in \Pi$ and $z \in \{1, 2, 3, ..., z_{max}\}$ (In the Appendix, we prove the existence of δ and the set of γ_z. In short, because Π is a finite set and $(\underline{\delta}, \bar{\delta})$ is an infinite set, there exists δ.). For convenience, we define

$$\bar{\epsilon} = \frac{\delta}{(1+\delta)\frac{\phi}{\underline{N}}}\frac{\phi}{l_{max}}. \tag{21}$$

We define *two-step cost change* as follows:

Definition 9 (Two-Step Cost Change)**.**

1. *Change the cost from γ to γ'_1, and define variables $z = 1$, $i = 1$ and $t_0 = 0$.*

2. a. *Wait for either*

$$(N(\boldsymbol{x}(t)) - l_{max}\epsilon_z) > \min(\phi, D) \tag{22}$$

or

$$(N(\boldsymbol{x}(t)) + l_{max}\epsilon_z) < \min(\phi, D) \tag{23}$$

and $t(\frac{\epsilon_z}{n}) < (t - t_{z-1})$ where $\bar{\epsilon} > \epsilon_z > 0$. (If neither appears forever, the limit is the intended outcome.) Define $t_z = t$. Update i s.t. $i = z$

is only a single $r \in R$ s.t. $\pi(r) > \gamma'_z$ but $\pi(r)\frac{\phi}{\underline{N}} < \gamma'_z$ where γ'_z is defined later.

[32] This minimum unit is required because we assume that we can change the cost only a finite number of times. Without this minimum unit, we sometimes meet technical problems. For example, if $\pi(r)\min(1, \frac{\phi}{j\omega}) = \gamma$ where $j \in \{0, ..., D/\omega\}$ for $r \in R$, and if $D = d(r) = j\omega$, $\boldsymbol{x}$ s.t. $x_1^r = d(r)$ is an equilibrium but not RTESS because $x_0^r = 0$. If we know ω, we can avoid the finite set of such cases where continuous parameters coincide.

b. If $N(\boldsymbol{x}(t)) - l_{max}\epsilon > \min(\phi, D)$ happens, stop this process and define $t_{z+1} = \infty$. Otherwise, update z s.t. $z = i+1$, change the cost from γ'_i to γ'_z and return to 2.a.

When $\phi > D$, typically, the policy-maker will not know the end of the cost change until the limit $t \to \infty$. However, the number of the required cost changes is at maximum finite, and the cost change is finished in finite time.

We claim the following main proposition.

Proposition 6. *From any initial state $\boldsymbol{x}(0)$, by the two-step cost change, $\boldsymbol{x}(t) \to \boldsymbol{x}^*$ and $\boldsymbol{x}^*$ is a LM equilibrium with $\bar{\sigma}$.*

Proof: See the Appendix.

The proof in the Appendix is based on the following logic. Since we assume the best response dynamics, a strictly best strategy for type r would be selected by at least $d(r) - \epsilon$ amount of senders if the strategy has been strictly best for r from t' to t'' s.t. $t(\epsilon) = t'' - t'$. Consider $\phi < D$. Step 1 makes sending 0 messages strictly best for all $r \in R$, and eventually $(N(\boldsymbol{x}(t_1)) + l_{max}\epsilon_1) < \phi$ where $t_1 = t(\epsilon_1/n)$ and $\bar{\epsilon} > \epsilon_1 > 0$. In step 2, for a positive integer z, if we slightly decrease the cost from γ'_z to γ'_{z+1} s.t. $\gamma'_z - \gamma'_{z+1} < MD/2$ at t_z, at maximum a single type $r \in R$ begins to send a message. If $(N(\boldsymbol{x}(t)) + l_{max}\epsilon_z) < \phi$, the amount of senders whose benefit is larger than γ_z is less than the capacity ϕ. Using this information, if we change the cost from γ_z to γ_{z+1}, all possible $N(\boldsymbol{x}(t))$ are lower than $\phi / \underline{\frac{\phi}{N}}$ after the cost change at t_z (Lemma 12 in the Appendix.). We repeat such a cost change and check the ϵ-convergence until $(N(\boldsymbol{x}(t)) - l_{max}\epsilon_k) \geqq \phi$ where k is a positive integer. If the condition appears, without any additional cost change, $\boldsymbol{x}(t)$ eventually enters the neighborhood of an LM equilibrium $\boldsymbol{x}^*$, and so it converges to $\boldsymbol{x}^*$. If both conditions never appear for γ_z, there exists the exact ϕ amount of senders whose benefit is larger than γ_z. As $t \to \infty$, in $\boldsymbol{x}(t)$, all such senders eventually select sending a message, and thus it converges to a LM equilibrium $\boldsymbol{x}^*$.

Proposition 6 means that we achieve LM equilibrium by the two-step cost change, regardless of both the initial condition and the distribution of senders' benefit. The policy-maker does not need to know the distribution of senders' benefit and does not have to observe the number of active senders and the benefit of marginal senders in each equilibrium.

The two-step cost change for LM equilibrium has three weak points. First, we need time to implement the policy. Second, we assume that the distribution of senders' benefit is unchanged; however, this may not be realistic in the long term. Third, this policy may need an external budget.

The discussion of the two-step cost change and LM equilibrium indicates that, if we have enough time and the cost is controllable, we can (almost) minimize both social losses, even if we cannot observe the private information of players, such as senders' benefit distribution, and cannot monitor the number of messages from each sender.

6.2 Full-Aid for First Messages and Universal Congestion Equilibrium

We define *full-aid for first messages* and the Universal Congestion (UC) equilibrium as follows:

Definition 10 (Full-Aid for First Messages)**.**
We define replacing the senders' utility function and the strategy set for each type $r \in R$ from (3) *and $L = \{0, ..., l_{max}\}$ to*

$$U(r, l, \boldsymbol{x}) = \pi(r) \min \Big((1 - (1 - \phi/N(\boldsymbol{x}))^l), 1 \Big) - \gamma(l - 1) \tag{24}$$

and $L' = \{1, ..., l'_{max}\}$ where $l'_{max} = l_{max} + 1$ as full-aid for first messages.

Under the full-aid for first messages, all senders stay in the competition, even if information congestion is extremely heavy. Thus, there is a lower bound for the congestion level under the aid. The pressure from the first messages from all senders can demotivate all senders to send additional messages if the senders' population is enough. We formally define such an equilibrium as follows.

Definition 11 (Universal Congestion Equilibrium)**.** *An equilibrium $\boldsymbol{x}$ is a universal congestion equilibrium if and only if all senders send a message ($x_1^r = d(r)\ \forall r \in R$) and $D = N(\boldsymbol{x}) > \phi$.*

We claim the following proposition:

Proposition 7. *If the population of potential senders (D) is large enough (s.t. $D > 2\phi$ and $\gamma > \pi_{max} \frac{\phi}{D} \left(1 - \frac{\phi}{D}\right)$), by the full-aid for first messages, $\boldsymbol{x}(t)$ converges to UC equilibrium $\boldsymbol{x}^*$ with global stability.*

Proof: Under the full aid for first messages, all senders are active, and cannot withdraw from message sending. If enough potential senders exist, by the full-aid, the probability of each message being received becomes so small that no senders want to send messages multiple times regardless of $\boldsymbol{x}$. If $D > 2\phi$ and $\gamma > \pi_{max}\frac{\phi}{D}\left(1 - \frac{\phi}{D}\right)$, no senders want to send multiple messages. When D increases, the right-hand side of the second condition approaches 0, and the conditions are satisfied. Further, sending a message is the strictly dominant strategy and eventually all senders select sending a message under the best response dynamics. This equilibrium satisfies the definition of UC equilibrium. Thus, after the full-aid begins, if enough potential senders exist, $\boldsymbol{x}(t)$ converges to UC equilibrium $\boldsymbol{x}$. In addition, if enough potential senders exist, UC equilibrium under the full aid is a RTESS because U is Lipschitz continuous in X and all players have a unique optimal strategy (It satisfies the stronger condition called *Globally Evolutionarily Stable State*. For this point, see the Appendix).

