

Hefti, Andreas; Lareida, Julia

Working Paper

Competitive attention, superstars and the long tail

Working Paper, No. 383

Provided in Cooperation with:

Department of Economics, University of Zurich

Suggested Citation: Hefti, Andreas; Lareida, Julia (2021) : Competitive attention, superstars and the long tail, Working Paper, No. 383, University of Zurich, Department of Economics, Zurich, <https://doi.org/10.5167/uzh-202542>

This Version is available at:

<https://hdl.handle.net/10419/234024>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



**University of
Zurich**^{UZH}

University of Zurich
Department of Economics

Working Paper Series
ISSN 1664-7041 (print)
ISSN 1664-705X (online)

Working Paper No. 383

Competitive Attention, Superstars and the Long Tail

Andreas Hefti and Julia Lareida

April 2021

Competitive Attention, Superstars and the Long Tail

Andreas Hefti* and Julia Lareida†

Abstract

We propose a model of competitive attention based on two key premises: i) People have limited information processing capacities and ii) consideration sets are formed according to relative salience. The equilibrium predictions we obtain can help to understand, and connect, diverse empirical phenomena, such as the Paradox of Choices, the Power Law dispersions of key market data (sales, profits, online clicks,...), the relation between advertising expenditures and market shares, the evolution of market inequality, or why evidence favoring a “Long Tail” effect is mixed at best.

Keywords Attention; Choice overload; Consideration Sets; Power Law; Superstar; Long Tail; Matthew Effect

JEL Classification D91, D40, D43, E71, L11, M37

*Andreas Hefti, Gertrudstrasse 15, CH-8400 Winterthur. Email: hefti@zhaw.ch. Web: <https://sites.google.com/site/heftinomics>. Author affiliations: I) School of Management and Law, Zurich University of Applied Sciences, II) Department of Economics, University of Zurich. Our special thanks go to the participants of the Sloan-Nomis-workshop on Attention and Choice 2019, and in particular to Ernst Fehr, Antonio Rangel and Andrew Caplin. Presented at seminars in Zurich, Shanghai (SUFU and ShanghaiTech), Nice, UECE Lisbon meeting 2019, SAEe 2019, EEA-ESEM 2020 and World Congress of the Econometric Society 2020

†Department of Economics, University of Zurich

1 Introduction

The modern, digitized competition looks like a textbook illustration for perfect competition. Markets seem to become increasingly contestable as expensive “brick-and-mortar” stores are replaced by cheaper online outlets, facilitating product entry. Consumers face an ever growing plethora of choice options, escorted by an omnipresence of product information. However, information becomes relevant only insofar it enters cognition. As mental resources are limited and information-processing consumes energy, only a subset of the available information may be passed on to cognition. The corresponding information selection is commonly referred to as an *attention process* in psychology (see, e.g., Pashler, 1998).

In this article, we study a setting with competing choice options, where the consumer attention process is described by two key factors emphasized by psychological and neuroscientific research: (i) People have limited information processing abilities, and (ii) their allocation of attention obeys stimulus-driven attentional control as described by Steven’s Law of Perception (Stevens, 1957). The equilibrium model that we derive from these presumptions can explain empirical phenomena observed in individual choices or in aggregate industry data, and connects these phenomena through the competitive allocation of attention.

First, we show that a “Paradox of Choices” emerges for individual consumers, meaning that choice satisfaction and the set of considered options diminish once the information load surrounding the choice options increases beyond a certain threshold. Choice overload arises in our model because the human brain is endowed with finite resources for information processing, and consequently needs to economize on the mental resources directed towards a decision task. How effective the brain is in processing the available information depends on the brain’s capacity relative to the information load that it needs to handle. The larger this load, the slower information processing becomes, much like more traffic on a given road causes congestion, reducing the traveling speed of each individual vehicle. The brain then economizes on its resources by “looking less carefully”, which is manifested in smaller consideration sets.

Second, we show that key market data, such as sales, profits or attention (e.g., online “clicks”) ought to follow Power Law dispersions in equilibrium, as a consequence of competitive forces and Steven’s Law of Perception. This empirically well-founded law describes the relation between the objective intensity of a stimulus (the luminance of a figure, the volume of a sound,...) and the subjective intensity of the mental sensation it evokes on the mind of the recipient. We embed its key observation – that equal stimulus ratios generate equal mental sensation ratios – in our

model by assuming that firms can influence the relative salience with which their products are imprinted on a consumer’s mind by investing into attention-seeking efforts, such as advertising. The choice alternatives which manage to evoke the strongest mental sensations relative to the rest have the largest chance of entering the consideration set. Firms are individually rational when choosing their attention efforts and correctly account for the consideration set formation by consumers. The equilibrium allocation of attention therefore is not random, but systematically reflects the competitive behavior of profit-oriented firms that differ in their abilities to attract (or to monetize) attention. Steven’s Law then implies that the competitively chosen attention efforts must be related by the principle of constant proportionality to each other, which necessarily results in a Power Law dispersion of the equilibrium quantities.

In addition, we find an intimate relation between the dispersion of perception chances and the average size of the consideration sets, as the latter must correspond to the sum of the products’ perception chances. This simple insight allows us to find the consideration set size of a representative consumer that generates this data. The perception chances can principally be deduced from attention data, such as an aggregate online click dispersion. Moreover, with an additional assumption we can even obtain the full dispersion of this representative consumer’s consideration sets that are consistent with the attention data.

Third, we study whether competitive attention works in favor of “Superstar products” or a “Long Tail” effect. According to the Long Tail hypothesis, the products in the tail of the sales distribution should benefit most from the improved economies of storage and distribution resulting from digitized markets, leading to a gradual de-concentration, e.g., in the sales distribution (Anderson, 2004). By contrast, theories of Superstar products suggests that the head of the distribution grows more successful due to quality advantages, thereby increasing market inequality (Rosen, 1981; Frank and Cook, 1995). Both theories do not account for consumer attention and its allocation. We find that competitive attention induces a Superstar effect in the empirically relevant cases where new products enter at the tail or total demand increases. Moreover, the same forces that cause a Superstar effect also inhibit a Long Tail effect. That is, the entry of new tail products tends to *increase* market concentration, and the firms with the largest market shares may benefit most, which contrasts with conventional competition results.

Our findings reflect the equilibrium interplay between competitive forces and the individual formation of consideration sets. Intuitively, the market forces forming the competition for attention can be described as firms shouting more for attention, while consumers respond by hearing less, which inclines individual firms to shout even louder. More precisely, firms compete in attention for becoming part of the consumers’ consideration sets. Product entry or an increase in demand intensifies this competition and increases the information load imposed

on consumers, which aggravates the choice overload effect. A stronger choice overload effect, in turn, intensifies the competition for attention as attention has become scarcer. While this increases attention expenditures, and more so for firms with larger market shares, it also allows firms to better monetize the attention they manage to attract due to a competition softening effect associated with smaller consideration sets. We show that the latter effect dominates if products are complex, such that their evaluation consumes a lot of mental resources, and strong substitutes in terms of true preferences. In such a case the firms that can afford the largest market shares must also be the ones to benefit most, in absolute terms, from product entry at the tail.

From the conceptual perspective, our framework offers a tractable theory of consideration set formation. Consideration sets have been a key concept in marketing, as these sets are thought of containing those alternatives from which a consumer makes her final purchase decision. We assume that consideration sets, as the central output of the information processing problem, are formed by the brain such that they contain as much information as can be analytically parsed. With the latter, we mean that consideration sets have been formed exactly such that consumers can identify the alternative that maximizes a fixed preference relation without mistakes. By contrast, the formation of a consideration set is cognitively more challenging, and may therefore involve an entirely different mental process, as emphasized by the relatively young literature about the “power of the subconscious” for decision-making in case of complex problems (Dijksterhuis et al., 2006). In this respect, our model draws a line between the mental operations involved in “choosing an option from a consideration set”, and “forming a consideration set”.

Article structure The attention model is developed in Section 2, and Section 3 derives the key properties of the attention equilibrium. In Section 4 we study whether a growing measure of tail firms or market entry of new consumers cause a Superstar or a Long tail effect in the attention model. Section 5 discusses the related literature and contains additional comments. Section 6 draws a conclusion. All proofs of the formal statements in the main text can be found in Appendix A. The supplementary material (Appendix B) contains additional results and extensions meant to strengthen and further explore some of our key points. The respective connections will be made explicit in the main text.

2 Choice Overload and Competitive Attention

On many occasions, decision-makers do not consider all available options when making a choice. Starting with Miller (1956), research in different disciplines has found clear evidence that people

fail to properly recognize all information items once total information exposure exceeds a certain threshold. Such observations are frequently summarized under “information overload”, meaning that somebody is exposed to “more information than the sensory or cognitive system can handle”.¹ In marketing, the idea that consumers choose from a “consideration set”, containing only a few out of all available alternatives, dates back to Howard and Sheth (1969). Hauser and Wernerfelt (1990) present a broad description of such consideration sets, emphasizing that such sets empirically are small ($\approx 4 - 5$ items) across very different product categories (soaps, cars,...). Similarly, De los Santos et al. (2012) observe that shoppers visit only a small number of websites for their online purchases. In social network research, several papers show that users fail to account for all incoming messages due to constraints on information processing, meaning that they only consider a subset of the information available to them when deciding which messages to forward or promote (see, e.g., Hodas and Lerman, 2014; Rodriguez et al., 2014; Feng et al., 2015).²

In the *finite capacity model*, developed in the next section, we formalize information overload and its connection to economic choice based on the well-established observation that humans are able to process the stream of available messages only with finite mental capacities. The outcome is a fairly tractable model of consideration set formation consistent with research on how the human brain appears to work. Intuitively, the consideration set is the result of a possibly subconscious deliberation of the brain that attempts to economize on its resources for information processing, and the relative salience of the various information stimuli that need to be processed.

2.1 Finite Capacity Model

Suppose that there are $n \in \mathbb{N}_+ \cup \{\infty\}$ options available in a decision task. For a decision-maker, these options matter only insofar they are represented on her mind, which requires mental information processing. How many of the choice option are perceived, i.e., become a part of the consideration set, depends on how much mental resource is dispatched, and how effective the brain can process the available information.

Regarding the latter, we suppose that the processing rate $\tau > 0$ quantifies how effective an additional unit of mental resources is in the production of a consideration set for a given decision situation. Specifically, if R measures consideration set size, we suppose that $R = \tau T$,

¹See Eppler and Mengis, 2004; Sutcliffe and Weick, 2009 for interdisciplinary surveys of information overload. The term “information overload” was coined by Toffler in the book “Future Shocks” (Toffler, 1984).

²For example, in case of Twitter Rodriguez et al. (2014) estimate the threshold to be about 30 incoming tweets per hour. Below this inflow rate, the likelihood that a users re-tweets an incoming tweet is constant, but above the threshold it drops substantially.

where $T \geq 0$ quantifies the amount of mental resource that has been dispatched by the brain for information processing. Lower values of τ or T thus imply, *ceteris paribus*, that fewer alternatives are perceived, and become part of the decision-problem.³ Intuitively, this means that the set of perceived options is smaller if the brain thinks less hard, or information processing is challenging.

We suppose that the human brain economizes on the mental resources T allocated to the decision task for a given processing rate $\tau > 0$ by solving

$$\max_{T \geq 0} \tau T - H(T), \quad s.t. \quad \tau T \leq n, \quad (1)$$

where $H(\cdot)$ is a strictly increasing, strictly convex C^2 -function with $H'(0) = 0$. We think of $H(\cdot)$ as the cost of allocating T units of mental resources to the decision problem, e.g., in terms of metabolic energy used by the brain (Lennie, 2003), or in form of a general opportunity cost (see Appendix B.3). The constraint in (1) captures that $R \leq n$, where $R = n$ means that the consideration set encompasses all available information. If $\varphi(\cdot)$ denotes the inverse of the function $H'(\cdot)$, and ignoring the integer-value problem,⁴ the optimal attention span T^* solves $T^* = \varphi(\tau)$, such that

$$R^* = \tau \varphi(\tau). \quad (2)$$

whenever $R^* < n$, i.e., an interior solution results (which occurs iff $\tau > H'(n/\tau)$). Condition (2) implies that consideration set size and the processing rate are positively related ($R'(\tau) > 0$) at interior solutions. That is, it is optimal to *decrease* the attention span once information processing is *slower*.⁵

Consistent with empirical evidence and mathematical information theory, we assume that the processing rate τ is inversely related to the total information load $\Sigma(n) \in \mathbb{R}_+$ the brain needs to handle when facing n alternatives. Specifically, we let

$$\tau = \min \left\{ \frac{C}{h(\Sigma)}, 1 \right\}, \quad (3)$$

where $h(\cdot) \geq 0$ is a strictly increasing C^1 -function. Expression (3) has its roots in information theory (Shannon, 1948), where C is a fixed channel capacity, and $h(\Sigma)$ quantifies the information

³See, e.g., Bettman et al. (1998) showing that the imposition of an external time constraint on the length of the information evaluation period forces consumers to engage in a less systematic, and hence less informative, evaluation of the available alternatives.

⁴To characterize the properties of the optimal solution R^* , it is analytically convenient to admit real-valued solution in problem (1). While allowing that $R \in \mathbb{R}_{++}$ greatly simplifies the formal analysis, in particular in our later equilibrium model, we do not see that the essential economic implications we derive hinge critically on this simplification.

⁵ $R'(\tau) > 0$ would also result if the attention span were exogenously fixed to $\bar{T} > 0$.

load that needs to be processed through the channel.⁶ We denote by $\bar{R} \equiv \varphi(1) > 0$ the maximal amount of information that can be processed, achieved if the brain works at full capacity, and accordingly normalize $C \equiv h(\bar{R})$. For simplicity, we normalize the minimal information load imposed by each option to one, which implies that $\Sigma(n) \geq n$.⁷

By (3), we assume that τ depends on a measure of information load $h(\Sigma)$, rather than just on the mere number of alternatives n . This reflects an overarching consensus that the information surrounding the choice options (e.g., marketing messages, the number of advertised attributes or the overall information dimensions of the options) and not the pure size of the grand choice set is the central determinant for the information (over-)load a decision-maker is exposed to (e.g., Jacoby, 1977; Payne, 1982; Payne et al., 1993; Lurie, 2004; Lee and Lee, 2004; Scheibehenne et al., 2010; Chernev et al., 2015).

Condition (3) implies that whenever the processing requirement $h(\Sigma)$ exceeds the capacity limit C , information overload occurs, resulting in information congestion and a reduced rate of information processing τ . Such a pattern has been robustly observed in humans.⁸

2.1.1 Choice Overload

We say that *information overload* occurs if one of the following equivalent conditions is met: $h(\Sigma) > C$, $\Sigma > \bar{R}$, or $\tau < 1$. That is, information load occurs iff information is processed below the maximal pace. It is easy to verify that $R^* < \bar{R}$ iff there is information overload, and that $\tau'(\Sigma) < 0$ whenever there is information overload.

Choice overload occurs whenever $R^* < n$, i.e., if less information is processed than the decision-maker is confronted with. The following elementary proposition shows that information overload is necessary for choice overload, and predicts a negative relation between consideration set size and information exposure of a choice-overloaded decision-maker.

Proposition 1 *Information overload is necessary for choice overload, and $R'(\Sigma) < 0$ whenever there is choice overload.*

Intuitively, if there is no information congestion ($\tau = 1$), then $\bar{R} \geq n$, such that it is optimal

⁶If τ is the transmission rate of a channel per unit of time, and an information source has a complexity as measured by $\text{Log}_2(\Sigma)$ of symbols, then $N(T) = \Sigma^{T\tau}$ corresponds to the possible inflow of information if the channel is open for duration T and symbols arrive randomly. The Shannon capacity is then defined as the limit $C = \lim_{T \rightarrow \infty} \frac{\text{Log}_2(N(T))}{T}$, which evaluates to $C = \tau \text{Log}_2(\Sigma)$. Thus, for a given capacity C , the transmission rate must be $\tau = C/\text{Log}_2(\Sigma)$.

⁷In the equilibrium model, the exposure $\Sigma(n)$ will be endogenously determined by the aggregate attention-seeking efforts, which can intuitively be thought of as the total number of messages sent to the consumer by the competing senders.

⁸Schroder et al. (1967) examine a large body of experimental evidence, and conclude that information load and information processing must be interdependent, where an increasing information load negatively influences an individual's processing capacity beyond a certain tipping point; also see Streufert and Driver (1965); Streufert et al. (1967); Streufert (1970), and more recently Eppler and Mengis (2004), concluding that a slowdown in information processing is a robust characteristic of information overload across many different decision tasks.

to process all available information, and a choice overload effect cannot arise. For the second result, an increase in the information load Σ intensifies information congestion given that choice overload already has arisen, and the corresponding reduction in information processing τ makes it optimal for the brain to dispatch less mental resources, resulting in smaller consideration sets. To illustrate Proposition 1 with an example, let $h(\Sigma) = \Sigma$ and $H(T) = \delta T^\mu$ with parameters $\delta > 0$ and $\mu > 1$. It is easily verified that $R(\cdot)$ then is of the form

$$R = \alpha \Sigma^{-\beta}, \quad \alpha > 0, \beta \equiv \frac{\mu}{\mu - 1} > 1, \quad (4)$$

whenever choice overload occurs.