Proposition 8. *If the population of potential senders is large enough, by the full-aid for first messages, $\boldsymbol{x}(t)$ converges to UC equilibrium $\boldsymbol{x}^*$ which induces a Pareto efficient allocation (conditional on $L' = \{1, ..., l'_{max}\}$). If $\pi_{min}\frac{\phi}{D} > \gamma$, $\boldsymbol{x}^*$ induces a Pareto efficient allocation in the original game.*

Proof: Since all senders send a message, this $\boldsymbol{x}^*$ and Φ induced by $\boldsymbol{x}^*$ satisfies Lemma 10. If $\pi_{min}\frac{\phi}{D} > \gamma$, this $\boldsymbol{x}^*$ and Φ induced by $\boldsymbol{x}^*$ satisfies Lemma 9.

Proposition 8 implies that, if there exist enough potential senders, a Pareto efficient resource allocation appears conditional on the policy. In addition, if the original cost of sending a message is small enough, the equilibrium after the policy satisfies the Pareto efficiency without the restriction by the policy.

From Definition 11, we have the following proposition.

Proposition 9. *In UC equilibrium, the social loss on receivers' side is 0, and the social loss on senders' side is $\gamma(D - \phi)$.*

Our discussion implies that, if there are enough potential senders (D is large enough), we can minimize the social loss on the receivers' side while allowing for the social loss on senders' side. The policy-maker would want the aid if the cost per message γ is relatively smaller than the cost per attention c.

The aid for making UC equilibrium has three weak points. First, we need enough potential senders. Second, it is not budget balanced. Third, the policy-maker is required to distinguish whether the message is the first or not for each sender if senders can fake their identity at little cost[33].

The discussion of UC equilibrium tells us that, if there are enough potential senders and enough budget for the aid, we can make $\boldsymbol{x}(t)$ converge to a stable Pareto efficient allocation with 0 social loss on the receivers' side by the full-aid for first messages.

6.3 Evolutionary Implementation

Both policies can be classified as an application of evolutionary implementation proposed by Sandholm (2002, 2005, 2007). When a policy-maker observes players' behaviors only at an aggregate level, it is sometimes difficult to achieve a desirable state. For example, if the policy-maker subsidies an action, the subsidy may invite too many types of players to select the action. Evolutionary implementation aims to solve inefficiencies in such anonymous situations by using evolutionary dynamics. In this study, we define evolutionary implementation as a policy which leads the target system, in which players can behave anonymously, to a desirable state by using evolutionary dynamics[34]. Our policies satisfy this definition.

However, our policies are completely different from the traditional approach which uses a *potential function*. Sandholm (2002, 2005, 2007) analyzes the situation where the externality among players is symmetric. In the symmetric-externality case, there exists a price scheme on actions such that the price scheme equalizes the private cost of the action with its social cost. Under such a price scheme, each player gradually behaves in accordance with the policy-maker's conditions. In the congestion game with such a price scheme, Sandholm (2002) finds the potential function which represents the total of players' utility. When such a potential function exists, standard evolutionary dynamics lead the system towards the direction in which the total of players' utility increases. In addition, the negative externality in the congestion game makes the potential function strictly concave, and thus there exists a unique equilibrium with global stability. In the similar negative-externality setting, Sandholm (2005) considers the additional factor, idiosyncratic payoffs which depend only on the type of players and the action. Sandholm (2007) discusses general settings where the

[33] Perhaps, the possibility of random sampling and penalty is enough to keep senders from cheating, because the incentive of cheating is smaller than γ in UC equilibrium.

[34] Our definition is slightly different from the original discussion in Sandholm (2002). Sandholm (2002) selects a large set of evolutionary dynamics (called admissible dynamics), and the price scheme invites the social optimum as the limit of consequence of all dynamics in the set. The discussions in previous literature focus on the robust implication among wide ranges of dynamics.

externality is not necessarily negative. Because the externality can be complicated, multiple equilibria can appear. By applying the stochastic dynamics, Sandholm (2007) shows that the price scheme makes the system almost always stay at an efficient state in the long run. As a recent advancement, Lahkar and Mukherjee (2019, 2021) find a methodology that broadens the application range of the potential-function approach. Their methodology requires a type-independent externality based on the *aggregate strategy level* instead of a symmetric externality[35]. The current literature of the potential-function approach assumes that, before policies are designed, the policy-maker knows/estimates the aggregate externality of each action at each state[36].

We focus on the situation where a policy-maker does not know the externality at each state, and, therefore, cannot apply the proposed approach with a potential function. Instead, our policies use the unique characteristics of information congestion. In the two-step cost change, the initial step leads the system to a certain condition by creating a strictly dominant strategy for players, and the second step makes the system move from the condition to the desirable equilibrium. This process is possible because the characteristic of the desirable equilibrium is well known. In the full-aid for first messages, a similar process occurs. We make the desirable strategy for senders the strictly dominant strategy, and the system converges to the desirable equilibrium. In our approach, the policy-maker does not have to understand the details of externalities among players. In addition, the number of required cost changes is finite. (However, the outcome of the two-step cost change is typically not globally stable.)

6.4 Welfare Implication

We briefly discuss the "welfare" implications of the presented policies[37]. The welfare implications depend on exogenous parameters and the relative weight of senders and receivers in the welfare function. ξ denotes the average value of a message's content for the receivers.

[35]If an externality of each action is independent of the type of the player who performs the action, there exists a price which represents the externality of each action at each state. Lahkar and Mukherjee (2019) focus on a public goods game, and Lahkar and Mukherjee (2021) analyze general settings where a variable, called aggregate strategy level, can represent the impact of each state. Symmetric externality is a special case of type-independent externality where the impact of the externality on each type is identical. However, the impact of each state in Sandholm (2002, 2005, 2007) is not necessarily represented by a single variable. Therefore, both methods complement each other.

[36]We do not claim that this requirement is necessary for the potential-function approach because the process of policies can reveal the externality and the type distribution.

[37]Our discussion focuses on the competition in advertising and is classified as partial equilibrium.

Suppose $\xi > 0$ and that each information transmission is beneficial for receivers. In this information congestion game, Pareto efficient allocation requires no duplication of messages. Conditional on this efficiency, if we apply γ instead of γ', the allocation achieved by the two-step cost change (approximately) maximizes the sum of the senders' and receivers' expected utility. The allocation achieved by the full-aid for first messages maximizes the players' minimum expected utility when enough potential senders exist conditional on the strategy set after the full aid. In particular, the allocation achieved by the former policy is often called first best allocation. Previous literature typically tries to achieve this first best allocation using a traditional single-step policy for advertising. Our result implies that we cannot always achieve the first best allocation through the traditional one-step cost-change policies if the structure of advertising is primitive.

7 Comparison with Anderson and de Palma (2013)

We compare our implications with those found in Anderson and de Palma (2013) (hereafter AD) because our discussion is based on the continuous repetition of the base model which is similar to the model presented in AD[38].

7.1 Anderson and de Palma (2013)'s Contribution

We summarize the results and implications of AD. First, AD show the existence of multiple equilibria. Each equilibrium is characterized by a minimum benefit for active senders. In the no congestion equilibrium, various senders send single messages. In the heavily congested equilibrium, only large-profit senders send multiple messages. This finding regarding multiple equilibria is new as the previous literature on information congestion did not consider multiple messages.

Second, AD discuss two possible interventions, namely, raising the cost of sending messages and ad caps. AD's Proposition 4 shows that when the total number of active senders is less than the receiver's capacity, if the cost per message increases appropriately[39], a (weak) Pareto improvement can happen without lump-sum transfers. This implies that, an

[38] AD assume that the number of the sender's benefit types is infinite. In our model, the number of the benefit types is finite.

[39] The proper raising cost makes a marginal sender in the original equilibrium indifferent between sending a message or not.

intervention for information congestion can drastically improve the efficiency of the system and this interesting result motivates us to write this paper.

Ad caps indicate the restriction regarding the maximum number of messages that senders can send. AD show that the ad caps can eliminate the equilibrium with heavy congestion, and it can increase the probability that the largest-profit senders fail to capture the receiver's attention.

7.2 Our Contribution

Our study answers the unsolved research questions of AD. First, which equilibrium is locally stable? Our result shows that multiple equilibria have local stability. This result strongly supports the implications of AD.