If $\Sigma(n)$ is strictly increasing in n , a property which holds in our later equilibrium model, the finite capacity model makes the behavioral prediction that a *choice overload effect* occurs, in thus that R^* follows an inverse-U in the number of available options n (see Figure 1).

Corollary 1 *Let $\Sigma(n)$ be a strictly increasing C^1 -function with $\Sigma(0) = 0$. Then there exists a unique $\bar{n} \leq \bar{R}$, such that $R^* = n$ for $n < \bar{n}$ and $R'(n) < 0$ for $n > \bar{n}$.*

The intuition is that an increase of n at some point must eventually lead to information overload whenever an increasing number of choice options also increases the information load. Beyond this threshold, a further increase of n leads to information congestion, resulting in a decreasing transmission rate and a decreasing attention span.

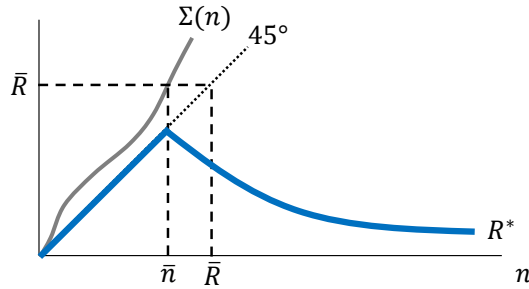


Figure 1: Choice Overload Effect

2.1.2 Discussion

Proposition 1 and Corollary 1 entail the essential prediction that information recipients “look less carefully” once overload occurs. The prediction of a diminished consideration in response to information overload is consistent with a large body of empirical evidence on what has been summarized as the negative effects of choice overload, sometimes also referred to as the “Paradox of Choices”. Eppler and Mengis (2004) present an extensive interdisciplinary survey

and conclude that an overloaded person has “difficulties in identifying the relevant information, becomes highly selective and ignores a large amount of information...” (p.333). Parra and Ruiz (2009) conduct a controlled experiment with a simulated online store for stereo systems, and show that an increase in the information load, measured jointly by the number of products and product attributes, leads to a reduction in the consideration set size. Pan et al. (2013) conduct eye-tracking studies in case of hotel selection, where the size of the information set (number of displayed hotels) is varied from 5 to 20. They find roughly double as many eye fixations on each hotel in the small information set, meaning that each hotel in the small set received much more processing. Moreover, many subjects missed to look at some hotels in the large set, but not so in the small set. Rodriguez et al. (2014) show with social network data from Twitter that the more incoming tweets an overloaded user is exposed to, the less information sources are considered by this users for re-tweeting. A diminished consideration in response to increasing information exposure has also been observed in case of catalog-based shopping (Eastlick et al., 1993), apartment selection (Helgeson and Ursic, 1993), online information search (Oulasvirta et al., 2009), the number of partner considers in online dating platforms (Finkel et al., 2012), click-through rates in news articles (Dellarocas et al., 2015), or online shopping (Li, 2016).

Besides diminished consideration, choice overload is manifested through decision-makers expressing general sentiments of choice dissatisfaction or insecurity with the choice, see, e.g., Malhotra (1982); Iyengar and Lepper (2000); Botti and Iyengar (2004); Oulasvirta et al. (2009); Haynes (2009); Inbar et al. (2011); Chernev et al. (2015). A simple Envelope-Theorem argument shows that the value function $\mathcal{V} = \tau T - H(T)$ in (1) is strictly decreasing in Σ for a choice-overloaded decision-maker, consistent with these negative sentiments.⁹

Complexity The cognitive complexity of a task is a central empirical aspect of choice overload (Chernev et al., 2015). According to Payne et al. (1993), the cognitive difficulty of a decision task amounts to conceptual or structural challenges, which likely are independent from the values of the option at stake. In the finite capacity model, the cognitive complexity of a task is incorporated by the (task-specific) function $h(\Sigma)$ in (3). To illustrate, suppose that $h(\Sigma) = \Sigma^\kappa$, where a larger value of the parameter $\kappa \geq 1$ indicates a more complex decision task. For $H(T) = \delta T^\mu$, (4) then is of the form $R(\Sigma) = \alpha \Sigma^{-\kappa\beta}$, implying that $\frac{\dot{R}(\Sigma)}{R(\Sigma)} = -\frac{\kappa\beta}{\Sigma}$ for an overloaded subject. Hence, considerations sets diminish more quickly in the information load, the more complicated the decision task is, *ceteris paribus*.¹⁰

⁹This also holds if the value function were an indirect utility function; see Appendix B.3.

¹⁰In view of this example, we can thus directly interpret a larger value of β in (4) as indicating a cognitively more demanding decision.

Consciousness One might ask whether our results depend on the fact that R enters linearly into the objective function (1). We show in Appendix B.3 that the linearity assumption is not decisive for the choice overload phenomenon to arise, nor for our later equilibrium predictions.¹¹ Nevertheless, the formation of a consideration set most likely is far more complex compared to “choosing an option” from a fixed and potentially small consideration set. Therefore, a model based on (1) may lead to a more reasonable description of human behavior as if one and the same mental processes were forced on each part of the decision-making process.

In this respect, a recent literature has argued that in case of utterly complex optimization problems the brain makes a subconscious guess, rather than analytically working all possible contingencies which, *inter alia*, requires to (correctly) anticipate all possible future contingencies that can arise from the choices made (Dijksterhuis, 2004).¹² For example, even the most proficient human chess players must rely on their educated intuition, rather than conducting a full analytical assessment, in sufficiently complex board situations. The simple form of (1) with $R = \tau T$ states that the brain associates a proportional value with an increased consideration set size R , which seems like a reasonable forward guess, which may be particularly reasonable if the possible alternatives in the consideration sets do not involve critical stakes (“movies rather than houses”).

The consideration sets in our model are formed such that they contain exactly as much information that can be analytically parsed. In particular, this means that consumers will be able to evaluate their (fixed) preference relation for the options in their consideration set. In this sense, we draw a line between the mental operations involved in decisions within a simple set of objects, and the more complicated operations required to reduce a problem to such a simple set. More generally, this procedure is compatible with the general notion of a two-phased decision process, see, e.g., Payne (1982); Gensch (1987); Hauser and Wernerfelt (1990). In a “System 1 and System 2” context (Kahneman, 2011), one can interpret the finite capacity mode as stating that the fast-thinking system prepares the information such that the slow-thinking system can penetrate it.

2.1.3 Allocation of Attention

If the set of available choice options is $\mathcal{N}_n \equiv \{1, 2, \dots, n\}$ and a consumer’s consideration set size is $R \in \mathbb{N}_+$, the space of all possible consideration sets is

$$\mathcal{A}(n, R) \equiv \{A \subset \mathcal{N}_n : |A| = \min\{n, R\}\}, \quad (5)$$

¹¹For example, we consider the case, where $\mathcal{V} = z(R) - H(T)$, $R = \tau T$, and $z(R)$ corresponds to an expected indirect utility function.

¹²We discuss the relation of this literature to our approach in Appendix B.3.

which we abbreviate by \mathcal{A} . It is easily verified that $\mathcal{A} = \mathcal{N}_n$ iff $R \geq n$. In particular, $\mathcal{A} = \mathcal{N}_n$, $\forall n$, results iff the standard premise of unlimited mental capacities ($R = \infty$) applies.

In psychology, attention is commonly defined as the *gating mechanism* selecting which information items are processed, and which excluded from consideration, if a recipient is exposed to more information than its sensory system can handle (see Pashler, 1998 for a survey). In our model, we describe this selection problem by a probability function $P_{\mathcal{A}} : \mathcal{P}(\mathcal{A}) \rightarrow [0, 1]$, where $\mathcal{P}(\mathcal{A})$ denotes the power set of \mathcal{A} . For any $A \in \mathcal{A}$, $P_{\mathcal{A}}$ quantifies the likelihood that consideration set A is formed.¹³ We refer to probability space $\mathbb{A} \equiv (\mathcal{A}, \mathcal{P}(\mathcal{A}), P_{\mathcal{A}})$ as an *attention space*, and to $P_{\mathcal{A}}$ as an *allocation of attention*.

Given any attention space, one can derive, for each choice option $j \in \mathcal{N}_n$, the probability π_j of being in a consideration set. If $B_j \equiv \{A \in \mathcal{A} : j \in A\}$ denotes the set of all consideration sets in \mathcal{A} containing option j , then $\pi_j \equiv P_{\mathcal{A}}(B_j) = \sum_{A \in B_j} P_{\mathcal{A}}(A)$. In our later model, π_j is the perception probability of a product, where π_j is determined by the *relative salience* of all competing choice options.

We end this section by stating a fundamental connection between the perception probabilities π_j with the consideration set size R that generates \mathcal{A}

Lemma 1 *Let \mathbb{A} be an attention space. Then $\sum_{j \in \mathcal{N}_n} \pi_j = \min\{R, n\}$.*

Thus, R always amounts to the sum of all individual perception chances. Let $n > R$ and define $\hat{\pi}_j \equiv \pi_j/R$ as the perception chance of j per unit of attention. Then, Lemma 1 implies that $\sum_j \hat{\pi}_j = 1$. This simple observation together with the fact that $\pi_j = \hat{\pi}_j R$ will play a central role in the equilibrium analysis. Moreover, as we show in Appendix B.4, this relation can be further exploited if one seeks to identify the allocation of attention $P_{\mathcal{A}}$ from empirical data (such as an online “click” dispersion).

2.2 Competitive Attention and Market Competition

In the following, we develop the notion of competitive attention based on the above attention framework. There is a measure I of consumers and $n > 1$ single-product firms. Let $\mathbb{A} = (\mathcal{A}, \mathcal{P}(\mathcal{A}), P)$ be the attention space of an average consumer where, for the moment, we exogenously assume that $n > R$, such that there is choice overload, and thus a non-trivial attention selection problem arises.

In Section 2.2.1 we define the average revenues that the firms can earn from any given attention they attract, and relate their chances of perception to attention data, such as an

¹³If $R \geq n$, then $\mathcal{A} = \mathcal{N}_n$, meaning that there is no selection problem, and the only possible probability function $P_{\mathcal{A}}$ is degenerate.

online “click” dispersion. In Section 2.2.2, we endogenize the firm-side attention by assuming that firms need to compete for attention in terms of relative salience, and show in Section 2.2.3 that Steven’s Law of perception pins down the relevant attention cost structure.

2.2.1 Attention Revenues and “Clicks”

Suppose that firm j can extract an attention revenue $V^j(A)$ from consideration set $A \in \mathcal{A}$, where the value function $V^j : \mathcal{A} \rightarrow \mathbb{R}_+$ verifies $V^j(A) > 0$ only if $A \in B_j$. In words, consideration is necessary for monetization. Thus, the expected attention revenue earned by firm j in the market is $\sum_{A \in B_j} P(A)V^j(A)I$.¹⁴ In what follows, we assume that for any given $\mathcal{A}(n, R)$ with $n > R$, firm j earns the same average revenue from each $A \in B_j$, such that $V^j(A) = V^j(A')$, $\forall A \in B_j \subset \mathcal{A}(n, R)$. Accordingly, we write (abusing notation) $V^j(R)$ instead of $V^j(A)$, and interpret $V^j(R)$ as the average attention revenue per consumer that firm j earns from consideration sets of size R . Then, the total expected attention revenue simplifies to

$$\sum_{A \in B_j} P(A)V^j(R)I = \pi_j IV^j(R), \quad (6)$$

which greatly increases the tractability of the equilibrium analysis. As we show in Appendix B.7, one foundation for the reduced form (6) is that $V^j(R)$ amounts to the average per-consumer revenue earned from strategic price competition, where $V^j(R) = (p_j - c)d^j(p, R)$ and $d^j(\cdot)$ a (conditional) demand function.

Clicks The quantity $C_j \equiv \pi_j I$ in (6) measures the total amount of attention attracted by firm j . To illustrate, suppose that the attention of each consumer is manifested by her “views” or “clicks” at different options, such that RI measures the total volume of clicks in the market. Then, firm j ’s chance of obtaining a click is π_j , the total measure of clicks going to j is $\pi_j I$, and the fraction of clicks going to j corresponds to $\hat{\pi}_j$, the perception chance per unit of attention.¹⁵

¹⁴In this expression, $P(A)$ and $V^j(A)$ correspond to population averages which, by assumption, is the only population information available to firms. Moreover, this expression would also apply if there were consumer-side heterogeneity in the attention allocation $P_i(\cdot)$ and the value function $V_i^j(\cdot)$ for given R , where these distributions are known by the firms, as long as P_i and V_i^j are uncorrelated in the consumer population. Then, $P(A)$ and $V^j(R)$ again would correspond to population averages. A simple example is obtained if either P_i or V_i^j is identical across consumers.

¹⁵Such “click” data is typically available in internet data (see, e.g. Optify.net. In Section B.4 we discuss how to retrace the attention space \mathbb{A} that matches the average consumer from such data.

2.2.2 Relative Salience

We suppose that the firms can influence their perception chances π_j by investing into attention-seeking activities, quantified by $f_j \geq 0$. Specifically, π_j is determined according to

$$\pi_j = \frac{f_j}{\Sigma} R, \quad (7)$$

where $\Sigma \equiv \sum_j f_j$ is the information load generated by the attention-seeking of all firms. An ad-hoc interpretation of (7) is that $\frac{f_j}{\Sigma}$ is the chance of being sampled in a single random draw from an urn, where f_j is the mass of ball j . Then, (7) says that the perception chance is proportional to the mass of ball j relative to all balls, and to the total number of draws R .¹⁶ That is, if one thinks of f_j as collating the total volume of firm j 's advertising campaign imposed on consumers, the chance of firm j to be “on a consumer’s mind” at the moment where the consumer makes her final purchase decision is proportional to the size of j 's campaign relative to total campaigning in the market.

Our main foundation for stipulation (7) that “standing out relative to the rest” determines ones perception chances follows from the psychological principle of *relative salience*. The literature has established that the salience value of an impression is determined by the strength of the impression *relative* to other impressions or a general background. Several papers document that items which are relatively more visible, e.g., by their prominence in list positions, their relative appearance (luminance, size, originality,...) or their relative frequency of occurrence, are more likely to be perceived.¹⁷

Expressed in our model, relative salience requires that f_j matters *relative* to the total information load Σ for determining the perception chance π_j . Accordingly, if $\pi_j = \pi_j(f_j, \Sigma)$, then the function $\pi_j(f_j, \Sigma)$ must be zero-homogeneous in (f_j, Σ) . As $\pi_j \equiv \hat{\pi}_j R$, also the perception chance per unit of attention, $\hat{\pi}_j$, must be zero-homogeneous in (f_j, Σ) . Thus $\hat{\pi}_j(f_j, \Sigma) = g(f_j/\Sigma)$, and in Appendix B.1 we prove that $g(\cdot)$ is linear, i.e., $g(f_j/\Sigma) \equiv \frac{f_j}{\Sigma} = \hat{\pi}_j$. Therefore, π_j must be given by (7).

¹⁶This argument can be made formally correct in the sense of an approximation. Let $\hat{\pi}_j = f_j/\Sigma$ be the one-shot sampling chance to draw j from an urn. The probability *not* to get any draw out of R independent draws with replacement then is $(1 - \hat{\pi}_j)^R$. Thus, the probability of obtaining (at least) one draw is $\pi_j = 1 - (1 - \hat{\pi}_j)^R$, the first-order approximation of which is $\pi_j = \hat{\pi}_j R$ at $\hat{\pi}_j \approx 0$.

¹⁷See, e.g., Smith and Brynjolfsson (2001); Hothkiss (2006); Ghose and Yang (2009); Baye et al. (2009); Ellison and Ellison (2009); Dellarocas et al. (2015) for evidence that the relative salience of various online links influences the attention they grab, or Hodas and Lerman (2013, 2014) for evidence that the relative visibility of online information explains which information is passed through social networks. Several marketing studies document the importance of “top-of-mind advertising” (see, e.g., Iyer et al., 2005; Hefti and Heinke, 2015; Astorne-Figari et al., 2019). At a lower level, such observations match the fact that the motion, color or luminance of an object matters relatively to the surrounding of the object in abstract visual search experiments (Pashler, 1998; Nothdurft, 2000; Treue, 2003; Pooresmaeli et al., 2014); likewise see Kahneman and Henik (1981); Yantis (1998); Maunsell and Treue (2006) for the “spotlight effect” of attention.