Second, we answer the following question: does an additional cost on sending messages always ease the information congestion problem? Our result shows that an additional cost on messages sometimes causes efficiency to fall. AD's Proposition 4 shows that, under specific conditions, the appropriate cost establishes the existence of relatively efficient equilibrium. However, the realization of the efficient equilibrium is not guaranteed because AD do not take into account dynamics. In addition, the policy-maker needs to know extensive non-aggregate information such as the benefit of the marginal sender and the number of active senders in the original equilibrium. Our discussion in Section 6 shows that, from any state $\boldsymbol{x}(0)$, if the aggregate factors such as D and $N(\boldsymbol{x}(t))$ are observable, we are able to make the system converge to Pareto efficient equilibrium with local stability by the two-step cost change.

As a third question, we ask: are there any situations in which aids/subsidies on messages are desirable? Our result shows that the full-aid for first messages achieves a Pareto efficient and globally stable equilibrium if there are enough potential senders. AD analyze neither aid/subsidies nor a restriction on the minimum number of messages.

8 Conclusion

Under the primitive structure of advertising, advertisers select an excessive amount of advertising in stable equilibrium even if the benefit of repetition is ignored. This structural inefficiency occurs because the private marginal cost of advertising does not take into account the decreasing performance in consumers' ability to process information. Traditional one-shot interventions do not work well, and we need costly interventions to efficiently utilize receivers' attention. Our discussion shows the difficulty in controlling advertising without an advanced structure.

Our main conclusion, the inefficiency of the primitive structure, potentially explains why we have seen recent advancements in advertising structures (e.g. Rust and Oliver 1994, Dahlen and Rosengren 2016). Typically, under new types of advertising, consumers are more active and the communication between consumers and advertisers is (relatively) interactive[40]. Additionally, when advertisements are not well targeted to consumers, consumers regard them as nuisances and try to avoid such advertising via new technology/legislations[41]. These advanced structures can overcome the limitation derived from information congestion.

This paper contributes to the field of information congestion. We analyze the theoretical model of information congestion in which heterogeneous senders can send multiple messages. We assume a finite number of senders' benefit types. Hence, the model fits the population game framework. We analyze equilibria by referring to the known concepts[42] of stability in the framework.

We find several new results which are useful for understanding information congestion. First, a suboptimal stable equilibrium sometimes appears. Thus, interventions are, at times, required to make the system efficient. Second, raising the cost sometimes increases social loss. Due to this possibility, we must think carefully before imposing interventions. Third, the two-step cost change is effective, regardless of the initial condition and the distribution of the senders' benefit. It achieves loss minimizing equilibrium with local stability, Pareto efficiency, and almost minimum social loss. Fourth, with enough potential senders, the full-aid for first messages achieves the stable universal congestion equilibrium with minimum social loss of the receivers' capacity and Pareto efficiency appear. Our main results support the main policy implications in previous studies such as Van Zandt (2004) and Anderson and de Palma (2013). The primitive structure of advertising can cause efficiency loss in the allocation of limited attention. Interventions sometimes significantly improve

[40]For example, behavior-based (conditional on consumption history) advertising in Shen and Villas-Boas (2018), advertising on nonretail platforms in Eliaz and Spiegler (2020) and skippable advertising and weak/strong conversion in Dukes and Qihong (2020).

[41]For example, targeting advertising and ad avoidance in Johnson (2013), and Do Not Call policy in Goh et al (2015). In particular, our discussion complements Johnson (2013) because both approaches get similar implications that the efficiency of advertising with simple structures is limited.

[42]For example, Regular Taylor evolutionary stable state is a sufficient condition of local stability under many types of dynamics.

the performance of advertising.

Our results can be interpreted as "gleaning" through the many discussions that use random assignment as a tie-breaking rule. For example, in a sealed-auction, if the bidders believe that other bidders are submitting the same bid price as theirs, they may bid the same price multiple times by borrowing their friends' names. When each bidder follows this logic, information congestion occurs. Our result supports an entry payment for auctions by revealing the repercussions that occur without the entry payment. In many mechanisms, senders can reveal their preferences, and we can achieve efficient allocation by using their preferences. Our study supplies the benchmark without the information regarding senders' (and receivers') preferences. Thus, this benchmark is useful to clarify the value of revealing/collecting private information in many studies with a random tie-breaking rule.

A Appendix

A.1 How to Derive Best Response Diagram (Figure 1)

Each curve in Figure 1 is the indifference boundary between two strategies of senders such as (0,1), (1,2) and (3,4)[43]. Consider a sender with type $r > 0$ and given $\boldsymbol{x}$. We can derive the indifference boundary between $(k, k+1)$ from the difference of the utility function as follows:

$$\begin{aligned} & U(r, k+1, \boldsymbol{x}) - U(r, k, \boldsymbol{x}) = 0 \\ \iff & \pi(r)\frac{\phi}{N(\boldsymbol{x})}\left(1 - \frac{\phi}{N(\boldsymbol{x})}\right)^k = \gamma \\ \iff & \frac{\gamma}{\pi(r)} = \frac{\phi}{N(\boldsymbol{x})}\left(1 - \frac{\phi}{N(\boldsymbol{x})}\right)^k . \end{aligned} \tag{25}$$

We only need to focus on the boundaries of two consecutive numbers because of the following lemma:

Lemma 11. *For a given $0 < \frac{\phi}{N(\boldsymbol{x})} < 1$, the best strategies for type r are either a single number or two consecutive numbers.*

[43]We focus on the strategies $l < 6$ in Figure 1.

Lemma 11 indicates that the set of best response strategies for each $\boldsymbol{x}$ includes at maximum two consecutive numbers. The other sets of strategies, such as $\{0, 2\}$ and $\{1, 3, 4\}$, cannot be a set of best strategies.

Proof of Lemma 11: Consider $0 < \phi/N < 1$. The utility function conditional on $\boldsymbol{x}$, denoted by $U_C(l)$, is strictly concave in $l \geqq 1$ if $\gamma \leqq \pi(r)\phi/N$. If and only if $\gamma < \pi(r)\phi/N$ is not satisfied, only 0 is the optimal strategy. When $\gamma = \pi(r)\phi/N$, $l = 1$ and $l = 0$ have the same utility ($= 0$) and both are optimal. When $\gamma < \pi(r)\phi/N$, $l = 1$ strictly dominates $l = 0$.

From now on, we focus on $l \geqq 1$ and $\gamma < \pi(r)\phi/N$. Since $U_C(l)$ is a single-variable function, when we focus on the discrete set of $l > 0$, we can simply do the convex extension. By the definition of strictly concave function, for any $t \in (0, 1)$ and $j, k \geqq 1$,

$$tU_C(j) + (1 - t)U_C(k) < U_C(tj + (1 - t)k). \tag{26}$$

This indicates that, if l is a continuous number and the two strategies have the same utility, the third strategy with the same utility cannot exist. When we consider the subset of the strategy set, this implication holds. Thus, in the discrete case, we do not have to consider the case in which three strategies have the same utility.

In addition, because of (26), the two strategies with the same maximum utility are consecutive numbers. Suppose both $k1$ and $k2$ maximize the utility function, but they are not consecutive. Then, we can find at least an integer $k3$ between $k1$ and $k2$ and t such that $k3 = tk1 + (1 - t)k2$. $U_C(k3)$ is larger than $U_C(k1) = U_C(k2)$ due to (26). This contradicts the assumption that $k1$ and $k2$ maximize the utility function.

A.2 How to Derive Bifurcation Diagram (Figure 2)

Consider homogeneous senders $R = \{1\}$. From the discussion in A.1, the set of optimal strategies in $\boldsymbol{x} \in X$ includes two consecutive numbers at maximum. We can then figure out the equilibria with the given parameters $\frac{\gamma}{\pi}$ and $\frac{\phi}{D}$ without complex calculations through the following way.

First, we can calculate the possible range of $\frac{\phi}{N(\boldsymbol{x})}$ with the pair of strategies. For instance, if $\frac{\phi}{D} = 1$, the pair (1,2) can achieve $\frac{1}{2} \leqq \frac{\phi}{N(\boldsymbol{x})} \leqq 1$. Second, from Figure 1, we know that the best response of senders in each $\frac{\phi}{N(\boldsymbol{x})}$ and $\frac{\gamma}{\pi}$. Third, by combining the boundaries in Figure 1 and the ranges, we can draw Figure 2 which illustrates all possible equilibria with the parameters.

A.3 Stability of Equilibria in Figure 2

Consider $\frac{\gamma}{\pi} = 0.2$ which is represented by the area in the orange rectangle in Figure 2. Figure 3 zooms in the orange rectangle. The arrows in Figure 3 illustrate the direction of movement in the system if we apply the best response or any impartial pairwise comparison dynamic to $\boldsymbol{x}$ in which $N = N(\boldsymbol{x})$ and all senders select the strategies in which the number of sending messages is either of two consecutive numbers like sending a message once and twice.

We check the stability of equilibria A and C in Figure 3. Consider the C's neighborhood $O = \{\boldsymbol{y} \in X : 0.4 < \phi/N(\boldsymbol{y}) < 0.6\}$. By the same logic in A.1, at least in $0.28 < \phi/N(\boldsymbol{y}) < 0.72$, the strategy of sending a message twice is the unique best response. Therefore, both (5) and (6) are satisfied. Further, equilibrium C is a regular Taylor evolutionarily stable state. By the similar logic, equilibrium A is a regular Taylor evolutionarily stable state, too.

A.4 Stability of LM equilibrium

We check whether a LM equilibrium meets the definition of Regular Taylor Evolutionarily Stable State (hereafter RTESS). For all $r \in R,\ l \in L$, $\boldsymbol{x} \in X$ and $i \in \{1, 2, ..., n\}$, all of U's partial derivatives $\frac{\partial U(r,l,\boldsymbol{x})}{\partial x_i}$ exist and are continuous in X so that U is continuously differentiable. Therefore, U is Lipschitz continuous in X.

We defined RTESS in equations (5) and (6) using the definition adopted from Sandholm (2010). In LM equilibrium, if $D \leqq \phi$, there does not exist any marginal type. When each type has a unique optimal strategy, both (5) and (6) are satisfied, and such an equilibrium is a RTESS.

If $D > \phi$ and $\phi/N(\boldsymbol{x}) \leqq 1$, and if there does not exist any marginal type, we can apply the similar logic in the previous paragraph, and $\boldsymbol{x}$ is a RTESS. If there exists a marginal type, from the definition of LM equilibrium, the active senders with the minimum benefit among the active senders (hereafter senders with r_{amin}) are indifferent between sending no messages and sending a message. In addition, from the definition, $x_0^{r_{amin}} > 0$ and $x_1^{r_{amin}} > 0$. Therefore, (5) is satisfied.

When (5) is satisfied, the support of $\boldsymbol{x}$ includes all optimal strategies for each type. Then, the condition $(\boldsymbol{y} - \boldsymbol{x})'U(\boldsymbol{x}) = 0$ in (6) can be satisfied only if the support of $\boldsymbol{x}$ is weakly larger than the support of $\boldsymbol{y}$. Thus, we focus on the changes in r_{amin}'s strategies between sending a message or not. We denote such a change by $\boldsymbol{z}$. Instead of (6), we use the following condition from p282 in Sandholm (2010).

$$\boldsymbol{z}'DU(\boldsymbol{x})\boldsymbol{z} < 0 \ \forall \, nonzero \ \boldsymbol{z} \in TX \cap \mathbb{R}^n_{L(\boldsymbol{x})} \tag{27}$$

where $L(\boldsymbol{x})$ is the support of $\boldsymbol{x}$, and $\boldsymbol{y} \in \mathbb{R}^n_{L(\boldsymbol{x})}$ indicates $\boldsymbol{y} \in \mathbb{R}^n$ and the support of y is a subset of $L(\boldsymbol{x})$.

Discuss $\boldsymbol{z}_{01} = \boldsymbol{e}_q - \boldsymbol{e}_s$ such that $\boldsymbol{e}_q/\boldsymbol{e}_s$ represents a unit population of type r_{amin} which sends 0 messages/ 1 message. ($\boldsymbol{e}_i$ is a standard basis of $\mathbb{R}^n$.). For example, x_q is the amount of the marginal-type senders who select 0. If

$$\frac{\partial U(r_{amin}, 0, \boldsymbol{x})}{\partial x_q} - \frac{\partial U(r_{amin}, 0, \boldsymbol{x})}{\partial x_s} < \frac{\partial U(r_{amin}, 1, \boldsymbol{x})}{\partial x_q} - \frac{\partial U(r_{amin}, 1, \boldsymbol{x})}{\partial x_s} \tag{28}$$

is satisfied, the equation (27) is satisfied.

$$\frac{\partial U(r_{amin}, 0, \boldsymbol{x})}{\partial x_q} - \frac{\partial U(r_{amin}, 0, \boldsymbol{x})}{\partial x_s} = 0 \tag{29}$$

and

$$\frac{\pi(r_{amin})\phi}{N(\boldsymbol{x})^2} = \frac{\partial U(r_{amin}, 1, \boldsymbol{x})}{\partial x_q} - \frac{\partial U(r_{amin}, 1, \boldsymbol{x})}{\partial x_s}. \tag{30}$$

Any other $nonzero$ $\boldsymbol{z} \in TX \cap \mathbb{R}^n_{L(\boldsymbol{x})}$ is the scalar multiplication of $\boldsymbol{z}_{01}$ and the sign of $\boldsymbol{z}'DU(\boldsymbol{x})\boldsymbol{z}$ is unchanged. Therefore, LM equilibrium is a RTESS.

A.5 Existence of δ and γ'_z s.t. $\gamma'_z \neq \pi$ for any $\pi \in \Pi$

For an arbitrary $\delta \in (\underline{\delta}, \bar{\delta})$, we define $\gamma'_z(\delta) = (1+\delta)\frac{\phi}{\underline{N}}\gamma'_{z-1}$ for any $z \in \{2, 3, ..., z_{max}\}$ where $\gamma'_z(\delta) < \pi_{min}$ if and only if $z = z_{max}$. We claim that there exists $\delta'' \in (\underline{\delta}, \bar{\delta})$ s.t. $\gamma'_z(\delta'') \neq \pi$ for all $\pi \in \Pi$ and $z \in \{1, 2, 3, ..., z_{max}\}$. Because $\gamma'_z(\delta)$ is differentiable in δ, there exists $(\underline{\delta}', \bar{\delta}') \subset (\underline{\delta}, \bar{\delta})$ s.t. z_{max} is constant if $\delta \in (\underline{\delta}', \bar{\delta}')$.

Consider a $\delta \in (\underline{\delta}', \bar{\delta}')$ with a certain z_{max}. If any $\delta'' \in (\underline{\delta}', \bar{\delta}')$ satisfies the condition, the claim is correct. If the condition is not satisfied by δ, there exists at least a pair of $\pi \in \Pi$ and $z \in \{1, 2, 3, ..., z_{max}\}$ s.t. $\gamma'_z(\delta) = \pi$. Among the other pairs (π, z) s.t. $\gamma'_z(\delta) \neq \pi$, we can find the minimum difference $|\gamma'_z(\delta) - \pi| \geqq M$. Because $\gamma'_z(\delta)$ is differentiable and strictly increasing in δ, we can slightly decrease from δ to $\delta'' \in (\underline{\delta}', \delta)$ s.t. $M > |\gamma'_z(\delta) - \gamma'_z(\delta'')| > 0$ for any $z \in \{1, 2, 3, ..., z_{max}\}$. As a result, we find $\delta'' \in (\underline{\delta}, \bar{\delta})$ s.t. $\gamma'_z(\delta'') \neq \pi$ for all $\pi \in \Pi$ and $z \in \{1, 2, 3, ..., z_{max}\}$.

A.6 Proof of Proposition 6

We fix $\frac{\phi}{\underline{N}}$ and δ which satisfy the required conditions (for a given $\bar{\sigma}$) explained in the main text. Then, a set of γ'_z and $\bar{\epsilon}$ are uniquely decided. The timing of switching from γ'_z to γ'_{z+1}, denoted by t_z, depends on the reaction of $\phi/N(\boldsymbol{x}(t))$. Consider $\phi \leqq D$ first.

We would like to show $\boldsymbol{x}(t) \to \boldsymbol{x}^*$ and $\boldsymbol{x}^*$ is LM equilibrium (with $\bar{\sigma}$) if the two-step cost change is conducted. When $\phi \leqq D$, an equilibrium $\boldsymbol{x}^*$ is an LM equilibrium if and only if $\frac{\phi}{N(\boldsymbol{x}^*)}$ is 1 or almost 1 ($\frac{\phi}{N(\boldsymbol{x}^*)} > (1-\bar{\sigma})$), all active senders send a single message, and all best strategies for each type are utilized, and the benefit is larger than the cost for all active senders. Because of $\frac{\phi}{\underline{N}}$'s definition, $\frac{\phi}{\underline{N}} > (1-\bar{\sigma})$, and the first condition is satisfied if $\frac{\phi}{N(\boldsymbol{x}^*)} > \frac{\phi}{\underline{N}}$ and $\frac{\phi}{N(\boldsymbol{x}^*)} \leqq 1$.