2.2.3 Attention Costs and Payoffs

Let attracting attention be a costly activity, captured by a cost function

$$C_j(f) = c_j f^\eta, \quad c_j > 0, \eta > 1. \quad (8)$$

As the revenue of attracting attention is (6), firm j 's payoff must be

$$\Pi^j(f_j) = \frac{f_j}{\Sigma} R V^j(R) I - c_j f_j^\eta. \quad (9)$$

The payoff function (9) shows that attention is a competitive resource, akin to a multi-prize contest, where $V^j(R)$ is the value j earns from obtaining one of R prizes, and f_j/Σ is the chance of seizing such a prize.¹⁸ Cost function (8) is not arbitrary. In Appendix B.2 we provide a technology-based and a psycho-physical foundation for (8). The main arguments are as follows.

The psycho-physical foundation derives (9) from *Steven's Law of perception* (Stevens, 1957). This empirical law relates the (objective) intensity of a sensory stimulus (sound, motion, luminance,...) to the intensity of the mental impression it evokes on the mind of an observer. In this interpretation, $f_j = f_j(e_j)$ is the intensity of the mental impression generated by exposing a consumer to an objective sensory stimulus of intensity e_j . Based on numerous observations, the law states that equal objective stimulus ratios generate equal subjective sensation ratios in an observers' minds, meaning that $f_j(e_j)$ must be of the bijective form $f_j(e_j) = k_j e_j^{1/\eta}$. Let the unit cost of producing an objective stimulus be normalized to one, such that e are the costs associated with sending at intensity e . Rational firms ultimately care only about how much attention they attract, i.e., how much sensory inputs it takes to achieve a mental impression that yields a certain perception chance. Then, expressing the costs of e_j in terms of the induced mental impressions f_j , where $e_j = f_j^{-1}(f_j)$, Steven's Law implies that the payoff structure must be given by (9). In this context, the assumption $\eta > 1$ captures that attention-seeking is subject to diminishing returns. Such diseconomies of scale are a well-known empirical regularity in case of advertising (e.g., Vakratsas and Ambler, 1999; Bagwell, 2007)

The technology-based foundation presumes that attention-seeking (e.g., advertising) is produced by a multi-input production process described by a homogeneous production function, where firms differ in their efficiency of producing attention. In this interpretation, the parameter η captures the economies of scale in the production of attention, and c_j capture efficiency

¹⁸We assume that $V^j(R)$ is independent of f_j , as our main focus is on "prominence on the consumer's mind", i.e., on the formation of the consideration sets themselves. The behavioral literature has emphasized that firms may also have ways to manipulate the evaluations of the alternatives *within* a fixed and given consideration set (see, e.g., Mullainathan et al., 2008; Bordalo et al., 2015, or Grubb, 2015 for a survey.). The two approaches are complements rather than substitutes: The value of manipulating any given consideration set simply enters the $V^j(\cdot)$ -function in our framework.

differences in the attention production across firms.

3 Equilibrium Analysis

We next characterize the market equilibrium and some of its key properties for a given number of firms n and a fixed consideration set size $R < n$. While we generally think of R as endogenously determined by the finite capacity model, it is a helpful intermediate step to first clarify the equilibrium implications of competitive attention for the case of a fixed R . The equilibrium implications if R responds to changing market conditions is the core subject of Section 4.

Regarding attention, payoff (9) shows that firms can differ in their abilities to attract attention ($c_j \neq c_k$) or to monetize the attracted attention ($V^j(R) \neq V^k(R)$). Here, we concentrate on the former and set $V^j(R) = V^k(R) \equiv V(R)$ for any two firms j, k . This does not require that all consumers have identical preferences. For example, we show in Appendix B.7 that such a symmetric situation results if consumers have heterogeneous preferences and horizontally differentiated firms strategically compete in prices. Appendix B.6 studies the case where firms differ in their abilities to monetize attention; the major equilibrium implications are similar.

3.1 Attention Equilibrium

Let $n > R$ be exogenously given, where $n, R \in \mathbb{R}_{++}$ for analytical convenience. Firms differ in their abilities to attract attention, and the attention costs are determined by a continuous and strictly increasing bijection $k : [0, n] \rightarrow [0, \bar{c}]$, such that firm j 's cost parameter is $c_j \equiv 1 + k(j)$. Thus, firms are ordered left-to-right, with $j = 0$ ($j = n$) as the most (least) efficient firm, and henceforth we identify firms directly by their cost types as these capture the relevant information about firm heterogeneity. Thus, the set of firms is $\mathcal{J} \equiv [1, 1 + \bar{c}]$, and $c \in \mathcal{J}$ is the firm with cost type c , and f_c denotes attention-seeking of this firm. Given the continuum structure of the type space, we replace summation by integration, such that $\Sigma = \int_{\mathcal{J}} f_c dc$.¹⁹

An *attention equilibrium* is a measurable function $f^* : \mathcal{J} \rightarrow \mathbb{R}_+$, where $\int f_c^* dc = \Sigma^*$ and the number f_c^* maximizes (9) at Σ^* for each firm $c \in \mathcal{J}$. The following proposition characterizes the unique attention equilibrium.

Proposition 2 *A unique attention equilibrium (f^*, Σ^*) exists, and the information load Σ^* , the perception chances π_c^* , profits Π_c^* , attention expenditures E_c^* and attention (clicks) C_c^* are*

$$\Sigma^* = \left(\frac{RV(R)I}{\eta} \right)^{\frac{1}{\eta}} K^{\frac{\eta-1}{\eta}} \quad K \equiv \int_1^{1+\bar{c}} s^{\frac{1}{1-\eta}} ds \quad (10)$$

¹⁹This formally implies that the perception chance per unit of attention, $\hat{\pi}_c = f_c/\Sigma$, is a (Lebesgue) density function of the firms types.

$$\hat{\pi}_c^* = \frac{f_c^*}{\Sigma^*} = \frac{c^{\frac{1}{1-\eta}}}{K}, \quad \pi_c^* = \hat{\pi}_c^* R, \quad (11)$$

$$\Pi_c^* = \hat{\pi}_c^* R V(R) I \frac{\eta-1}{\eta}, \quad E_c^* = \hat{\pi}_c^* \frac{R V(R)}{\eta} I, \quad C_c^* = \pi_c^* I. \quad (12)$$

In equilibrium, each firm understands the principle of relative salience, and optimally chooses its attention-seeking f_j while forming rational expectations about the total information load Σ^* .

3.2 Equilibrium Properties

We now discuss the key equilibrium implications of competitive attention, and highlight their relation to various equilibrium regularities.

Advertising The equilibrium identified by Proposition 2 is consistent with various empirical regularities. Expressions (11)-(12) show that f_c^*, π_c^* and Π_c^* are ordered according to cost types, where the firm with the lowest costs invests most, attracts most attention, and obtains the largest payoffs. Moreover, the top firms also are those to incur the highest attention expenditures, despite being most efficient in attracting attention.²⁰ As such, Proposition 2 is consistent with three central empirical regularities in advertising research: (i) There is a positive relation between profitability and attention expenditures (Bagwell, 2007), (ii) advertising expenditures are heavily skewed,²¹ and (iii) market and attention expenditure shares move in lockstep (Jones, 1990).

Power Law Property In our view, the most important prediction entailed in Proposition 2 is that perception chances, profits, attention (“clicks”), attention expenditures and sales are dispersed by (right-truncated) *power laws*. This follows as $\hat{\pi}_c$ in (11) adopts a power law distribution over the interval of firms \mathcal{J} , and profits Π_c^* , expenditures E_c^* , attention C_c , revenues and sales are scaled version of $\hat{\pi}_c^*$.²² If $\eta = 2$, then $\hat{\pi}(c)$ is inversely proportional in the cost c , in which case the resulting distribution is called “Zipf’s Law”. Figure 2 illustrates the dispersion of the aggregate attention data $\pi_j I$ (“clicks”), its characteristic linear shape in a log-log plot, as well as the distribution of the firm-side market shares (revenues or sales) and attention expenditures.

²⁰This is not obvious because more efficient firms could also achieve a higher market share than less efficient firms with lower attention expenditures.

²¹See Chauvin and Hirschey (1993) for such evidence across different industries; a nice exemplification is ad spending in the US beer market.

²²The revenue earned by firm c is $\pi_c V(R)I$, and total revenue is $R V(R)I$, which as $\hat{\pi}_c = \pi_c/R$ implies that $\hat{\pi}_c$ is the fraction of total revenue that goes to firm c . If $V(R)I = (p-c)d(p,R)I$, then $\pi_c d(p,R)I$ are the total sales of firm c , and it follows that $\hat{\pi}_c$ is the fraction of total sales that go to firm c .

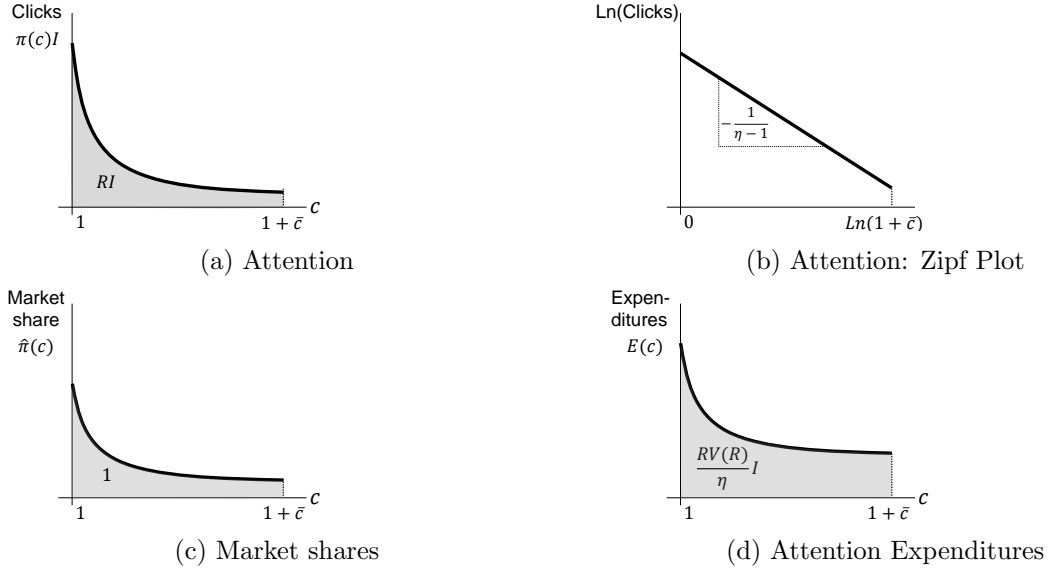


Figure 2: Equilibrium distribution of attention, market shares and expenditures

Power law dispersions are among the best documented empirical phenomena, and “fundamental in understanding the marketplace economy”.²³ Besides sales or revenues, quantities related to the attention of consumers also seem to follow a Power Law. For example, power laws were found in the online clicking or viewing-time distributions (Granka et al., 2004), in the click-through and conversion rates of keywords (Skiera et al., 2010), or internet site visits Adamic and Huberman (2000); Webster and Lin (2002). Such power laws are generally viewed as a part of the “laws of the web”.²⁴

To see why competitive attention implies such a distributional pattern, note first that a firm with a comparably very strong attention ability (low c) can afford to invest more into attention-seeking. The corresponding market share $\hat{\pi}(c)$ in turn burdens the firm with high equilibrium expenditures, despite a low c . Now, if c were to increase slightly, this would therefore have a severe impact on the expenditures needed to maintain the large market share, and a rational firm would then economize on costs by strongly cutting back its attention-seeking. This explains intuitively why the $\hat{\pi}$ -curve must be steep to its left. By contrast, the same hypothetical cost increase would have almost no effect if c is comparably high, meaning that such a firm invests only little in attention-seeking. Therefore, the $\hat{\pi}$ -curve must become flat to its right.

²³See Marketplaces Power Law. Power Laws have been documented, e.g., in the sales distribution of online books Chevalier and Goolsbee (2003), music purchases Newman (2005), mobile apps Zhong and Michahelles (2013), consumer electronics Hisano and Mizuno (2011) or movies Tan et al. (2017), as well as in consumption patterns in digital media marketplaces Webster and Ksiazek (2012), online news Hindman (2018) and mobile news Nelson (2019). Kohli and Sah (2006) report Power Laws in the market shares of a vast array of branded goods Zhang et al. (2009) and (Crawford et al., 2015) observe them in firm revenues, and Gaffeo et al. (2003) document them in case of firm sales across G7 countries. Also see Brynjolfsson et al. (2003, 2010a); Andriani and McKelvey (2011).

²⁴See Huberman (2003); Sinha and Pan (2006); Clegg et al. (2010) for overviews.

Now, the market share dispersion adopts a power law, rather than an arbitrary convex distribution, as a consequence of Steven’s Law and competitive forces. To see this, note that the characteristic property of a power law density $f(\cdot)$ is its *constant proportionality*, i.e., for any $\lambda > 0$, $\frac{f(\lambda z)}{f(z)}$ is independent of $z \in \mathbb{R}_{++}$. Steven’s Law of perception and $\hat{\pi}_c = f_c \Sigma$ imply that marginal costs MC_c must be power functions of the current market share, where $MC_c = \eta c \hat{\pi}_c^{\eta-1} \Sigma^{\eta-1}$. Thus, a firm with $c' = \lambda c$, $\lambda > 0$, could afford a market share equal to $\hat{\pi}_{c'} = \lambda^{\frac{1}{1-\eta}} \hat{\pi}_c$ at the same level of marginal costs. But competitive forces exactly equate marginal costs in equilibrium across all firms, meaning that market shares necessarily adopt the constant proportionality property $\frac{\hat{\pi}_{\lambda c}}{\hat{\pi}_c} = \lambda^{\frac{1}{1-\eta}}$. This implies that the market share distribution $\hat{\pi}(\cdot)$ must obey a power law.

In our static equilibrium model, Power Laws emerge as a consequence of competition and Steven’s Law of perception, which contrasts with previous contributions that derive Power laws as the limiting distribution of certain stochastic processes.²⁵ In that literature, it is a frequently posed question, whether Power Law patterns in the social sciences emerge from sheer chance or from rationality (Perc, 2014), where our analysis clearly points towards the latter. This is further strengthened in such that the attention equilibrium of our model can also arise as the steady-state of a dynamic attention process, as we show in Appendix B.5.

3.3 Attention Effects

Attention enters the equilibrium equations by means of the variables R and η , where we interpret the latter as capturing how sensitive consumer attention responds to the attention-seeking stimuli. In the following, we discuss the equilibrium effects of these two variables.

Sensitivity Effects The equilibrium equations in Proposition 2 show that the value of η matters for aggregate quantities as well as the distribution of attention, market shares and profits. To illustrate how, consider an exogenous reduction of η , meaning that consumer attention responds more sensitively to attention-seeking. First, this intensifies the competition for attention, which is manifested in higher aggregate attention expenditures and lower industry profits (see (12)). Second, the fact that in equilibrium $\hat{\pi}_c / \hat{\pi}_{c'} = (c' / c)^{\frac{1}{\eta-1}}$ shows that the relative differences in attention abilities are magnified if attention becomes more sensitive, such that there can be large differences in the perception chances despite small differences in the ability to attract attention.

As a consequence of the second effect, the market share, profit share, and the click dispersion rotate clockwise, which can easily be seen in Figure 2 (b), showing that a decrease in η implies a steeper Zipf plot. Thus, the more sensitively attention responds, the more market

²⁵See, e.g., Blank and Solomon, 2000; Reed, 2001; Newman, 2005; Gabaix, 2016.

inequality arises in equilibrium.²⁶ Several empirical papers have estimated the slope coefficient in a respective Zipf plot, e.g., of sales or attention data (see, e.g., Brynjolfsson et al., 2010b or Tan et al. (2017)). Our model thus provides a new explanation for what such coefficients could actually measure.

Regarding profitability, we note from (12) that the equilibrium return to attention is equal to η for any firm.²⁷ Thus, profitability diminishes if attention becomes more sensitive, which is reflected in lower aggregate profits. Nevertheless, this does not affect the profits for all firms equally. The profits of the bottom firms necessarily dwindle as a consequence of their loss in market shares. By contrast, the top firms are harmed least, as their gain in market share tends to compensate for the higher attention expenditures. The former effect may even dominate, leading to an upswing in profits for the top firms.²⁸

Consideration Set Size Effects We now consider the equilibrium effects of a reduction in consideration set size R . Intuitively, such a reduction has two opposing effects on the expected attention revenues. On the one hand, it is likely that firms manage to extract a higher revenue from consideration if the consideration sets are smaller, and thus the firm is compared to fewer competitors. On the other hand, a reduction of R means a lower chance of consideration, which tends to decrease the expected attention revenues.

Formally, these two opposing forces of R enter payoffs (9) via the expression $RV(R)$. In the following analysis, we assume the first effect to dominate, such that $RV(R)$ depends negatively on R . This equivalently means that $V(\cdot)$ is sufficiently elastic, i.e.,

$$-\frac{V'(R)R}{V(R)} > 1, \quad (13)$$

where $V'(R) < 0$. Standard models of strategic price competition verify (13), mainly due to a competition softening effect (see Section B.7). Intuitively, the consideration of additional competitors intensifies the price competition among the firms in a given consideration set, leading to lower prices and a loss of demand for each individual firm.

Given (13), a reduction in consideration set size ($dR < 0$) makes attracting attention more valuable. Accordingly, firms respond by boosting their attention-seeking activities, which implies

²⁶A clockwise rotation implies that the new distribution features more inequality as measured by any Lorenz-consistent inequality measure (Hefti and Teichgräber, 2020).

²⁷The equilibrium attention revenue is $\hat{\pi}_j RV(R)I$. Dividing this by E_j^* gives the return to attention.

²⁸To see that bottom firms lose, note that $\frac{\Pi_c}{\Pi_{c'}} = \frac{\hat{\pi}(c)}{\hat{\pi}(c')}$. Thus, by (11) relative profits increase if $d\eta < 0$ and $c < c'$. As aggregate profits diminish, this implies that the bottom firms must necessarily realize lower profits. To see that the top firms may gain, let $\eta < 2$, and note that for the lowest cost firm with $c = 1$, $\Pi_1 = RV(R)I^{\frac{2-\eta}{\eta}}$ as $\bar{c} \rightarrow \infty$. This shows that $\Pi_1(\eta') > \Pi_1(\eta)$ for $\eta' < \eta$.

that the total information load Σ as well as expenditures E_c^* increase, and most so for the firms with the largest market shares.

Moreover, the equilibrium forces are such that individual attention-seeking increases in proportion to Σ , implying that the equilibrium dispersion of market shares $\hat{\pi}_c$ is invariant to R . Nevertheless, the dispersion of attention in Figure 2 (a) is shifted downwards, because the area under the figure is fixed to RI (see Lemma 1). Thus, all firms lose the same constant fraction of clicks (attention), equal to dR/R , meaning that the top firms lose most clicks. That is, a lower consideration set size leads to a more equal distribution of clicks in terms of absolute differences.

The equilibrium profits evolve in a substantially different manner. All profits increase if $dR < 0$, despite higher expenditures. The reason is that, by competition, all firms must achieve the same equilibrium return to attention, equal to η . Consequently, expenditures and revenues must increase by the same proportion if $dR < 0$. Because revenues already exceeded expenditures in the previous equilibrium, the proportionality argument implies that revenues must increase more than expenditures in the new equilibrium, leading to higher profits.

Moreover, a lower consideration set size increases the absolute differences in profits between different firms, inducing what has been frequently called a *Matthew Effect* (see Section 4). To see this, note from (12) that relative profits are equal to relative market shares. Because the latter are invariant to R but profits increase as $dR < 0$, profits must increase more for firms with a larger market share - and thus the profit gap between any two different firms necessarily widens.

4 Superstar and Long Tail Effects

The digital transition, in particular the rise of the Internet, brought major changes to markets, among which are an accretion in the number of people having access to the various markets and the entry of new products, mostly at the tail of the distribution.²⁹ Regarding the latter, there are two major views about who may gain or lose from an increasing product variety caused by a digitized competition.

According to the *Superstar effect*, first introduced by Rosen (1981), quality differences in the supplied services jointly with scale effects in production and consumption imply that a small number of price-taking firms – a few “superstars” – manage to earn an ever larger share of

²⁹For example, the number of book titles bought each year, e.g., in Australia has risen from 275'000 to 450'000 from 2004 to 2007 (see “A world of hits” (The Economist, Nov 26, 2009), or Brynjolfsson et al., 2003; Bar-Isaac et al., 2012.) The number of scripted TV shows in the US has more than doubled from 2010 to 2016 (see “The paradox of Choice”, The Economist, Feb 11, 2017). The number of different movies rented on Netflix has more than quadrupled from 2001 to 2005 (Tan et al., 2017), and similar for video sales (Elberse and Oberholzer-Gee, 2008), where by far most product entry has occurred at the tail of the distribution as measured by sales.

sales or profits as new firms enter at the bottom of the distribution.³⁰ By contrast, the *Long Tail Effect*, coined by Anderson (2004), refers to a supposed increasing profitability of the tail products as a consequence of technology-based changes associated with an increasingly digitized economy and the structure of demand. In particular, the high costs of storage and distribution in the pre-digital era made it imperative for companies to fill their shelves or magazine spaces predominantly with mainstream products, despite a “taste-of-variety” in consumer preferences. In addition, consumers may fail to discover various products of interest due to high search costs, e.g., associated with physically visiting several outlets.³¹ Accordingly, a growing measure of tail products in a digitized economy should “drive demand from the head to the tail” (Anderson, 2004), leading to less concentrated markets and “ultimate fragmentation”.

Neither the Superstar nor the Long tail effect includes attention in their analysis, which practitioners have seen as a key component missing in the debate about the Long Tail.³² Our model predicts that, as a consequence of competitive attention, sales and other quantities are dispersed according to power laws, which is the type of market where the Long Tail should be most effective (Anderson, 2004). We use our equilibrium model to study whether competitive attention rather promotes the occurrence of a Superstar or a Long Tail effect (or both) as the market size or the number of tail products increases.

4.1 Equilibrium Forces

Analyzing the effects of a growing tail or consumer entry amounts to studying the comparative-statics of our equilibrium model, where consideration sets are determined by the Finite Capacity Model of Section 2, and firms compete for attention as in Section 3. The equilibrium forces are illustrated in Figure 3. For any given consideration set size $R < n$, individual firms ratio-

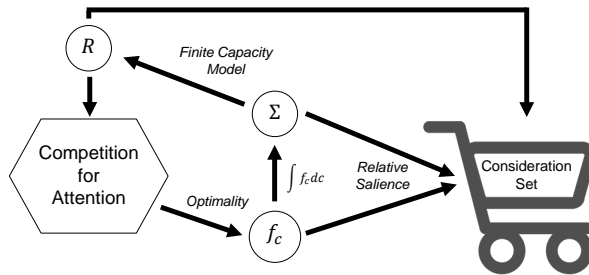


Figure 3: Equilibrium Forces

nally choose their attention-seeking f_j , taking into account that consideration sets are formed

³⁰Such effects may be reinforced if consumers have tastes for conformity (Frank and Cook, 1995).

³¹See, e.g., Brynjolfsson et al., 2010b for an overview of the supply and demand side arguments favoring a Long Tail effect.

³²See, e.g., “The battle for consumers’ attention”, The Economist, Feb 09, 2017.

according to relative salience. The corresponding competition for attention produces an information load Σ , which affects the consideration formation process according to the finite capacity model. Thus, two countervailing forces arise. If the “shouting” for attention becomes louder, choice overload is aggravated, and consumers respond by “hearing less”. By contrast, smaller consideration sets allow firms to better monetize the attention they manage to attract, which increases the shouting for attention. These two forces are illustrated in Figure 4 (a) by the functions $R(\Sigma)$ and $\Sigma(R)$.

An equilibrium is a triple (f^*, Σ^*, R^*) , where $R^* = R(\Sigma^*)$ is determined by (2), and (f^*, Σ^*) constitutes the attention equilibrium from Section 3.1 given that $R = R^* < n$. By Proposition 1, $R(\Sigma)$ is strictly decreasing if $n > R$, and from Section 3.3 we know that $\Sigma(R)$ is also strictly decreasing for $n > R$. An equilibrium with choice overload ($R^* < n$) then is determined by an intersection of the two functions.

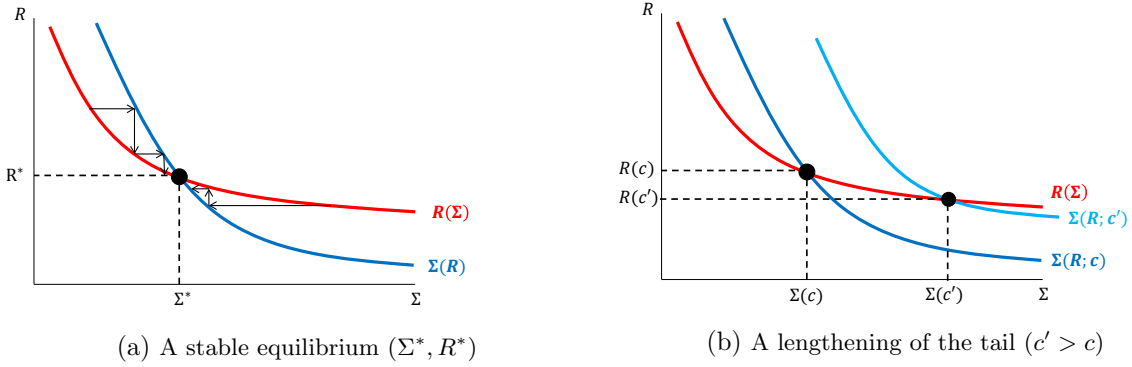


Figure 4: Equilibrium consideration set size R^* and information load Σ^* with choice overload

Changes in the number of choice alternatives or the measure of consumers affect the competition for attention and formation of consideration sets because it alters individual attention-seeking and the information load produced by the market, which both affect the formation of the consideration set. In the following comparative-static analysis, we concentrate on the case where a unique and stable equilibrium (f^*, Σ^*, R^*) with choice overload exists, which requires that the $R(\Sigma)$ -locus intersect the $\Sigma(R)$ -locus from below, as illustrated in Figure 4 (a).³³

4.2 Definitions

To study whether a Superstar or a Long Tail effect arises, we first need to clarify these concepts in the context of our framework. We base our respective notions on the definitions invoked

³³ If $R(\Sigma)$ intersects $\Sigma(R)$ -locus from below, this assures asymptotic stability for several standard dynamics $R(t), \Sigma(t)$ that can be defined via the equilibrium system (Hefti, 2016), where a stable equilibrium requires that $|\frac{\mathcal{V}'(R)R}{\mathcal{V}(R)}| < \eta |\frac{R(\Sigma)}{R'(\Sigma)\Sigma}|$, $\mathcal{V}(R) \equiv RV(R)$. Further, it is principally possible to derive sufficient conditions for such an equilibrium to arise, which however have no further implications for our equilibrium predictions, in particular because the changes we study aggravate rather than diminish the choice overload problem.

frequently by the empirical literature. Several empirical papers define the Long Tail and Superstar effect, respectively, in relative terms, such as the top $x\%$ or bottom $y\%$ of firms (sales or revenues).³⁴ For example, Tan et al. (2017) refer to the Superstar effect as a relative increase in profits of the top 1%, or 10% of movies on Netflix, and to the Long Tail effect as an increase in the relative sales of the bottom 1% and 10% of movies; also see Cho and Roy (2004); Elberse (2008); Elberse and Oberholzer-Gee (2008); Brynjolfsson et al. (2011); Zhou and Duan (2012); Zhong and Michahelles (2013). We follow these papers, and define the Superstar and Long Tail effects in relative terms.³⁵

In the following definitions, $F(\cdot)$ denotes a measure of interest, such as profits, revenues, sales or attention (“clicks”). Consistent with the equilibrium characterization in Proposition 2, we identify different firms with their cost type c .

Definition 1 Let $\mathcal{J} \equiv [1, 1 + \bar{c}]$ be the set of firms, and $F(c, \chi)$ be a C^1 -function, where $c \in \mathcal{J}$ and $\chi \in \mathbb{R}$ is a parameter. Fix $x \in (0, 1/2)$ and denote by $\mathcal{J}_T \equiv [1, 1 + x\bar{c}]$ and $\mathcal{J}_B \equiv [1 + (1 - x)\bar{c}, 1 + \bar{c}]$ the top and bottom $x\%$ of firms, respectively. Then

$$Z_T(\chi) \equiv \frac{\int_{\mathcal{J}_T} F(c, \chi) d_c}{\int_{\mathcal{J}} F(c, \chi) d_c}, \quad Z_B(\chi) \equiv \frac{\int_{\mathcal{J}_B} F(c, \chi) d_c}{\int_{\mathcal{J}} F(c, \chi) d_c}, \quad (14)$$

denote the $F(\cdot)$ -share of the top and bottom $x\%$ of firms. A **Superstar effect** in χ occurs if $Z'_T(\chi) > 0$, and a **Long Tail effect** in χ occurs if $Z'_B(\chi) > 0$.

For example, if $F(\cdot) \equiv \Pi(\cdot)$ corresponds to equilibrium profits, and $\chi \equiv I$ is the measure of consumers, then a Superstar effect (Long Tail effect) occurs if the share of profits earned by the top $x\%$ of firms (bottom $x\%$ of firms) increases in I . The firms in $[1 + x\bar{c}, 1 + (1 - x)\bar{c}]$ constitute the “middle ground” and, $Z_M(\chi) \equiv 1 - Z_T(\chi) - Z_B(\chi)$ is the share of profits earned by the middling firms. Note that Superstar and Long Tail effect need not be mutually exclusive, and should they both arise, the middling firms always lose.³⁶

While we follow most of the empirical literature in our relative definitions of the Superstar and Long Tail effects, we also want to consider absolute changes, e.g., of profits or clicks. Specifically, we shall study whether the absolute gap between the top and bottom firms increases, to which

³⁴One challenge is that different empirical metrics or definitions of these effects have been implemented, making a conclusive comparison of the results difficult (see, e.g., Brynjolfsson et al., 2010b; Tan et al., 2017).

³⁵In the taxonomy of Brynjolfsson et al. (2010b) we thus consider a “Relative Long Tail” measure. Also see Tucker and Zhang (2007); Hinz et al. (2011); Skiera et al. (2010); Peltier and Moreau (2012); Gu et al. (2013) for related empirical contributions.

³⁶One might note that the $x\%$ lowest cost types, i.e., $\mathcal{J}_T = [1, 1 + x\bar{c}]$ does not necessarily coincide with $x\%$ of the firms. For example, if $n = 10$ and $c_j = 1 + k(j) = 1 + j^2$, such that $\bar{c} = k(n) = 100$ then the set of the 10% lowest cost types is $\mathcal{J}_T = [1, 11]$, which amounts to $\frac{\sqrt{10}}{10} \% \approx 31,6\%$ of all firms. Clearly, given the bijection $k(j)$ between j and c_j , any given percentage $x\%$ of cost types can be converted into a corresponding percentage of firms, and vice-versa. This is irrelevant for our equilibrium analysis because, as we shall see, our results do not rely on any particular value of x .

we refer to as a Matthew effect.³⁷

Definition 2 Let $x \in (0, 1/2)$ and $F(c, \chi)$ be as in Definition 1. Define

$$M_x(\chi) \equiv \int_{\mathcal{J}_T} F(c, \chi) dc - \int_{\mathcal{J}_B} F(c, \chi) dc. \quad (15)$$

as the absolute gap between the top and bottom $x\%$ of firms. Then the parameter χ induces a (inverse) **Matthew effect** in $F(\cdot)$ if $M'_x(\chi) > (<) 0$.

A (inverse) Matthew effect occurs, if the gap in $F(c, \chi)$ between the top $\chi\%$ and the bottom $\chi\%$ of firms widens (narrows). The Matthew effect, in general, does neither imply nor preclude the occurrence of a Superstar or a Long Tail effect.

4.3 Equilibrium Effects of a Longer Tail

Rosen (1981) and Anderson (2004) consider a lengthening of the tail, i.e., the entry of new products at the bottom of the distribution, as an elementary part of their expositions.³⁸ In our model, “entry at the tail” means that the set of cost types increases from $\mathcal{J}(\bar{c}) = [1, 1 + \bar{c}]$ to $\mathcal{J}(\bar{c}') = [1, 1 + \bar{c}']$, $\bar{c}' > \bar{c}$.

4.3.1 Superstar or Long Tail Effect?

To analyze whether a Superstar or Long Tail effect arises once $\bar{c} \rightarrow \bar{c}'$, we first consider equilibrium profits, i.e., $F(c) = \Pi(c)$, and study the tractable case where $\eta = 2$, such that $\Pi(c)$ (12) obey Zipf’s Law. For a given $x \in (0, 1/2)$, using (12) in (14) yields

$$Z_T(\bar{c}) = \frac{\int_{\mathcal{J}_T} \Pi_c dc}{\int_{\mathcal{J}} \Pi_c dc} = \int_1^{1+x\bar{c}} \hat{\pi}_c dc = \int_1^{1+x\bar{c}} \frac{1}{cK(\bar{c})} dc = \frac{\text{Ln}(1+x\bar{c})}{\text{Ln}(1+\bar{c})}. \quad (16)$$

Standard algebra gives $Z'_T(\bar{c}) > 0$, $\forall x \in (0, 1)$, $\bar{c} > 0$, showing that a growing tail causes a Superstar effect in profits. The intuition is as follows. If the tail grows longer, the joint profits of the top $x\%$ of firms must increase more (or diminish less) than the total market profit for a Superstar effect in profits to emerge. By (12) and (16), this boils down to how the joint market shares of the top $x\%$ firms evolve, i.e., the dispersion of market shares and profits follow

³⁷The Matthew effect, a term coined by sociology, refers to the observation that the “the rich get richer and the poor get poorer”. Originally, the term was used to describe a pattern in scientific citations, where already famous authors receive far more citations, thereby becoming even more famous, than their less known counterparts (Merton, 1968). The notion of a Matthew effect has been used in the economics literature predominantly in the context of a rising income inequality, but similar observations have been subsumed under a large class of names, such as first mover advantage, path-dependent increasing returns, increasing dominance, the rich-get-richer or reputational effects; see DiPrete and Eirich (2006); Perc (2014) for surveys.