After Step 1, since sending 0 messages is the strictly dominant strategy among all senders for γ'_1, $(N(\boldsymbol{x}(t)) + l_{max}\epsilon_1) < \phi$ eventually happens in 2.a. Further, there exist at least $D - \epsilon_1$ senders who send 0 messages and at maximum ϵ_1 senders who are going to select 0 messages. In 2.b., at t_1, we change the cost from γ'_1 to γ'_2. We return to 2.a.

Before checking each process in Step 2, we want to confirm a fact about this process. Because of the definition of $\frac{\phi}{\underline{N}}$, if we keep $\frac{\phi}{N(\boldsymbol{x}(t))} > \frac{\phi}{\underline{N}}$, for any $r \in R$, sending a message is strictly better than any $l > 1$. We claim that $\frac{\phi}{N(\boldsymbol{x}(t))} > \frac{\phi}{\underline{N}}$ for all $t \geqq t_1$ in this policy, and in the following paragraphs, we check this claim.

When we change the cost from γ'_1 to γ'_2 at t_1, $N(\boldsymbol{x}(t_1)) < \phi$. Under γ'_2, because $\gamma'_z - \gamma'_{z+1} < MD/2$[44], only senders with π_{max} can have a larger benefit than γ'_2. Suppose $\pi_{max} > \gamma'_2$ (Or we can repeat the cost change finite times until the condition is satisfied and each suffix for γ'_z and t_z in the following discussion can be increased by the number of the repetition.). For any other type, the unique optimal strategy is sending 0 messages regardless of $\boldsymbol{x}$, and ϵ-convergence happens. Later, as Lemma 12, we prove that, for any $z \in \{1, 2, 3, ..., z_{max}\}$ s.t. t_z exists, $\frac{\phi}{N(\boldsymbol{x}(t))} > \frac{\phi}{\underline{N}}$ for any $t \geqq t_z$ if we stop the cost change after t_z.

If we observe $N(\boldsymbol{x}(t_2)) + l_{max}\epsilon_2 < \phi$ at the end of 2.a, in the whole population D, there exist less than $N(\boldsymbol{x}(t_2)) + \epsilon_2$ amount of senders whose benefit is strictly larger than γ'_2. (Otherwise, some portion of senders in $(D - \epsilon_2)$ have to select the suboptimal strategy.) At t_2, we change the cost from γ'_2 to γ'_3.

We can repeat the similar discussion about the cost change from γ'_2 to γ'_3. At t_2, $N(\boldsymbol{x}(t_2)) < \phi$. Because $\gamma'_z - \gamma'_{z+1} < MD/2$, there are no types $r \in R$ s.t. $\gamma'_2 > \pi(r) > \gamma'_3$.

[44] Since $(1-(1+\delta')\frac{\phi}{\underline{N}})(\pi_{max} + MD/3) < MD/2$, $\gamma'_z - \gamma'_{z+1} < MD/2$.

Thus, there exist less than $N(\boldsymbol{x}(t_2)) + \epsilon_2$ amount of senders whose benefit is strictly larger than γ_3' in the whole population D. For such senders, sending a message is the best strategy after t_2 as long as $N(\boldsymbol{x}(t)) < \phi/\frac{\phi}{\underline{N}}$. In the group (at minimum $(D - \epsilon_2)$) with the optimal strategy for γ_2', no players have incentive to change the strategy after t_2 (for γ_3') at least as long as $N(\boldsymbol{x}(t)) < \phi/\frac{\phi}{\underline{N}}$. In addition, even if ϵ_2 senders change their strategy from 0 to l_{max}[45], $N(\boldsymbol{x}(t_2)) + l_{max}\epsilon_2 < \phi$. Therefore, $N(\boldsymbol{x}(t)) < \phi < \phi/\frac{\phi}{\underline{N}}$ after t_2 (if any additional cost change does not happen after γ_3'). In addition, after t_2, there exists a strictly best strategy for each type, and ϵ-convergence happens for each type. We can repeat the same discussion about the cost change from γ_z' to γ_{z+1}' as long as there are no types $r \in R$ s.t. $\gamma_z' > \pi(r) > \gamma_{z+1}'$.

Suppose there exists a marginal type $r' \in R$ s.t. $\gamma_z' > \pi(r') > \gamma_{z+1}'$. Consider the upper bound (hereafter $\bar{N}$) of $N(\boldsymbol{x}(t))$ after t_z if we ignore the further cost changes. We suppose $d(r') > \pi(r')\phi/\gamma_{z+1}'$ because this setting would maximize the possible N. If we focus on $D - \epsilon_z$ with the optimal strategy for γ_z', any senders except type r' does not have any incentive to change their strategies under γ_{z+1}' and as long as $\frac{\phi}{N(\boldsymbol{x}(t))} > \frac{\phi}{\underline{N}}$. At t_z, $\frac{\phi}{N(\boldsymbol{x}(t_z))} > \frac{\phi}{\underline{N}}$.

After t_z, as long as $\frac{\phi}{N(\boldsymbol{x}(t))} > \frac{\phi}{\underline{N}}$, only the marginal r' senders in $D - \epsilon_z$ and ϵ_z senders can increase messages under the best response dynamics. After t_z, as long as $\gamma_{z+1}' < \frac{\phi}{N(\boldsymbol{x}(t))}\pi(r')$, r' senders change their strategy from 0 to 1. Because we assume enough $d(r')$, if we do not consider ϵ_z senders, the minimum inverse congestion ratio would be $\frac{\phi}{N}$ s.t. $\gamma_{z+1}' = \frac{\phi}{N}\pi(r')$, and thus $N = \frac{\pi(r')\phi}{\gamma_{z+1}'}$. Even if we consider ϵ_z, N cannot be larger than $\frac{\gamma_z'\phi}{\gamma_{z+1}'} + l_{max}\epsilon_z$ (We can replace $\pi(r')$ by γ_z' because $\gamma_z' > \pi(r')$). Therefore, $\bar{N} = \frac{\gamma_z'\phi}{\gamma_{z+1}'} + l_{max}\epsilon_z$ as long as $\frac{\phi}{N(\boldsymbol{x}(t))} > \frac{\phi}{\underline{N}}$ after t_z. From the definition of $\bar{\epsilon} > \epsilon_z$ and γ_z, $\bar{N} < \phi/\frac{\phi}{\underline{N}}$. $\frac{\phi}{N(\boldsymbol{x}(t))} > \frac{\phi}{\underline{N}}$ is satisfied at t_z, and the possible $\bar{N}$ is too small to break the condition from $\boldsymbol{x}(t_z)$ under the best response dynamics. In conclusion, $\frac{\phi}{N(\boldsymbol{x}(t))} > \frac{\phi}{\underline{N}}$ if $t \geqq t_z$ and if there are no additional cost changes.

We can repeat the same argument for each $[t_z, t_{z+1}]$ as long as $t_z < \infty$. Therefore,

Lemma 12. *In Step 2 after t_1, $\frac{\phi}{N(\boldsymbol{x}(t))} > \frac{\phi}{\underline{N}}$ under the best response dynamics.*

Next, we check whether $\boldsymbol{x}(t)$ converges to LM equilibrium in Step 2. Suppose that the cost change continues until t_k when we change the cost from γ_k to γ_{k+1}. If $N(\boldsymbol{x}(t_{k+1})) + l_{max}\epsilon_{k+1} < \phi$, there are no enough senders under γ_{k+1} to fill the capacity ϕ, and thus we move to the next step by changing the cost from γ_{k+1} to γ_{k+2}.

[45]This is the worst case, and there are no incentives which make such a strategy change.