³⁸Such entry could occur, e.g., due to the decreasing fixed costs typically associated with a digital economy, allowing higher cost firms to enter the market.

essentially the same pattern. There are two opposing effects.³⁹

First, there is an *attention dilution effect*. The lengthening of the tail intensifies the competition for attention and increases attention expenditures. This amplifies consumer overload as captured by smaller consideration sets ($R'(\bar{c}) < 0$) (see Figure 4), which results in a downward shift of the market share function in Figure 2 (c). By this effect, the top firms lose most, reflecting that their large equilibrium expenditures are most exposed to an intensified attention competition. Second, there is a *composition effect*. An expansion of firms, say by $q\%$, must increase the measure of firms belonging to the top $x\%$. As the firms around the top by definition are the ones with the largest market shares, the aggregate market share of the top $x\%$ firms tends to increase as more firms belong to the top $x\%$.

With respect to these two effects, the power law shape of $\hat{\pi}(\cdot)$ implies that the dilution effect must be dominated by the composition effect, which means that the reduction of market shares due to attention dilution must be compensated by the aggregate market shares generated through the additional firms entering the top $x\%$ segment. Therefore, a Superstar effect must arise as a consequence of the power law dispersion of the market shares.

The same forces are at play in case of the Long Tail effect. In particular, the entrant firms become part of the tail, and as such increase the market share captured by tail firms. However, all previous tail firms lose some market share due to the entry, and some firms cease to be part of the tail firms as defined by the Long Tail effect (those at the left margin of $[1 + (1 - x)\bar{c}, 1 + \bar{c}]$). It is easy to see that $Z_B(\bar{c}) = 1 - \frac{\text{Ln}(1+(1-x)\bar{c})}{\text{Ln}(1+\bar{c})}$, and $Z'_B(\bar{c}) < 0$ follows from the same algebra assuring that $Z'_T(\bar{c}) > 0$. Thus, the equilibrium forces leading to a Superstar effect, in particular the power law shape of $\hat{\pi}(\cdot)$ exactly prevent a Long Tail effect to occur from a growing tail. As a consequence, we see that the tail must grow thinner as it grows longer.

The fact that the head of the distribution gains market shares while the bottom loses market share makes the analysis of the middle ground non-obvious, as the middle ground could end up as winner or loser. For any $x \in (0, 1/2)$ the share of profits of the firms in the middle ground is $Z_M(\bar{c}) = \frac{\text{Ln}(1+(1-x)\bar{c}) - \text{Ln}(1+x\bar{c})}{\text{Ln}(1+\bar{c})}$, and differentiation reveals that $Z'_m(\bar{c}) < 0$. Again, this result is a consequence of the forces leading to a power law distribution of the market shares, which imply that the loss of some large firms to the head jointly with the generally reduced market shares due to diluted attention dominates the gains of the middle ground from some firms previously in the tail.

³⁹See Brynjolfsson et al. (2010b) for a simple example where the sales share of the top $x\%$ of products increases mechanically if a large number of tail products with the same low levels of sales enters the market and the sales of the incumbent products remain fixed. In our model, sales, profits and market shares of the entrants and the incumbents are endogenously determined by the competition for attention, making the relation between profit shares and tail entry far from trivial.

In sum, the above analysis shows that for $\eta = 2$ and any $x\%$ -definition of the head, the tail and the middle ground lose profit shares to the head. The same pattern emerges in case of attention (clicks), revenues or sales shares, as these shares are determined by the market share function $\hat{\pi}(\cdot)$ alone.⁴⁰ These patterns generalize for any $\eta > 1$, as the next proposition summarizes.

Proposition 3 (Superstar and Long Tail effects) *Let $x \in (0, 1/2)$ and $\eta > 1$. A growing tail ($d\bar{c} > 0$) increases the share of profits, attention, expenditures, revenues or sales of the top $x\%$ of firms, while the shares of the bottom $x\%$ of firms and of the middling firms deteriorates. Thus, a growing tail yields a Superstar effect in the dispersion of these quantities, but not a Long Tail effect.*

Figure 5 (a) illustrates Proposition 3 for the case of the Zipf distribution ($\eta = 2$) and $x = 0.2$, showing how the profit share of the top 10% ($Z_T(\bar{c})$), the bottom 10% ($Z_B(\bar{c})$) and the middle 80% of firms evolve as the tail grows. Most remarkable is that the profit share of the 10% top firms surpasses the one of the 80% middle firms once the tail is long enough. Regarding

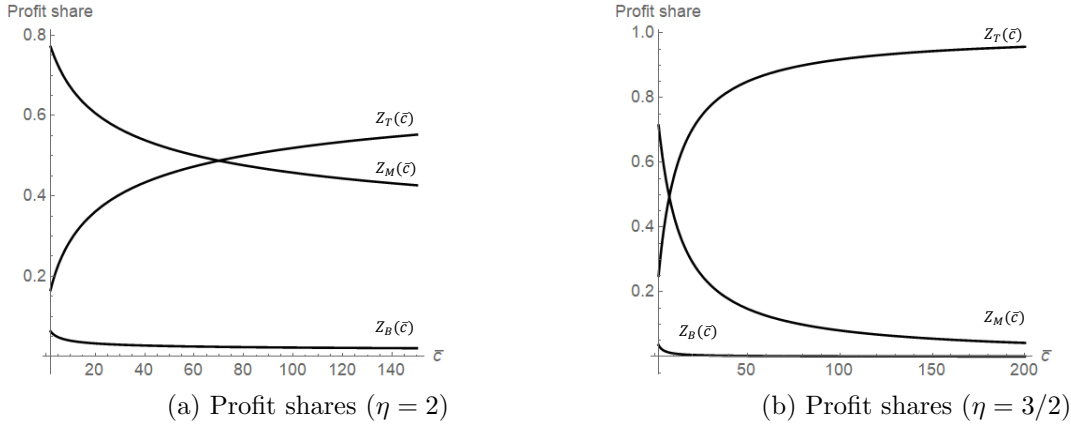


Figure 5: The effects of a growing tail on profits

magnitudes, using (12) in (16) shows that the consumer sensitivity to attention-seeking, η , affects the growth of $Z_T(\bar{c})$. Intuitively, we expect a growing tail to show more pronounced effects on $Z_T(\bar{c})$ the more sensitive consumer attention is (i.e., the lower η is). The following result confirms that the attention sensitivity indeed plays a key role for how the market shares evolve if the tail grows arbitrarily long.

Corollary 2 *Let $Z_T(\bar{c})$, $Z_B(\bar{c})$, $Z_M(\bar{c})$ denote the the share of profits, attention or revenues held by the top $x\%$, bottom $x\%$ and middling $(1-2x)\%$ of the firms, respectively. If $\eta > 2$, then $\lim_{\bar{c} \rightarrow \infty} Z_T(\bar{c}) = x^{\frac{\eta-2}{\eta-1}} \in (0, 1)$, $\lim_{\bar{c} \rightarrow \infty} Z_B(\bar{c}) = 1 - (1-x)^{\frac{\eta-2}{\eta-1}} \in (0, 1)$ and $\lim_{\bar{c} \rightarrow \infty} Z_M(\bar{c}) = (1-x)^{\frac{\eta-2}{\eta-1}} - x^{\frac{\eta-2}{\eta-1}} \in (0, 1)$. If instead $\eta \in (1, 2)$, then $\lim_{\bar{c} \rightarrow \infty} Z_T(\bar{c}) = 1$ and $\lim_{\bar{c} \rightarrow \infty} Z_B(\bar{c}) = \lim_{\bar{c} \rightarrow \infty} Z_M(\bar{c}) = 0$.*

⁴⁰For the case where $V(R) = (p(R) - c)d(p, R)$, sales correspond to $\pi(c)d(p, R)$.

If consumer attention is not too sensitive ($\eta > 2$), the profit shares Z_T , Z_M , Z_B all stabilize asymptotically at some positive number. In fact, in the extreme case where $\eta \rightarrow \infty$, such that consumer attention does not react to the stimuli at all, the profit shares of the top and bottom $x\%$ remain constant at x . By contrast, the Superstar effect literally turns into a “winner-takes-all” effect if consumer attention is sufficiently sensitive ($\eta \leq 2$), meaning that the profit shares of the middle and the tail firms are completely eroded as the tail grows longer; see Figure 5(b).

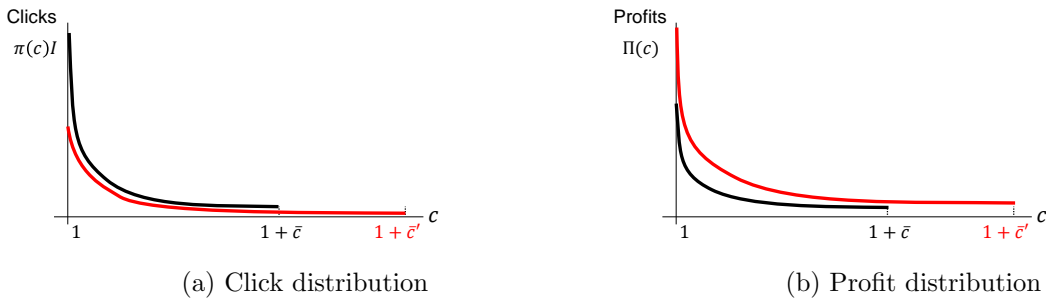
4.3.2 Absolute Effects

Regarding the absolute effects of a growing tail, our perhaps most remarkable finding, derived below, is that the equilibrium profits of the incumbents may increase due to a growing tail. Moreover, this can occur despite that some attention is diluted due to entry, and despite that the standard economic forces at play would suggest a reduction in profits absent limited attention.

To illustrate, consider the following simple example, where $\eta = 2$, $R(\Sigma) = \alpha \Sigma^{-\beta}$ by (4), and $V(R) = R^{-2}$.⁴¹ It is easily verified that $\beta < 2$ in a stable equilibrium, which we assume in the following. Further, (10) gives $\Sigma^2 = I \ln(1 + \bar{c}) / (2R)$, from which

$$R(\bar{c}) = \left(\frac{I \ln(1 + \bar{c})}{2\alpha} \right)^{\frac{-\beta}{2-\beta}}$$

follows. Then, equilibrium profits are $\Pi_c(\bar{c}) = Q(c) \ln(1 + \bar{c})^{\frac{2(\beta-1)}{2-\beta}}$, while attention is $C_c(\bar{c}) = W(c) \ln(1 + \bar{c})^{\frac{-2}{2-\beta}}$, where $Q(c) > 0$ and $W(c)$ are constants. It follows that $C'_c(\bar{c}) < 0$ unambiguously while $\Pi'_c(\bar{c}) > 0$ iff $\beta > 1$. This shows two things. First, either all or no incumbent benefits from a growing tail, while all incumbents lose on attention. Second, for the incumbents to benefit, β must be sufficiently large, meaning that the overload effect due to the increasing information load must be strong enough. Such a situation is depicted in Figure 6.



The figure shows schematically how the click (a) and profit (b) distribution change as the tail grows from \bar{c} (black) to \bar{c}' (red) for a situation with $\beta > 1$.

Figure 6: Effects of a growing tail on attention and profits.

⁴¹This particular revenue function is obtained from a standard random utility model (see Appendix B.7).

This result, while perhaps surprising relative to conventional competition theory, is highly intuitive from the perspective of the underlying attention effects. By (12), equilibrium profits of an incumbent firm is $\Pi_c = \frac{1}{2}\hat{\pi}_c RV(R)I$, showing that firm entry at the tail imposes two conflicting effects on an incumbent's profits. First, the entry of new firms dilutes some attention away from the incumbents, and decreases attention in general due to an increased choice overload ($dR < 0$). This effect, *ceteris paribus*, reduces incumbents' profits. Second, the increasing choice overload itself softens competition and allows firms to better monetize the attention they attract ($RV(R)$ increases). This effect, *ceteris paribus*, increases profits. Moreover, this competition softening effect is a central reason for why firm entry reduces profits of the incumbents absent limited attention.⁴²

The question of interest here is whether equilibrium forces dilute attention so much as to compete away the additional revenues from consumer inattention. In our simple example, this happens if and only if the choice overload effect is sufficiently weak, captured by a small β . More generally, we would also expect the details of the economic competition to matter, as captured by how sensitive $V(R)$ responds to changes in R . The following proposition makes this intuition precise, and shows that the patterns of our simple example generalize.

Proposition 4 *Let $\eta > 1$ and consider a stable equilibrium, where profits $\Pi_c(\bar{c})$ and attention $C_c(\bar{c}) = \pi_c I$ are determined by Proposition 2. Let $\mathcal{J}(\bar{c}) = [1, 1 + \bar{c}]$ denote the set of incumbent firms, and $c, c' \in \mathcal{J}(\bar{c})$ with $c < c'$. A growing tail*

- (i) *decreases each incumbent's attention, $C'_c(\bar{c}) < 0, \forall c \in \mathcal{J}(\bar{c})$, and decreases the attention gap $C_c - C_{c'}$*
- (ii) *increases industry-level profits, and increases profits of an incumbent firm iff the profit of each incumbent increases,*
- (iii) *increases the profit gap $\Pi_c - \Pi_{c'}$ between any two incumbent firms iff the profit of any incumbent increases,*
- (iv) *increases the profit for an incumbent firm iff $\left| \frac{\mathcal{V}'(R)R}{\mathcal{V}(R)} \right| > \left| \left(\frac{R'(\Sigma)\Sigma}{R(\Sigma)} \right)^{-1} \right|$, where $\mathcal{V}(R) \equiv RV(R)$, and*
- (v) *yields a Matthew effect in profits.*

Condition (iv) clarifies our previous intuition, and shows that if either competition softens enough due to inattention ($|\frac{\mathcal{V}'(R)R}{\mathcal{V}(R)}|$ is large) or the consideration set size responds highly sensitive to information overload ($|\left(\frac{R'(\Sigma)\Sigma}{R(\Sigma)} \right)^{-1}|$ is small), then the gains from choice overload earned

⁴²Without limited attention, both $V(n)$ and $nV(n)$ are strictly decreasing in n by (13), yielding the standard effect that firm entry reduces both incumbents' and industry profits.

from considering consumers exceed the losses due to attention dilution. Looking at the details, competition softening due to inattention likely is substantial if products are strong substitutes, i.e., we are in an economic environment where competition is intense and margins low by standard economic forces⁴³ Likewise, example (4) shows that $\left| \left(\frac{R'(\Sigma)\Sigma}{R(\Sigma)} \right)^{-1} \right|$ is small if β is large, which is more likely to hold, e.g., with complex products. Overall, this suggests that incumbents are likely to benefit from a growing tail if products effectively are strong substitutes in terms of underlying tastes, but their comparison involves a lot of effort, e.g., due to true product complexity or because of firm-side obfuscation (Hefti et al., 2020). Finally, (v) shows that a Matthew effect in profits emerges, meaning that the top firms either earn strictly more than the bottom firms, or at least lose less. In any case, this shows that market inequality increases in absolute terms.

Taking stock, the above results show a clear deviation from conventional models of economic competition, where firm entry at the tail typically is harmful for incumbents and reduces industry profits due to an intensified competition. Moreover, we find a growing tail to yield both a Superstar and a Matthew effect in profits, suggesting an increasing concentration on a few rather than a leveling of the playing field as it occurs in conventional models, or as suggested by Anderson’s Long Tail.

Sales Several empirical studies are concerned about the distribution of sales. Regarding the corresponding market shares, we have already seen that sales show a Superstar effect if the tail grows. By contrast, the gap between different incumbent firms narrows upon firm entry, as we show next. Let $d(p, R)$ denote the average per-consumer demand if firm c belongs to a consideration set, such that the total sales are $S_c = \pi_c d(p, R)I$. In the following, we assume that $d(p, R) = w(p)/R$, where $w'(p) \leq 0$ and $p'(R) < 0$. As we show in Appendix B.7 this functional form is implied by standard competition models.⁴⁴

Corollary 3 (Sales) *Let $S_c = \pi_c \frac{w(p)}{R}I$ denote the sales of incumbent $c \in \mathcal{J}$. Then, $S'_c(\bar{c}) < 0$, and $\frac{\partial}{\partial \bar{c}}(S_c - S_{c'}) < 0$ for any two incumbents $c < c'$.*

Intuitively, entry causes the sales of each incumbent firm to diminish, because the dilution of attention dominates the potential gain in demand due to smaller consideration sets if $d(p, R) = w(p)/R$. Moreover, market forces imply that relative sales of any two incumbents must remain constant. As sales decrease, this then implies that they must decrease more for incumbents

⁴³For example, it is a simple exercise to show that with CES-demand $\left| \frac{v'(R)R}{v(R)} \right|$ is strictly increasing in the substitution parameter.

⁴⁴See Appendix B.7. In discrete choice models, $d(p, R) = 1/R$, such that $w(p) = 1$. In Euclidean models, $d(p, R) = m/(pR)$, where $m > 0$ is disposable wealth, such that $w(p) = m/p$.

with larger market shares.

4.4 Market Size Effects

We next study the implications of a growing market size, i.e., an increase in the measure of consumers $dI > 0$, for a given set of firms \mathcal{J} . Proposition 2 immediately shows that the equilibrium dispersion of market shares $\hat{\pi}_c$ is invariant to I , meaning that I neither a Superstar nor a Long Tail effect occurs in profits, attention, sales or revenues. Moreover, the fact that an increase in the audience size I alone does not affect the shape of the attention distribution is consistent with evidence from online news consumption (Nelson, 2019).

The main reason for these results goes back to competitive forces and the structure of the attention costs. Intuitively, all firms strive for a larger market share if the amount of total disposable consumer income increases. This, however, increases attention expenditures, and more so for firms that already hold a large market share. In equilibrium, marginal costs are equated, which by the power function form of these costs implies that the additional attention investments must increase by the same proportion for all firms, leaving market shares unaltered.

Although market shares remain constant, $dI > 0$ induces absolute effects because all firms ramp up their attention-seeking activities.

Proposition 5 *Let $\eta > 1$ and consider a stable equilibrium, where profits $\Pi_c(\bar{c})$ and attention $C_c(\bar{c}) = \pi_c I$ are determined by Proposition 2. Let $\mathcal{J}(\bar{c}) = [1, 1 + \bar{c}]$ denote the set of firms, and $c, c' \in \mathcal{J}(\bar{c})$ with $c < c'$. A growing market size*

- (i) *decreases each firm's perception chance per consumer, π_c , and more so for larger firms,*
- (ii) *increases each firm's attention $C_c = \pi_c I$ iff total market attention RI increases,*
- (iii) *increases total market attention iff $1 + \left| \frac{\nu'(R)R}{\nu(R)} \right| < \eta \left| \frac{R'(\Sigma)\Sigma}{R(\Sigma)} \right|^{-1}$*
- (iv) *yields a Matthew effect in attention if market attention increases, and an inverse Matthew effect otherwise*
- (v) *increases the profit for any firm, and yields a Matthew effect in profits.*

A growing market increases the value of any given market share, which intensifies the competition for attention. By our analysis in Section 3, the concomitant choice overload effect implies that the top firms must lose most attention per consumer (i) among all firms. However, such losses may be more than compensated by the additional consumers entering the market (ii), where the latter effect dominates by (iii) if the attention competition is not too intense (low η), the information overload effect is not too pronounced, or product competition is not too intense

(low $\left| \frac{V'(R)R}{V(R)} \right|$). In our example with $V(R) = R^{-2}$ and $R(\Sigma) = \alpha \Sigma^{-\beta}$, this holds iff $\beta < \eta/2$. Finally, (iv) and (v) are a consequence of the fact that market shares $\hat{\pi}_c$ remain constant, but all profits increase while aggregate attention may or may not increase.

4.5 Discussion

We analyzed the effects of a growing tail or a growing market in a ceteris-paribus manner, while both effects are likely to occur simultaneously in reality. For example, a growing measure of consumers increases profits, which allows less efficient firms to cover a fixed entry cost, leading to a growing tail. The equilibrium predictions then are a combination of both effects, where our model predicts a Superstar effect but not a Long Tail effect in profits to arise if I increases and the tail lengthens at the same time.

Our analysis shows that if attention is competitively allocated through the actions taken by rational firms, a rising inequality in revenues, sales or profits is a likely outcome as markets grow larger. In particular, the tail is predicted to grow thinner as it grows longer, with the head benefiting most. Such a pattern may be hard to rationalize, e.g., with classical theories of product differentiation (e.g. Lancaster, 1990), where an increasing product differentiation generally reduces market concentration by fostering product competition. Nevertheless, our predicted patterns fit the data in many cases.

Regarding the head, the attention predictions resonate well with a substantial empirical literature observing a Superstar effect in digital sales or revenues data. For example, Tan et al. (2017) provide a careful econometric analysis based on movie rental data, strongly supporting that the substantial increase in the number of available movies caused a Superstar effect, measured by relative sales, but not a Long Tail effect. A similar observation is made by Elberse and Oberholzer-Gee (2008) who find that sales are increasingly concentrated on the top titles, speaking in favor of a Superstar effect. Similar superstar effects, following an increase in product variety, have been documented by Skiera et al. (2010), Zhou and Duan (2012) or Ordanini and Nunes (2016). Further, Elberse and Oberholzer-Gee (2008) report that the entry of new movies reduces the sales of incumbent movies, and more so for movies with larger sales. This is consistent with the general dilution effect of attention and sales as well as with the corresponding inequality effects identified by Proposition 4 and Corollary 3. A similar finding is reported by Zhou and Duan (2012) in case of software downloads.

Regarding the tail, the empirical evidence is not fully conclusive, in parts because frequently incongruent empirical metrics have been used (Brynjolfsson et al., 2010b). However, the empirical findings surveyed by Napoli (2016) or Tan et al. (2017) strongly speak against a success of the tail following a lengthening of the tail. Likewise, Elberse and Oberholzer-Gee (2008)

conclude that in the movie industry events are best described as a tail which has grown *longer but thinner*, which is exactly what the competitive allocation of limited attention predicts. A similar pattern also has been observed in the music industry, and Webster and Ksiazek (2012) find that fragmentation, in the sense of the Long Tail, has not occurred on digital media marketplaces, while popular brands have fostered their popularity.⁴⁵

Taking stock, the evidence from different industries strongly suggests a Superstar effect to emerge jointly with a substantially longer but thinner tail. The fact that an increasing product diversity leads to an increasing sales concentration is hard to reconcile with the contrary effects predicted by classical models of product differentiation, including models of consumer search, but consistent with the equilibrium predictions offered by competitive attention. More generally, the laws of attention and competition jointly can explain why a substantial increase in the size of the audience and the diversity of the available choice options – two central premises of the Long Tail effect that have been satisfied throughout the last two decades – have not qualitatively altered the distributional appearance of consumption or revenue patterns (Hindman, 2018; Nelson, 2019).

5 Related Literature

Rational Inattention and Consumer Search In economics, the predominant approach towards attention is given by the rational inattention framework, starting with Sims (2003). The key presumption of this approach is that the allocation of attention is under the mental control of a rational decision-maker, and as such the outcome of a rational cost-benefit analysis.⁴⁶ Besides formalistic differences, this is a conceptual similarity to search models, in that the information acquisition problem is solved by the consumer via rational cost-benefit analysis.

In psychology, rational inattention or optimal search would be viewed as a *goal-driven* (or top-down) attention process. By contrast, our paper explores the implications of *stimulus-driven* (or bottom-up) attentional control. Nevertheless, the equilibrium allocation of attention still is a rational outcome, in that it reflects the optimal decisions of the competing firms how much to invest into attention. Such attraction effects have no existing counterpart in the rational inattention literature.⁴⁷

⁴⁵See, e.g., Elberse, 2008; Dellarocas and Narayan, 2007; Tucker and Zhang, 2007; Gu et al., 2013; Zhong and Michahelles, 2013 or “The paradox of choice” for evidence favoring a Superstar rather than a Long Tail effect.

⁴⁶One scientific merit is that this approach yields a rational benchmark that can be used to identify systematic anomalies, much in the same way as behavioral economics identified deviations from the standard rational choice model.

⁴⁷Hefti (2018) studies the welfare implications in an entirely symmetric model if strategic firms need to compete for attention; also see De Clippel et al. (2014) for a cross-market welfare analysis with inattentive consumers. See Falkinger (2007) for an early emphasis about the importance of stimulus-driven attention for economics.

However, the formation of a consideration set does involve a goal-driven aspect in our model, in that the human brain responds to an increasing information load by tightening the attention span (Section 2.1). In this respect, one may note that our approach is less vulnerable to a possible criticism held against the rational inattention framework, according to which the key presumption motivating rational inattention – humans have information processing constraints – contradicts the assumption that consumers still manage to solve highly complex decision problems correctly.

In context of digitization, some have advocated decreasing search costs, which induces more search, intensifies competition and possibly promotes the firms in the tail (e.g., Brynjolfsson et al., 2003; Cachon et al., 2008; Bar-Isaac et al., 2012). Our model of competitive attention differs in at least two central aspects. First, the finite capacity model entails the opposite prediction that consumers respond by a diminished attention to an increasing information load, where we found the interrelated equilibrium forces not to be conducive to the tail firms. Evidence indicates that the empirical consideration sets in online shopping may be astonishingly small (Gu et al., 2012), and the prediction that consumers should search more (or at least threaten to do so) is generally hard to reconcile with the evidence presented in Section 2. Second, a central assumption of search models is that the consumer sampling probabilities are exogenously given, meaning that firms have no means of influencing their perception chances, which is a key difference to our model of competitive attention. The sampling chances in search models typically are assumed to be uniformly distributed, which is rejected by empirical data Granka et al., 2004; Skiera et al., 2010, and inconsistent with the power law dispersion of attention.

Informative Advertising According to the theory of informative advertising (see Bagwell, 2007 for a comprehensive survey), the information burden rests on the firms and not on the consumers, which is a difference to consumer search and rational inattention, but at most a superficial similarity to competitive attention. The key difference is that all information eventually reaching a consumer becomes a part of the consideration set in informative advertising. By contrast, in our model only the information that passes through the attention gate can enter consideration, where the amount of perceived information is endogenous and the selection described by the principle of relative salience. Intuitively, informative advertising works like throwing out many balls to hit as many urns with unbounded capacity as possible. The attention problem means throwing out sufficiently “thick” balls that must pave their way against other balls to enter an urn of finite capacity.

With respect to empirical data, competitive attention, requiring firms to become more salient than others, seems a better fit particularly with digitized competition, where becoming or

remaining “top-of-mind” has been persistently voiced as the key concern of advertising firms (Iyer et al., 2005).

Regarding theoretical predictions, our model claims that market entry increases i) aggregate attention expenditures, independent of the pre-existing market concentration, and ii) increases a firm’s expenditures if its profit increases. This is a strong contrast to Grossman and Shapiro (1984), where market entry decreases individual advertising expenditures, and also decreases aggregate expenditures if advertising costs are of the type (8). The key reason for these differences is rooted in the assumption of unbounded consumer attention, where market entry implies that, on average, consumers always become aware of more suitable products which reinforces competition and lowers the propensity to advertise, independent of how the market actually has become.

6 Conclusion

In a digitized economy, every second stands witness to an abundance of information produced, keeping track of which is simply impossible for the human brain. As a recent article in “The Economist” put it: “There is almost no limit to the supply of entertainment choices in every category, but people’s awareness of these products and their ability to find them is constrained by the time and attention they can spare”.⁴⁸

In this article, we ask how individual choices, competition and market structure evolve if attention is a competitive resource. Our model is based on two central tenets from psychology and neuroscience: People have limited information processing capacities and their allocation of attention is guided by relative salience. The former implies that attention is a rival, depletable resource, a consequence of which is that the set of objectively given information items and the set of subjectively perceived options can disagree. The latter implies that attention is a competitive resource, where the profit-maximizing firms are forced to make their choice options salient relative to the market. The equilibrium outcome then is a consequence of the choices made by rational firms that need to respect the laws of attention by which the consumer-side considerations sets are formed.

The model of competitive attention can account for – and connect – various empirical phenomena at the micro and macro level. Specifically, it offers an explanation why the “Paradox of Choices” prevails in a digitized economy, why firms continuously state “being on the consumer’s mind” as their top concern regarding advertising, why sales, profits or attention data persistently adopt power law dispersions, why there is a positive relation between profitability and marketing expenditures, why market shares and advertising expenditures move in lockstep,

⁴⁸ “Monetizing eyeballs”, Feb 11, 2017.

or why the evidence on the “Long Tail effect” is mixed at best.

We therefore believe that the competitive allocation of limited attention can and should play a major role for empirical assessments of the matter, and the scope for such examination seems large and promising. In this respect, we hope that our framework and its theoretical predictions may provide some guidance for future empirical work.

References

- ADAMIC, L. A. AND B. A. HUBERMAN (2000): “Power-law distribution of the world wide web,” *science*, 287, 2115–2115.
- ANDERSON, C. (2004): “The Long Tail,” *Wired Magazine*.
- ANDRIANI, P. AND B. MCKELVEY (2011): “From skew distributions to power-law science,” *The Sage handbook of complexity and management*, 254–273.
- ASTORNE-FIGARI, C., J. J. LÓPEZ, AND A. YANKELEVICH (2019): “Advertising for consideration,” *Journal of Economic Behavior & Organization*, 157, 653–669.
- BAGWELL, K. (2007): “The Economic Analysis of Advertising,” in *Handbook of Industrial Organization*, ed. by M. Armstrong and R. Porter, Amsterdam: Elsevier B.V., vol. 3.
- BAR-ISAAC, H., G. CARUANA, AND V. CUÑAT (2012): “Search, design, and market structure,” *American Economic Review*, 102, 1140–60.
- BAYE, M. R., J. R. J. GATTI, P. KATTUMAN, AND J. MORGAN (2009): “Clicks, Discontinuities, and Firm Demand Online,” *Journal of Economics Management Strategy*, 18, 935–975.
- BETTMAN, J. R., M. F. LUCE, AND J. W. PAYNE (1998): “Constructive consumer choice processes,” *Journal of consumer research*, 25, 187–217.
- BLANK, A. AND S. SOLOMON (2000): “Power laws in cities population, financial markets and internet sites (scaling in systems with a variable number of components),” *Physica A: Statistical Mechanics and its Applications*, 287, 279–288.
- BORDALO, P., N. GENNAIOLI, AND A. SHLEIFER (2015): “Competition for Attention,” *Reviews of Economic Studies*, 1–33.
- BOTTI, S. AND S. IYENGAR (2004): “The psychological pleasure and pain of choosing: when people prefer choosing at the cost of subsequent outcome satisfaction,” *Journal of personality and social psychology*, 87, 312–326.
- BRYNJOLFSSON, E., Y. HU, AND M. D. SMITH (2003): “Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers,” *Management Science*, 49, 1580–1596.
- (2010a): “Long tails vs. superstars: The effect of information technology on product variety and sales concentration patterns,” *Information Systems Research*, 21, 736–747.
- (2010b): “Long tails vs. superstars: The effect of information technology on product variety and sales concentration patterns,” *Information Systems Research*, 21, 736–747.
- BRYNJOLFSSON, E., Y. J. HU, AND D. SIMESTER (2011): “Goodbye pareto principle, hello long tail: The effect of search costs on the concentration of product sales,” *Management Science*, 57, 1373–1386.
- CACHON, G. P., C. TERWIESCH, AND Y. XU (2008): “On the effects of consumer search and firm entry in a multiproduct competitive market,” *Marketing Science*, 27, 461–473.
- CHAUVIN, K. W. AND M. HIRSCHY (1993): “Advertising, R&D expenditures and the market value of the firm,” *Financial management*, 128–140.
- CHERNEV, A., U. BÖCKENHOLT, AND J. GOODMAN (2015): “Choice overload: A conceptual review and meta-analysis,” *Journal of Consumer Psychology*, 25, 333–358.
- CHEVALIER, J. AND A. GOOLSBEE (2003): “Measuring Prices and Price Competition Online: Amazon.com and BarnesandNoble.com,” *Quantitative Marketing and Economics*, 1, 203–222.
- CHO, J. AND S. ROY (2004): “Impact of Search Engines on Page Popularity,” in *Proceedings of the 13th International Conference on World Wide Web*, New York: ACM Press, 20–29.