If $N(\boldsymbol{x}(t)) - l_{max}\epsilon_{k+1} > \phi$, this policy stops the cost change (and any additional intervention is not required). Because of Lemma 12, $\frac{\phi}{N(\boldsymbol{x}(t))} \geqq \frac{\phi}{\underline{N}}$. Since we take $\frac{\phi}{\underline{N}}$ s.t. $\pi_{max}(1/\frac{\phi}{\underline{N}} - 1) < \frac{MD}{2}$, there is only a single type r_{mar} whose optimal strategy fluctuates between 0 and 1 after the stop. Let N_{mar} denote the total amount of messages s.t. $\pi(r_{mar})\frac{\phi}{N_{mar}} = \gamma_{k+1}$. When $\frac{\phi}{N(\boldsymbol{x}(t))} < \frac{\phi}{N_{mar}}$, $\dot{x}_1^{r_{mar}} > 0$, and vice versa. Let $\epsilon'(t)$ in $t \in [t_k, \infty)$ denote the upper bound of the total population who selects suboptimal strategies in $r \in R - \{r_{mar}\}$[46].

As $t \to \infty$, at least $\sum_{r \in R - \{r_{mar}\}} d(r) - \epsilon'(t)$ senders select the unique optimal strategy for each type. Let $N_{-r_{mar}}$ denote the limit amount of messages from all $r \neq r_{mar}$. From the process of the two-step cost change, we know $N_{-r_{mar}} < \phi$[47]. Since $\epsilon'(t)$ is monotonically decreasing, the total amount of messages from the non-marginal types must be between $N_{-r_{mar}} - l_{max}\epsilon'(t)$ and $N_{-r_{mar}} + l_{max}\epsilon'(t)$. In addition, since $N(\boldsymbol{x}(t)) - l_{max}\epsilon_{k+1} > \phi$, we claim that

$$N_{-r_{mar}} + d(r_{mar}) > \phi. \tag{31}$$

This is because if $N_{-r_{mar}} + d(r_{mar}) \leqq \phi$, more than ϵ_{k+1} senders have to select the suboptimal strategy at t_{k+1}[48].

If $d(r_{mar})$ is small s.t. $N_{mar} - N_{-r_{mar}} > d(r_{mar})$, there exists $\bar{t}$ s.t. for any $t > \bar{t}$,

$$\frac{\phi}{N(\boldsymbol{x}(t))} > \frac{\phi}{N_{mar}}. \tag{32}$$

First, since $N_{-r_{mar}} + l_{max}\epsilon'(t)$ is a strictly decreasing function in t, we can find t s.t. the possible total amount of messages from the non-marginal types would be arbitrarily close to $N_{-r_{mar}}$. Second, for the marginal type, the best strategy is either 0 or 1 (by Lemma 12). The other strategies are always strictly suboptimal for r_{mar}, and thus the possible maximum amount of messages converges to $d(r_{mar})$ as $t \to \infty$. Since $N_{mar} - N_{-r_{mar}} > d(r_{mar})$, we can find $\bar{t}$ s.t. (32) for any $t > \bar{t}$. Thus, after $t > \bar{t}$, for all type, there is a unique best strategy, and the system converges to a unique point $\boldsymbol{x}^* \in X$. This point satisfies the definition of LM equilibrium.

[46]Such $\epsilon'(t)$ exists because there exists a unique optimal strategy for each $r \in R - \{r_{mar}\}$.

[47]Otherwise, the process would stop before r_{k+1}.

[48]Even for the marginal type, the population selecting $l > 1$ cannot be larger than ϵ_{k+1}. When $N_{-r_{mar}} + d(r_{mar}) = \phi$, even if $d(r_{mar}) - \epsilon_{k+1}/n$ senders in r_{mar} selects 1 and all ϵ_{k+1} senders select l_{max}, the maximum total amount $N(\boldsymbol{x}(t_{k+1}))$ is lower than $N_{-r_{mar}} + d(r_{mar}) + l_{max}\epsilon_{k+1} = \phi + l_{max}\epsilon_{k+1}$. This is not enough for satisfying $N(\boldsymbol{x}(t_{k+1})) - l_{max}\epsilon_{k+1} > \phi$, and thus $N_{-r_{mar}} + d(r_{mar}) > \phi$.

If there is $d(r_{mar})$ s.t. $N_{mar} - N_{-r_{mar}} \leqq d(r_{mar})$, the system eventually reaches to $N_{mar} + l_{max}\epsilon'(t') \geqq N(\boldsymbol{x}(t)) \geqq N_{mar} - l_{max}\epsilon'(t')$ at a finite time after t'. For example, if $N(\boldsymbol{x}(t)) < N_{mar} - l_{max}\epsilon'(t')$, it must be broken in finite time. For the sake of contradiction, suppose the inequality remains forever. Because of the best response dynamics and because sending one message is the strict best strategy for r_{mar} (Lemma 12), $x_1^{r_{mar}}(t) \to d(r_{mar})$ as $t \to \infty$. $x_1^{r_{mar}}(t)$ is strictly increasing in t and bounded, and thus $x_1^{r_{mar}}(t)$ converges to a point. For the other type and the other strategies, because of a unique optimal strategy for each type, they converge to a point. Therefore, $\boldsymbol{x}(t)$ converges to a point denoted by $\boldsymbol{x}^* \in X$. If $d(r) = N_{mar} - N_{-r_{mar}}$[49], $\boldsymbol{x}(t)$ converges to $\boldsymbol{x}^*$ s.t. $N_{mar} = N(\boldsymbol{x}^*)$. Thus, for any $\mu > 0$ there exists $\bar{t}$ s.t. $|N(\boldsymbol{x}(t)) - N_{mar}| < \mu$ for any $t > \bar{t}$. This contradicts our initial assumption that $N(\boldsymbol{x}(t)) < N_{mar} - l_{max}\epsilon'(t')$ remains after t'. If $d(r) > N_{mar} - N_{-r_{mar}}$, $\boldsymbol{x}(t)$ converges to $\boldsymbol{x}^*$ s.t. $N_{mar} < N(\boldsymbol{x}^*)$. However, under the best response dynamics, $\boldsymbol{x}(t)$ can converge to a point if and only if the point is a Nash equilibrium. $\boldsymbol{x}^*$ is not a Nash equilibrium[50], and thus we get the contradiction. We can apply the similar discussion to the case $N(\boldsymbol{x}(t)) > N_{mar} + l_{max}\epsilon'(t')$. Therefore, at finite time, $N_{mar} + l_{max}\epsilon'(t') \geqq N(\boldsymbol{x}(t)) \geqq N_{mar} - l_{max}\epsilon'(t')$ for any $t' > t_k$ happens.

A LM equilibrium $\boldsymbol{x}^*$ exists for γ_{k+1} and $d(r_{mar})$ s.t. $N_{mar} - N_{-r_{mar}} \leqq d(r_{mar})$. Since there can exist at maximum a single marginal type in LM equilibrium, for any $r \in R - \{r_{mar}\}$, $x_1^r = d(r)$ if $\pi(r) > \gamma_{k+1}$, and otherwise $x_0^r = d(r)$ in $\boldsymbol{x}^*$. Thus, the total amount of messages from all non-marginal types is $N_{-r_{mar}}$. Suppose in $\boldsymbol{x}^*$, $x_1^{r_{mar}} = N_{mar} - N_{-r_{mar}}$ and $x_0^{r_{mar}} = d(r_{mar}) - x_1^{r_{mar}}$. This $\boldsymbol{x}^*$ satisfies the all requirements for a LM equilibrium if $d(r_{mar}) \neq N_{mar} - N_{-r_{mar}}$. Because of the definition of γ_z, $\pi \neq \gamma_z$ for all $\pi \in \Pi$, and thus $d(r_{mar}) \neq N_{mar} - N_{-r_{mar}}$. In addition, $\phi < N(\boldsymbol{x}^*) < \phi / \frac{\phi}{\underline{N}}$. Therefore, there exists a LM equilibrium $\boldsymbol{x}^*$ for γ_{k+1}. Because LM equilibrium is locally asymptotically stable under the best response dynamics, if $\boldsymbol{x}^*$ is a LM equilibrium, there exists a neighborhood O around $\boldsymbol{x}^*$ in which the system is attracted to $\boldsymbol{x}^*$.

As $t \to \infty$, $\epsilon'(t) \to 0$, and thus all of the other types selects the optimal strategy in a LM equilibrium $\boldsymbol{x}^*$ at the limit of $\boldsymbol{x}(t)$ ($N_{mar} - l_{max}\epsilon'(t) \to N_{mar}$, $N_{mar} + l_{max}\epsilon'(t) \to N_{mar}$). Except $x_1^{r_{mar}}$ and $x_0^{r_{mar}}$, all x_l^r converges to the value in $\boldsymbol{x}^*$. In addition, from the previous paragraph, $N(\boldsymbol{x}(t))$ approaches to any neighborhood of $N(\boldsymbol{x}^*) = N_{mar}$ at least in a finite time. Therefore, $\boldsymbol{x}(t)$ eventually enters O, where O is the neighborhood of LM equilibrium $\boldsymbol{x}^*$ which has the local asymptotic stability under the best response dynamics. Therefore,

[49] Later, we see that this case does not happen because of our requirements on γ_z.