- CLEGG, R. G., C. DI CAIRANO-GILFEDDER, AND S. ZHOU (2010): “A critical look at power law modelling of the Internet,” *Computer Communications*, 33, 259–268.
- CRAWFORD, G. C., H. AGUINIS, B. LICHTENSTEIN, P. DAVIDSSON, AND B. MCKELVEY (2015): “Power law distributions in entrepreneurship: Implications for theory and research,” *Journal of Business Venturing*, 30, 696–713.
- DARDANONI, V., P. MANZINI, M. MARIOTTI, AND C. J. TYSON (2018): “Inferring cognitive heterogeneity from aggregate choices,” *Working paper*.
- DE CLIPPEL, G., K. ELIAZ, AND K. ROZEN (2014): “Competing for consumer inattention,” *Journal of Political Economy*, 122, 1203–1234.
- DE LOS SANTOS, B., A. HORTACSU, AND M. R. WILDENBEEST (2012): “Testing models of consumer search using data on web browsing and purchasing behavior,” *The American Economic Review*, 102, 2955–2980.
- DELLAROCAS, C. AND R. NARAYAN (2007): “Tall heads vs. long tails: Do consumer reviews increase the informational inequality between hit and niche products?” *Robert H. Smith School of Business Research Paper*.
- DELLAROCAS, C., J. SUTANTO, M. CALIN, AND E. PALME (2015): “Attention allocation in information-rich environments: the case of news aggregators,” *Management Science*, 62, 2543–2562.
- DIJKSTERHUIS, A. (2004): “Think different: the merits of unconscious thought in preference development and decision making,” *Journal of personality and social psychology*, 87, 586.
- DIJKSTERHUIS, A., M. W. BOS, L. F. NORDGREN, AND R. B. VAN BAAREN (2006): “On making the right choice: The deliberation-without-attention effect,” *Science*, 311, 1005–1007.
- DIJKSTERHUIS, A. AND L. F. NORDGREN (2006): “A theory of unconscious thought,” *Perspectives on Psychological science*, 1, 95–109.
- DIPRETE, T. A. AND G. M. EIRICH (2006): “Cumulative advantage as a mechanism for inequality: A review of theoretical and empirical developments,” *Annu. Rev. Sociol.*, 32, 271–297.
- EASTLICK, M. A., R. FEINBERG, AND C. TRAPPEY (1993): “Information overload in mail catalog shopping: how many catalogs are too many?” *Journal of Direct Marketing*, 7, 14–19.
- ELBERSE, A. (2008): “Should You Invest in the Long Tail?” *Harvard Business Review*, 86, 88–96.
- ELBERSE, A. AND F. OBERHOLZER-GEE (2008): “Superstars and Underdogs: An Examination of the Long-Tail Phenomenon in Video Sales,” *Working paper*.
- ELLISON, G. AND S. F. ELLISON (2009): “Search, Obfuscation and Price Elasticities on the Internet,” *Econometrica*, 77, 427–452.
- EPPLER, M. J. AND J. MENGIS (2004): “The concept of information overload: A review of literature from organization science, accounting, marketing, MIS, and related disciplines,” *The information society*, 20, 325–344.
- FALKINGER, J. (2007): “Attention economies,” *Journal of Economic Theory*, 133, 266–294.
- FENG, L., Y. HU, B. LI, H. E. STANLEY, S. HAVLIN, AND L. A. BRAUNSTEIN (2015): “Competing for attention in social media under information overload conditions,” *PloS one*, 10, e0126090.
- FINKEL, E. J., P. W. EASTWICK, B. R. KARNEY, H. T. REIS, AND S. SPRECHER (2012): “Online dating: A critical analysis from the perspective of psychological science,” *Psychological Science in the Public interest*, 13, 3–66.
- FLEDER, D. AND K. HOSANAGAR (2009): “Blockbuster culture’s next rise or fall: The impact of recommender systems on sales diversity,” *Management science*, 55, 697–712.
- FRANK, R. AND P. COOK (1995): “The-Winner-Takes-All-Society: Why the Few at The Top Get So Much More Than the Rest of Us,” *New York: Penguin Books*.
- GABAIX, X. (2016): “Power laws in economics: An introduction,” *Journal of Economic Perspectives*, 30, 185–206.
- GAFFEO, E., M. GALLEGATI, AND A. PALESTRINI (2003): “On the size distribution of firms: additional evidence from the G7 countries,” *Physica A: Statistical Mechanics and its Applications*, 324, 117–123.
- GENSCH, D. H. (1987): “A two-stage disaggregate attribute choice model,” *Marketing Science*, 6, 223–239.
- GHOSE, A. M. AND S. YANG (2009): “An Empirical Analysis of Search Engine Advertising: Sponsored Search in Electronic Markets,” *Management Science*, 55, 1605–1622.

- GRANKA, L. A., T. JOACHIMS, AND G. GAY (2004): “Eye-tracking analysis of user behavior in WWW search,” in *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*, 478–479.
- GROSSMAN, G. M. AND C. SHAPIRO (1984): “Informative Advertising with Differentiated Products,” *The Review of Economic Studies*, 51, 63–81.
- GRUBB, M. D. (2015): “Failing to choose the best price: Theory, evidence, and policy,” *Review of Industrial Organization*, 47, 303–340.
- GU, B., P. KONANA, AND H.-W. M. CHEN (2012): “Identifying consumer consideration set at the purchase time from aggregate purchase data in online retailing,” *Decision Support Systems*, 53, 625–633.
- GU, B., Q. TANG, AND A. B. WHINSTON (2013): “The influence of online word-of-mouth on long tail formation,” *Decision Support Systems*, 56, 474–481.
- HAUSER, J. R. AND B. WERNERFELT (1990): “An Evaluation Cost Model of Consideration Sets,” *The Journal of Consumer Research*, 16, 393–408.
- HAYNES, G. A. (2009): “Testing the boundaries of the choice overload phenomenon: The effect of number of options and time pressure on decision difficulty and satisfaction,” *Psychology & Marketing*, 26, 204–212.
- HEFTI, A. (2016): “On the relationship between uniqueness and stability in sum-aggregative, symmetric and general differentiable games,” *Mathematical Social Sciences*, 80, 83–96.
- (2017): “Equilibria in symmetric games: Theory and Applications,” *Theoretical Economics*, 12, 979–1002.
- (2018): “Limited Attention, Competition and Welfare,” *Journal of Economic Theory*, 178.
- HEFTI, A., S. LIU, AND A. SCHMUTZLER (2020): “Preferences, Confusion and Competition,” *CEPR Working Paper*.
- HEFTI, A. AND J. TEICHGRÄBER (2020): “Inequality in Models with a Competition for Market Shares,” *ECON Working Paper No. 375*.
- HEFTI, A. M. AND S. HEINKE (2015): “On the economics of superabundant information and scarce attention,” *Oeconomia*, 5, 37–74.
- HELGESON, J. G. AND M. L. URSIC (1993): “Information load, cost/benefit assessment and decision strategy variability,” *Journal of the Academy of Marketing Science*, 21, 13–20.
- HINDMAN, M. (2018): *The Internet trap: How the digital economy builds monopolies and undermines democracy*, Princeton University Press.
- HINZ, O., J. ECKERT, AND B. SKIERA (2011): “Drivers of the long tail phenomenon: an empirical analysis,” *Journal of management information systems*, 27, 43–70.
- HISANO, R. AND T. MIZUNO (2011): “Sales distribution of consumer electronics,” *Physica A: Statistical Mechanics and its Applications*, 390, 309–318.
- HODAS, N. O. AND K. LERMAN (2013): “Attention and visibility in an information-rich world,” in *2013 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, IEEE, 1–6.
- (2014): “The simple rules of social contagion,” *Scientific reports*, 4, 4343.
- HOTHKISS, G. (2006): “Eye Tracking Report: Google, MSN, and Yahoo! Compared,” *Enquiro, Kelowna, BC, Canada*.
- HOWARD, J. A. AND J. N. SHETH (1969): *The Theory of Buyer Behavior*, New York: John Wiley.
- HUBERMAN, B. A. (2003): *The laws of the Web: Patterns in the ecology of information*, MIT Press.
- INBAR, Y., S. BOTTI, AND K. HANKO (2011): “Decision speed and choice regret: When haste feels like waste,” *Journal of Experimental Social Psychology*, 47, 533–540.
- IYENGAR, S. S. AND M. R. LEPPER (2000): “When choice is demotivating: Can one desire too much of a good thing?” *Journal of personality and social psychology*, 79, 995.
- IYER, G., D. SOBERMAN, AND J. M. VILLAS-BOAS (2005): “The Targeting of Advertising,” *Marketing Science*, 24, 461–476.
- JACOBY, J. (1977): “Information load and decision quality: Some contested issues,” *Journal of Marketing Research*, 14, 569–573.

- JONES, J. P. (1990): "Ad spending: maintaining market share." *Harvard business review*, 68, 38–42.
- KAHNEMAN, D. (2011): *Thinking, fast and slow*, Macmillan.
- KAHNEMAN, D. AND A. HENIK (1981): "Perceptual organization and attention," in *Perceptual organization*, ed. by M. Kubovy and J. R. Pomerantz, Hillsdale, NJ: Lawrence Erlbaum Associates Inc.
- KELLO, C. T., G. D. BROWN, R. FERRER-I CANCHO, J. G. HOLDEN, K. LINKENKAER-HANSEN, T. RHODES, AND G. C. VAN ORDEN (2010): "Scaling laws in cognitive sciences," *Trends in cognitive sciences*, 14, 223–232.
- KOHLI, R. AND R. K. SAH (2006): "Market shares: Some power law results and observations," *Management Science*, 52, 1792–1798.
- LANCASTER, K. (1990): "The economics of product variety: A survey," *Marketing science*, 9, 189–206.
- LEE, B.-K. AND W.-N. LEE (2004): "The effect of information overload on consumer choice quality in an on-line environment," *Psychology & Marketing*, 21, 159–183.
- LENNIE, P. (2003): "The cost of cortical computation," *Current biology*, 13, 493–497.
- LI, C.-Y. (2016): "The more, the better? Why abundant information leads to unanticipated outcomes," *Telecommunications and Informatics*, 33, 834–847.
- LURIE, N. H. (2004): "Decision making in information-rich environments: The role of information structure," *Journal of Consumer Research*, 30, 473–486.
- MALHOTRA, N. K. (1982): "Information load and consumer decision making," *Journal of consumer research*, 8, 419–430.
- MANZINI, P. AND M. MARIOTTI (2014): "Stochastic choice and consideration sets," *Econometrica*, 82, 1153–1176.
- MASATLIOGLU, Y., D. NAKAJIMA, AND E. OZBAY (2012): "Revealed Attention," *American Economic Review*, 102, 2183–2205.
- MAUNSELL, J. AND S. TREUE (2006): "Feature-based attention in visual cortex," *TRENDS in Neurosciences*, 29, 317–322.
- MERTON, R. K. (1968): "The Matthew effect in science: The reward and communication systems of science are considered," *Science*, 159, 56–63.
- MILLER, G. A. (1956): "The magical number seven, plus or minus two: some limits on our capacity for processing information," *Psychological review*, 63, 81.
- MUKHERJEE, S. AND N. SRINIVASAN (2013): "Attention in preferential choice," in *Progress in brain research*, Elsevier, vol. 202, 117–134.
- MULLAINATHAN, S., J. SCHWARTZSTEIN, AND A. SHLEIFER (2008): "Coarse thinking and persuasion," *The Quarterly Journal of Economics*, 123, 577–619.
- NAPOLI, P. M. (2016): "Requiem for the long tail: Towards a political economy of content aggregation and fragmentation," *International Journal of Media & Cultural Politics*, 12, 341–356.
- NELSON, J. L. (2019): "The Persistence of the Popular in Mobile News Consumption," *Digital Journalism*, 1–16.
- NEWMAN, M. E. (2005): "Power laws, Pareto distributions and Zipf's law," *Contemporary physics*, 46, 323–351.
- NOTHDURFT, H. C. (2000): "Salience from feature contrast: additivity across dimensions," *Vision Research*, 40, 1183–1201.
- ORDANINI, A. AND J. C. NUNES (2016): "From fewer blockbusters by more superstars to more blockbusters by fewer superstars: how technological innovation has impacted convergence on the music chart," *International Journal of Research in Marketing*, 33, 297–313.
- OULASVIRTA, A., J. P. HUKKINEN, AND B. SCHWARTZ (2009): "When more is less: the paradox of choice in search engine use," in *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, 516–523.
- PAN, B., L. ZHANG, AND R. LAW (2013): "The complex matter of online hotel choice," *Cornell Hospitality Quarterly*, 54, 74–83.
- PARRA, J. F. AND S. RUIZ (2009): "Consideration sets in online shopping environments: the effects of search tool and information load," *Electronic Commerce Research and Applications*, 8, 252–262.
- PASHLER, H., ed. (1998): *Attention*, Psychology Press.

- PAYNE, J. W. (1982): "Contingent decision behavior." *Psychological bulletin*, 92, 382.
- PAYNE, J. W., J. R. BETTMAN, AND E. J. JOHNSON (1993): *The Adaptive Decision Maker*, Cambridge University Press.
- PELTIER, S. AND F. MOREAU (2012): "Internet and the Long Tail versus superstar effect debate: Evidence from the French book market," *Applied Economics Letters*, 19, 711–715.
- PERC, M. (2014): "The Matthew effect in empirical data," *Journal of The Royal Society Interface*, 11, 20140378.
- POORESMAEILI, A., D. R. BACH, AND R. J. DOLAN (2014): "The effect of visual salience on memory-based choices," *Journal of neurophysiology*, 111, 481–487.
- REED, W. J. (2001): "The Pareto, Zipf and other power laws," *Economics letters*, 74, 15–19.
- RODRIGUEZ, M. G., K. GUMMADI, AND B. SCHOELKOPF (2014): "Quantifying information overload in social media and its impact on social contagions," in *Eighth International AAAI Conference on Weblogs and Social Media*.
- ROSEN, S. (1981): "The economics of superstars," *The American Economic Review*, 71, 845–858.
- SCHEIBEHENNE, B., R. GREIFENEDER, AND P. M. TODD (2010): "Can there ever be too many options? A meta-analytic review of choice overload," *Journal of consumer research*, 37, 409–425.
- SCHRODER, H. M., M. J. DRIVER, AND S. STREUFERT (1967): *Human information processing: Individuals and groups functioning in complex social situations*, Holt, Rinehart and Winston.
- SHANNON, C. E. (1948): "A mathematical theory of communication," *The Bell System Technical Journal*, 27, 379–423.
- SIMS, C. A. (2003): "The Implications of rational inattention," *Journal of Monetary Economics*, 50, 665–690.
- SINHA, S. AND R. K. PAN (2006): "How a hit is born: The emergence of popularity from the dynamics of collective choice," *Econophysics and Sociophysics: Trends and Perspectives*, 417–447.
- SKIERA, B., J. ECKERT, AND O. HINZ (2010): "An analysis of the importance of the long tail in search engine marketing," *Electronic Commerce Research and Applications*, 9, 488–494.
- SMITH, M. D. AND E. BRYNJOLFSSON (2001): "Consumer Decision-Making at an Internet Shopbot: Brand still matters," *Journal of Industrial Economics*, 49, 541–558.
- STEVENS, S. S. (1957): "On the psychophysical law." *Psychological review*, 64, 153.
- STREUFERT, S. (1970): "Complexity and complex decision making: Convergences between differentiation and intergration approaches to the prediction of task performance," *Journal of Experimental Social Psychology*, 6, 494–509.
- STREUFERT, S. AND M. J. DRIVER (1965): "Conceptual structure, information load and perceptual complexity," *Psychonomic Science*, 3, 249–250.
- STREUFERT, S., M. J. DRIVER, AND K. W. HAUN (1967): "Components of response rate in complex decision-making," *Journal of Experimental Social Psychology*, 3, 286–295.
- SUTCLIFFE, K. AND K. E. WEICK (2009): "Information overload revisited," in *The Oxford handbook of organizational decision making*, Oxford University Press.
- TAN, T. F., S. NETESSINE, AND L. HITT (2017): "Is Tom Cruise threatened? An empirical study of the impact of product variety on demand concentration," *Information Systems Research*, 28, 643–660.
- TOFFLER, A. (1984): *Future shock*, vol. 553, Bantam.
- TREUE, S. (2003): "Visual attention: the where, what, how and why of saliency," *Current opinion in neurobiology*, 13, 428–432.
- TUCKER, C. AND J. ZHANG (2007): "Long tail or steep tail? A field investigation into how online popularity information affects the distribution of customer choices," *Working paper*.
- VAKRATSAS, D. AND T. AMBLER (1999): "How advertising works: what do we really know?" *Journal of Marketing*, 63, 26–43.
- WEBSTER, J. G. AND T. B. KSIAZEK (2012): "The dynamics of audience fragmentation: Public attention in an age of digital media," *Journal of communication*, 62, 39–56.
- WEBSTER, J. G. AND S.-F. LIN (2002): "The Internet audience: Web use as mass behavior," *Journal of Broadcasting & Electronic Media*, 46, 1–12.