[50] some r_{mar} selects the suboptimal strategy 1.

$\boldsymbol{x}(t)$ converges to $\boldsymbol{x}^*$.

We show that, if $N(\boldsymbol{x}(t)) - l_{max}\epsilon \leqq \phi \leqq N(\boldsymbol{x}(t)) + l_{max}\epsilon$ for all $t \in [t_k, \infty)$ where ϵ satisfying $t(\frac{\epsilon}{n}) < (t - t_k)$ and $\bar{\epsilon} > \epsilon > 0$, the limit of $\boldsymbol{x}(t)$ is a LM equilibrium $\boldsymbol{x}^*$. This happens only if $\sum_{r \in R'} d(r) = \phi$ where $R' = \{r \in R | \pi(r) > \gamma_{k+1}\}$[51] (Otherwise, eventually the condition must be broken.). As same as the discussion for the case $N(\boldsymbol{x}(t)) - l_{max}\epsilon_{k+1} > \phi$, there exists a single type r_{mar} whose optimal strategy is fluctuated between 0 and 1. However, because $\pi(r_{mar}) > \gamma_{k+1}$, as $\epsilon \to 0$, there eventually exists a unique best strategy for each type. $\boldsymbol{x}(t) \to \boldsymbol{x}^*$, and in $\boldsymbol{x}^*$, $\phi = N(\boldsymbol{x}^*)$ and all active senders strictly select sending a message. Thus, $\boldsymbol{x}^*$ is a LM equilibrium.

When $k = 1$, if $N(\boldsymbol{x}(t_2)) + l_{max}\epsilon_2 < \phi$, there are no enough senders under γ_2' to fill the capacity ϕ, and thus we move to the next step by changing the cost from γ_2' to γ_3'.

If $N(\boldsymbol{x}(t_2)) - l_{max}\epsilon_2 > \phi$, this policy stops the cost change. In such a case, $\gamma_1' > \pi_{max} > \gamma_2'$ and $d(r) > \phi$ for $r \in R$ s.t. $\pi_{max} = \pi(r)$. For any other $r' \in R - \{r\}$, sending 0 messages is strictly dominant. Then, by the similar discussion for general $k+1$, the system eventually moves into the neighborhood of a LM equilibrium $\boldsymbol{x}^*$ where only r-type senders send a message and either $\pi_{max}\frac{\phi}{N(\boldsymbol{x}^*)} = \gamma_2'$ or $N(\boldsymbol{x}^*) = d(r)$ is satisfied.

If $N(\boldsymbol{x}(t)) - l_{max}\epsilon \leqq \phi \leqq N(\boldsymbol{x}(t)) + l_{max}\epsilon$ for all $t \in [t_1, \infty)$ where ϵ satisfying $t(\frac{\epsilon}{n}) < (t - t_1)$ and $\bar{\epsilon} > \epsilon > 0$, this implies $d(r) = \phi$ where $r \in R$ s.t. $\pi_{max} = \pi(r)$. We can apply the similar discussion for $k+1$ again. When we consider r as the marginal type, because $\pi(r) > \gamma_2$, as $\epsilon \to 0$, there eventually exists a unique best strategy for each type. Thus, $\boldsymbol{x}(t) \to \boldsymbol{x}^*$ where $\boldsymbol{x}^*$ is a LM equilibrium.

Next consider $\phi > D$. We can repeat the similar argument. Since we require a stricter condition $N(\boldsymbol{x}(t) - l_{max}\epsilon_z) > \phi > D$ for changing the cost in Step 2, Lemma 12 works even when $\phi > D$. In addition, (22) is never satisfied[52]. Then, there exists $z \in \{1, 2, 3, ..., z_{max}\}$ s.t. $N(\boldsymbol{x}(t)) - l_{max}\epsilon \leqq D \leqq N(\boldsymbol{x}(t)) + l_{max}\epsilon$ for all $t \in [t_z, \infty)$ s.t. $t(\frac{\epsilon}{n}) < (t - t_z)$ where $\bar{\epsilon} > \epsilon > 0$. This implies that $N(\boldsymbol{x}(t))$ converges to D. Since $D < \phi$, for each type, there exists $\bar{t}$ s.t. a unique optimal strategy (sending a message) exists for each type if $t > \bar{t}$. Therefore, the limit satisfies the definition of LM equilibrium for $\phi > D$.

In conclusion,

[51] From the definition, $\gamma_{k+1} \neq \pi(r)$ for any $r \in R$.

[52] This is because otherwise a marginal type $r_{mar} \in R$ s.t. $\gamma_{z+1} > \pi(r) > \gamma_z$ exists and $N_{-r_{mar}} + d(r_{mar}) > \phi$ by the discussion for (31). Because of Lemma 12, $N_{-r_{mar}}$ is constructed by $N_{-r_{mar}}$ amount of the population sending a message. This contradicts $\phi > D$.

Proposition 6. *From any initial state $\boldsymbol{x}(0)$, by the two-step cost change, $\boldsymbol{x}(t) \to \boldsymbol{x}^*$ and $\boldsymbol{x}^*$ is a LM equilibrium with $\bar{\sigma}$.*

A.7 Global stability of UC Equilibrium with the full-aid for first messages

We use *Globally Evolutionarily Stable State* (hereafter GESS) as a criterion of global stability. From Sandholm (2010),

Definition 12. *State $\boldsymbol{x} \in X$ is a globally evolutionarily stable state if*

$$(\boldsymbol{y} - \boldsymbol{x})' U(\boldsymbol{y}) < 0 \quad \forall \boldsymbol{y} \in X - \{\boldsymbol{x}\}. \tag{33}$$

This definition indicates that $\boldsymbol{x} \in X$ is GESS if $\boldsymbol{x}$ strictly invades any other $\boldsymbol{y} \in X - \{\boldsymbol{x}\}$.

Global stability of UC equilibrium with the aid depends on the amount of potential senders and the maximum benefit of senders. If all senders do not want to send a second message regardless of $\boldsymbol{x}$, UC equilibrium is a GESS. If the senders with the maximum benefit (π_{max}) do not want to send a second message, this condition is satisfied. Thus, if $D \geqq 2\phi$, the following condition is sufficient for GESS.

$$\gamma > \pi_{max} \frac{\phi}{D} \left(1 - \frac{\phi}{D}\right) \tag{34}$$

If D is large enough, the right-hand side goes to 0, and this condition[53] is satisfied. Therefore, UC equilibrium with the aid is a GESS.

Reference

Ali, S.N., 2018. Herding with Costly Information. Journal of Economic Theory 175, 713–729.

Anderson, S.P., de Palma, A., 2009. Information Congestion. RAND Journal of Economics, 40(4), 688–709.

[53]In addition, this condition is necessary for the state in which all senders send a message is a RTESS.

Anderson, S.P., de Palma, A., 2012. Competition for Attention in the Information (Overload) Age. RAND Journal of Economics, 43(1), 1–25.

Anderson, S.P., de Palma, A., 2013. Shouting to Be Heard in Advertising. Management Science, 59, 1545–1556.

Anderson, S.P., Peitz, M., 2022, Ad Clutter, Time Use, and Media Diversity. American Economic Journal: Microeconomics. Forthcoming.

Bagwell, K., 2007. The Economic Analysis of Advertising. Handbook of Industrial Organization, 3, 1701–1844.

Bandhu, S., Lahkar, R., 2022. An Evolutionary Interpretation of Dominant Strategy Implementation, Mimeo

Butters, G., 1977. Equilibrium Distributions of Sales and Advertising Prices. Review of Economic Studies. 44(3), 465–491.

Callander, S., 2011. Searching for Good Policies. American Political Science Review 105, 643–662.

Cisco, 2019. Email: Click with Caution - How to protect against phishing, fraud, and other scams. https://www.cisco.com/c/dam/en/us/products/collateral/security/email-security/email-threat-report.pdf

Dukes, Anthony J., Liu, Qihong, Advertising Content and Viewer Attention: The Role of Ad Formats (August 22, 2020). Available at SSRN: https://ssrn.com/abstract=3679335 or http://dx.doi.org/10.2139/ssrn.3679335

Dahlen, M., Rosengren, S., 2016. If Advertising Won't Die, What Will It Be? Toward a Working Definition of Advertising. Journal of Advertising 45, 334–345.

Eliaz, K., Spiegler, R., 2020. Incentive-Compatible Advertising on Nonretail Platforms. RAND Journal of Economics 51, 323–345.

Eppler, M.J., Mengis, J., 2004. The Concept of Information Overload: A Review of Literature from Organization Science, Accounting, Marketing, MIS, and Related Disciplines. Information Society 20, 325–344.

Falkinger, J., 2007. Attention economies. Journal of Economic Theory 133, 266–294.

Falkinger, J., 2008. Limited Attention as a Scarce Resource in Information-Rich Economies. The Economic Journal 118, 1596–1620.

Gilboa, I., Matsui, A., 1991. Social Stability and Equilibrium. Econometrica 59, 859.

Goh, K-Y., Hui, K-L., Png, I.P.L., 2015. Privacy and Marketing Externalities: Evidence from Do Not Call. Management Science 61, 2982–3000.

Hammer, P., Riebe, E., Kennedy, R., 2009. How Clutter Affects Advertising Effectiveness. Journal of Advertising Research 49, 159–163.

Hefti, A. 2018., Limited Attention, Competition and Welfare. Journal of Economic Theory 178, 318–359

Hirshleifer, J., 1973. Where Are We in the Theory of Information?. American Economic Review, 63(2), Papers and Proceedings of the Eighty-fifty Annual Meeting of the American Economic Association (May, 1973), 31-39

Johnson, J.P., 2013. Targeted Advertising and Advertising Avoidance. The RAND Journal of Economics 44, 128–144.

Lahkar, R., Mukherjee, S., 2019. Evolutionary implementation in a public goods game. Journal of Economic Theory 181, 423–460.

Lahkar, R., Mukherjee, S., 2021. Evolutionary implementation in aggregative games. Mathematical Social Sciences 109, 137–151.

Malhotra, N.K., 1984. Reflections on the Information Overload Paradigm in Consumer Decision Making. Journal of Consumer Research, 10(4), 436-440.

Niehaus, P., 2011. Filtered Social Learning. Journal of Political Economy 119, 686–720.

Rao, J.M., Reiley, D.H. 2012. The Economics of Spam. Journal of Economic Perspectives, 26(3), 87-110.

Renault, R., 2015. Advertising in Markets. Handbook of Media Economics, 2, pp. 121–204.

Riebe, E., Dawes, J., 2006. Recall of Radio Advertising in Low and High Advertising Clutter Formats. International Journal of Advertising 25, 71–86.

Rust, R.T., Oliver, R.W., 1994. The Death of Advertising. Journal of Advertising 23, 71–77.

Sandholm, W.H., 2002. Evolutionary Implementation and Congestion Pricing. The Review of Economic Studies 69, 667–689.

Sandholm, W.H., 2005. Negative Externalities and Evolutionary Implementation. The Review of Economic Studies 72, 885–915.

Sandholm, W.H., 2007. Pigouvian pricing and stochastic evolutionary implementation. Journal of Economic Theory 132, 367–382.

Sandholm, W.H., 2010. Population Games and Evolutionary Dynamics. MIT

Shen, Q., Villas-Boas, J.M., 2018. Behavior-Based Advertising. Management Science 64, 2047–2064.

Sims, C.A., 2003. Implications of Rational Inattention. Journal of Monetary Economics 50, 665–690

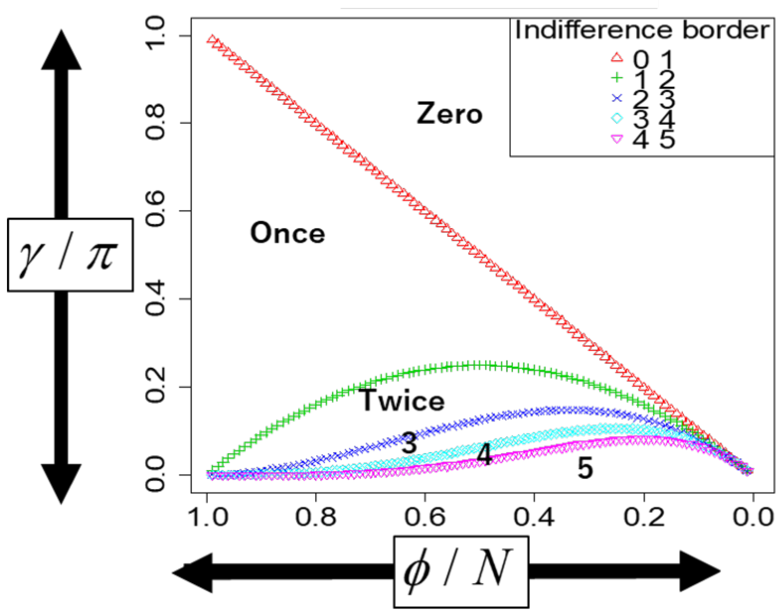

Figure 1: Best Response Diagram

Stigler, G.J., 1961. The Economics of Information. Journal of Political Economy 69, 213–225.

Symantec, 2019. Internet Security Threat Report Volume 24. https://docs.broadcom.com/doc/istr-24-2019-en

Taylor, J., Kennedy, R., Sharp, B., 2009. Is Once Really Enough? Making Generalizations about Advertising's Convex Sales Response Function. Journal of Advertising Research 49, 198–200.

Vakratsas, D., Feinberg, F.M., Bass, F.M., Kalyanaram, G., 2004. The Shape of Advertising Response Functions Revisited: A Model of Dynamic Probabilistic Thresholds. Marketing Science 23, 109–119.

Van Zandt, T., 2004. Information Overload in a Network of Targeted Communication. RAND Journal of Economics, 35(3), 542–560.

Wardrop, J., 1952. Some Theoretical Aspects of Road Traffic Research. ICE Proceedings: Part 2, Engineering Division 1, 325-378

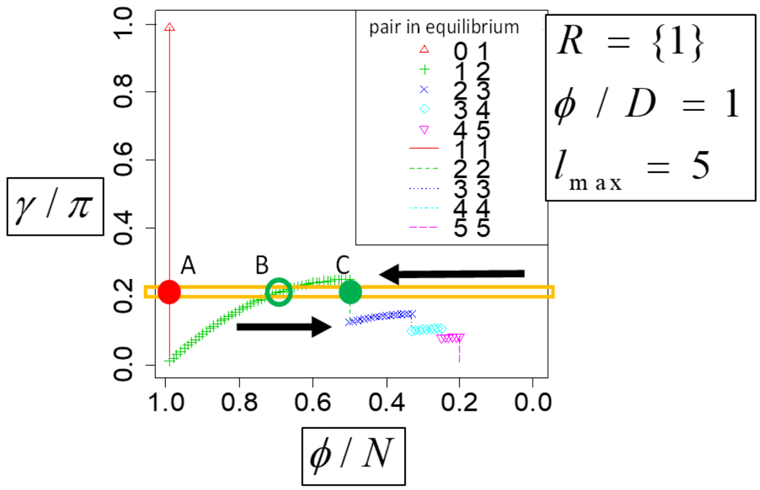

Figure 2: Bifurcation Diagram

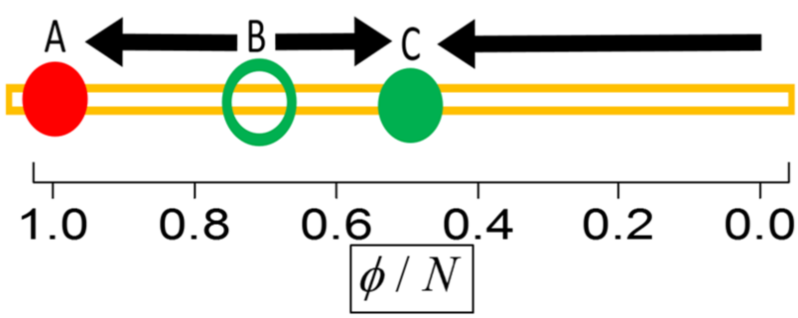

Figure 3: An Example of Stable Equilibria