- YANTIS, S. (1998): “Control of Visual Attention,” in *Attention*, ed. by H. Pashler, Psychology Press.
- ZHANG, J., Q. CHEN, AND Y. WANG (2009): “Zipf distribution in top Chinese firms and an economic explanation,” *Physica A: statistical Mechanics and its applications*, 388, 2020–2024.
- ZHONG, N. AND F. MICHAHELLES (2013): “Google play is not a long tail market: an empirical analysis of app adoption on the Google play app market,” in *Proceedings of the 28th Annual ACM Symposium on Applied Computing*, 499–504.
- ZHOU, W. AND W. DUAN (2012): “Online user reviews, product variety, and the long tail: An empirical investigation on online software downloads,” *Electronic Commerce Research and Applications*, 11, 275–289.

A Proofs

Proof Proposition 1 By contradiction, let $n > R^*(\tau)$ but $\Sigma \leq \bar{R}$; hence $\tau = 1$. As $R^*(1) \leq \min\{\bar{R}, n\}$, we must have $R^*(1) = \bar{R}$, as $R^*(1) = n$ contradicts $n > R^*(1)$. But then $R^*(1) = \bar{R} < n \leq \Sigma$, or $\Sigma > \bar{R}$, which is impossible. Hence information overload is necessary for choice overload. Next, if $R^*(\tau) < n$, then $\Sigma > \bar{R}$ by the previous claim, and thus $\tau < 1$. Defining $R^*(\Sigma) \equiv R^*(\tau(\Sigma))$ shows that $R'(\Sigma) < 0$, because $R'(\tau(\Sigma)) > 0$ and $\tau'(\Sigma) < 0$ by (3). ■

Proof Corollary 1 Let $\bar{n} > 0$ be the unique number implicitly defined by $\Sigma(\bar{n}) = \bar{R}$. Note that $\bar{n} \leq \bar{R}$. If $n < \bar{n}$, then also $n < \bar{R}$ and $\Sigma(n) < \bar{R}$, meaning that there is no information overload; hence also no choice overload. Thus $R^* = n$. If $n > \bar{n}$, then also $\Sigma(n) > \bar{R}$, implying that $\tau < 1$, and there must be choice overload ($R^* < n$). Defining $R^*(n) \equiv R^*(\tau(\Sigma(n)))$ shows that $R'(n) < 0$ as $R'(\tau) > 0$, $\tau'(\Sigma) < 0$ and $\Sigma'(n) > 0$. ■

Proof Lemma 1 : If $R \geq n$, then $\pi_j = 1 \forall j \in \mathcal{N}_n$, and hence $\sum_{j \in \mathcal{N}_n} \pi_j = n$. Let $R < n$, and define $B_j \equiv \{A \in \mathcal{A}(n, R) : j \in A\}$. Hence $\pi_j = P_{\mathcal{A}(n, R)}(B_j)$. As $\mathbb{A}_{n, R}$ is a discrete probability space, we can assume wlog that there is a (density) function $\varphi : \mathcal{A} \rightarrow [0, 1]$ with $\sum_{A \in \mathcal{A}} \varphi(A) = 1$, such that $P_{\mathcal{A}(n, R)}(B) = \sum_{A \in B} \varphi(A)$ for any $B \in \mathcal{P}$. Then, $P(B_j) = \sum_{A \in \mathcal{A}} \varphi(A) \mathbf{1}_{[j \in A]}(A)$, where $\mathbf{1}_{[j \in A]}$ is an indicator variable. Thus

$$\sum_{j \in \mathcal{N}_n} \pi_j = \sum_{j \in \mathcal{N}_n} P(B_j) = \sum_{j \in \mathcal{N}_n} \sum_{A \in \mathcal{A}} \varphi(A) \mathbf{1}_{[j \in A]}(A) = \sum_{A \in \mathcal{A}} \varphi(A) \sum_{j \in \mathcal{N}_n} \mathbf{1}_{[j \in A]}(A)$$

from which $\sum_{j \in \mathcal{N}_n} \pi_j = R$ follows, as $\sum_{j \in \mathcal{N}_n} \mathbf{1}_{[j \in A]}(A) = R \forall A \in \mathcal{A}$ and $\sum_{A \in \mathcal{A}} \varphi(A) = 1$. ■

Proof Proposition 2 In equilibrium, $f_c^* > 0$ is characterized by the first-order condition pertaining to (9), evaluated at Σ^* . This directly yields

$$f_c^* = \left(\frac{1}{\Sigma^*} \frac{RV(R)I}{\eta c} \right)^{\frac{1}{\eta-1}}$$

Integrating this equation and using $\int_{\mathcal{J}} f_c^* dc = \Sigma^*$ and rearranging gives (10), and $\hat{\pi}_c^* = f_c^*/\Sigma^*$ as well as $\pi_c^* = \hat{\pi}_c^* R$ follow immediately. Plugging (10) and (11) into (9) gives Π_c^* . Finally, $E_c^* = c(\hat{\pi}_c^* \Sigma^*)^\eta$. ■

Proof Proposition 3 For reasons of tractability, we split the proof in various steps.

Lemma A1 Let $\mathcal{J} = [1, 1 + \bar{c}]$, $x \in (0, 1)$, and $h : \mathcal{J} \rightarrow \mathbb{R}_{++}$ integrable. Define

$$S_x(\bar{c}) \equiv \frac{\int_1^{1+x\bar{c}} h(s) ds}{\int_1^{1+\bar{c}} h(s) ds}. \quad (\text{A.1})$$

Then $S'_x(\bar{c}) > 0 \forall x \in (0, 1)$ if $H(c) \equiv \int_1^{1+c} h(s) ds$ has a strictly decreasing c -elasticity on $(0, \bar{c})$.

Proof: Note that

$$S_x(\bar{c}) = \frac{\int_1^{1+x\bar{c}} h(s)ds}{\int_1^{1+\bar{c}} h(s)ds} = \frac{\int_0^{x\bar{c}} h(1+u)du}{\int_0^{\bar{c}} h(1+u)du} = \frac{\int_0^{x\bar{c}} g(s)ds}{\int_0^{\bar{c}} g(s)ds},$$

where $g(s) \equiv h(1+s)$. Differentiation then shows that $S'_x(\bar{c}) > 0 \forall x \in (0, 1)$ if and only if

$$\frac{\frac{cg(c)}{\int_0^c g(s)ds}}{\frac{\bar{c}g(\bar{c})}{\int_0^{\bar{c}} g(s)ds}} > \frac{\bar{c}g(\bar{c})}{\int_0^{\bar{c}} g(s)ds}. \quad (\text{A.2})$$

This inequality holds if $\frac{cg(c)}{\int_0^c g(s)ds}$ is strictly decreasing in c , i.e., $H(c)$ has a strictly decreasing c -elasticity.

□

Lemma A2 *Let $\mathcal{J} = [1, 1 + \bar{c}]$, $x \in (0, 1)$, and suppose that market shares are of the form*

$$\hat{\pi}_c = \frac{c^m}{\int_{\mathcal{J}} s^m ds}, \quad m < 0, \quad (\text{A.3})$$

Then $Z'_T(\bar{c}) > 0$, i.e., the market shares display a Superstar effect as the tail grows.

Proof: Note first that

$$Z_T(\bar{c}) = \frac{\int_1^{1+x\bar{c}} \hat{\pi}_c dc}{\int_1^{1+\bar{c}} \hat{\pi}_c dc} = \frac{\int_1^{1+x\bar{c}} c^m dc}{\int_1^{1+\bar{c}} c^m dc} = \frac{\int_1^{1+x\bar{c}} h(c)dc}{\int_1^{1+\bar{c}} h(c)dc},$$

for $h(c) \equiv c^m$. Let $H(c) \equiv \int_1^{1+c} h(s)ds$. To prove that $Z'_T(\bar{c}) > 0$, it suffices to show that $\tilde{H}(c) \equiv \frac{H'(c)c}{H(c)} = \frac{ch(1+c)}{H(c)}$ is strictly decreasing in c by Lemma A1. Integration yields

$$H(c) = \int_1^{1+c} s^m ds = \frac{(1+c)^{1+m} - 1}{1+m}$$

for $m \neq -1$.⁴⁹ Then

$$\tilde{H}(c) = \frac{c(1+c)^m(1+m)}{(1+c)^{1+m} - 1},$$

and $\tilde{H}'(c) < 0$ if and only if

$$\varepsilon((1+c)^\varepsilon - c\varepsilon - 1) < 0, \quad \varepsilon \equiv 1+m. \quad (\text{A.4})$$

If $m \in (-1, 0)$, then $\varepsilon \in (0, 1)$, and (A.4) holds if and only if $(1+c)^\varepsilon < 1 + c\varepsilon$. The last inequality is true because i) the LHS and RHS both are 1 as $c = 0$, ii) the LHS is increasing and strictly concave in c , and iii) its maximal slope is ε , which occurs at $c = 0$. Thus (A.4) holds for $m > -1$ and hence $Z'_T(\bar{c}) > 0$. If $m < -1$, then $\varepsilon < 0$ and (A.4) is equivalent to $(1+c)^\varepsilon > 1 + c\varepsilon$, where the truth of the latter inequality can be verified as before. □

⁴⁹The case where $m = -1$ corresponds to $\eta = 2$, which is already solved in the main text, hence let $m \neq -1$ in the following.

Corollary A1 *A growing tail induces a Superstar effect in $\hat{\pi}_c \forall x \in (0, 1)$.*

Proof: By (11), $\hat{\pi}_c$ has the form in Lemma A2. \square

Corollary A2 *A growing tail never induces a Long tail effect in $\hat{\pi}_c \forall x \in (0, 1)$.*

Proof: If $\hat{\pi}_c$ is of the form (A.3), a Long Tail Effect in market shares occurs if

$$Z_B(\bar{c}) = \frac{\int_1^{1+\bar{c}} c^m dc}{\int_1^{1+\bar{c}} c^m dc} = 1 - \frac{\int_1^{1+(1-x)\bar{c}} c^m dc}{\int_1^{1+\bar{c}} c^m dc} \quad (\text{A.5})$$

is strictly increasing in \bar{c} or, equivalently, $Z_T(\bar{c}) \equiv \frac{\int_1^{1+y\bar{c}} c^m dc}{\int_1^{1+\bar{c}} c^m dc}$, $y \equiv (1-x)$, is strictly decreasing in \bar{c} . But Lemma A2 implies that $Z'_T(\bar{c}) > 0 \forall y \in (0, 1)$, hence a Long Tail Effect cannot occur. \square

Corollary A3 *A growing tail decreases the joint market shares of the middle firms $\forall x \in (0, 1/2)$.*

Proof: The middle ground loses market share if

$$M_x(\bar{c}) \equiv \frac{\int_1^{1+(1-x)\bar{c}} c^m dc}{\int_1^{1+\bar{c}} c^m dc} = \frac{\int_1^{1+(1-x)\bar{c}} c^m dc}{\int_1^{1+\bar{c}} c^m dc} - \frac{\int_1^{1+x\bar{c}} c^m dc}{\int_1^{1+\bar{c}} c^m dc} = S_{1-x}(\bar{c}) - S_x(\bar{c})$$

is strictly decreasing in \bar{c} . For $m \neq -1$ integration yields

$$M_x(\bar{c}) = \frac{(1 + c(1-x))^\varepsilon - (1 + cx)^\varepsilon}{(1 + c)^\varepsilon - 1}, \quad \varepsilon \equiv 1 + m.$$

Proceeding similarly as in the last part of the proof of Lemma A2, $M'_x(\bar{c}) < 0$ follows. Further, as $S'_x(\bar{c}), S'_{1-x}(\bar{c}) > 0$ (Lemma A2), $M'_x(\bar{c}) < 0$ shows that the middle firms lose because the Superstar effect is stronger for smaller values of x ($S'_x(\bar{c}) > S'_{1-x}(\bar{c})$). \square

By (11)-(12) and (14), the shares of profits, attention, expenditures, revenues or sales of the top $x\%$ of firms are determined by the corresponding joint market shares of these firms. As $m \equiv \frac{1}{1-\eta}$, the claim in Proposition 3 now follows from Corollaries A1 - A3. \blacksquare

Proof Corollary 2 As before, it suffices to show the claim for market shares $\hat{\pi}_c$. Define $m \equiv \frac{1}{1-\eta} < 0$. Let $m \neq -1$. Then

$$Z_T(\bar{c}) = \frac{\int_{\mathcal{J}_x(\bar{c})} \hat{\pi}_c d_c}{\int_{\mathcal{J}(\bar{c})} \hat{\pi}_c d_c} = \frac{\int_1^{1+x\bar{c}} c^m dc}{\int_1^{1+\bar{c}} c^m dc} = \frac{(1 + x\bar{c})^\varepsilon - 1}{(1 + \bar{c})^\varepsilon - 1}, \quad \varepsilon \equiv 1 + m.$$

Taking the limit then directly shows that $\lim_{\bar{c} \rightarrow \infty} Z_T(\bar{c}) = x^\varepsilon \in (0, 1)$ if $\varepsilon \in (0, 1)$, and $\lim_{\bar{c} \rightarrow \infty} Z_T(\bar{c}) = 1$ if $\varepsilon < 0$. If $\varepsilon = 0$, such that $\hat{\pi}_c$ follows Zipf's Law, then $Z_T(\bar{c})$ is given by (16), from which $\lim_{\bar{c} \rightarrow \infty} Z_T(\bar{c}) = 1$ follows. The claim for $Z_B(\bar{c})$ is proved similarly, and the claim for $Z_M(\bar{c})$ follows from $Z_M(\bar{c}) = 1 - Z_T(\bar{c}) - Z_B(\bar{c})$. \blacksquare

Proof Proposition 4 (i): $C'_c(\bar{c}) < 0$ follows because $C_c(\bar{c}) = \hat{\pi}_c(\bar{c})RI$, where $\hat{\pi}'_c(\bar{c}) < 0$ by (11) and $R'(\bar{c}) < 0$.

$\frac{\partial}{\partial \bar{c}}(\pi_c(\bar{c})I - \pi_{c'}(\bar{c})I) < 0$ holds if $\hat{\pi}_c(\bar{c}) - \hat{\pi}_{c'}(\bar{c})$ is strictly decreasing in \bar{c} , which holds by (11).

(ii): Define $RV(R) \equiv \mathcal{V}(R)$, where $\mathcal{V}'(R) < 0$ by (13). By (12), the profit of firm c is $\Pi_c = \hat{\pi}_c(\bar{c})\mathcal{V}(R(\bar{c}))I^{\frac{\eta-1}{\eta}}$, and industry profits therefore are $\Pi(\bar{c}) = \int_{\mathcal{J}} \Pi_c dc = \mathcal{V}(R(\bar{c}))I^{\frac{\eta-1}{\eta}}$. Then $\Pi'(\bar{c}) > 0$ follows from $R'(\bar{c}) < 0$ and $\mathcal{V}'(R) < 0$. Consider now any incumbent firm c . Noting from (11) that $\frac{\hat{\pi}'_c(\bar{c})}{\hat{\pi}_c(\bar{c})} = -\hat{\pi}_{1+\bar{c}}$, where $\hat{\pi}_{1+\bar{c}}$ denotes the market share of the highest-cost incumbent, it can easily be verified that, for any $c \in \mathcal{J}(\bar{c})$, $\Pi'_c(\bar{c}) > 0$ iff $\hat{\pi}_{1+\bar{c}} < \frac{\mathcal{V}'(R)}{\mathcal{V}(R)}R'(\bar{c})$. Thus *all* incumbents either gain or lose from a growing tail.

(iii): Define $\Delta(\bar{c}) \equiv \Pi_c(\bar{c}) - \Pi_{c'}(\bar{c})$, for any two incumbents. Then by (12)

$$\Delta(\bar{c}) = \mathcal{V}(R(\bar{c}))I^{\frac{\eta-1}{\eta}} \frac{c^{\frac{-1}{\eta-1}} - c'^{\frac{-1}{\eta-1}}}{\int_{\mathcal{J}} c^{\frac{-1}{\eta-1}} dc},$$

and differentiation and (11) lead to $\Delta'(\bar{c}) > 0$ iff $\hat{\pi}_{1+\bar{c}} < \frac{\mathcal{V}'(R)}{\mathcal{V}(R)}R'(\bar{c})$, which is the same condition assuring increasing profits for each incumbent by (ii).

(iv): From $\Pi_c = \hat{\pi}_c\mathcal{V}(R)I^{\frac{\eta-1}{\eta}}$ we obtain that $\Pi'_c(\bar{c}) > 0$ iff

$$\hat{\pi}_{1+\bar{c}} < \frac{\mathcal{V}'(R)}{\mathcal{V}(R)}R'(\bar{c}). \quad (\text{A.6})$$

Let $\Sigma(R)$ denote the inverse function of $R = R(\Sigma)$ from Section 2.⁵⁰ Then, plugging $\Sigma(R)$ into (10) yields

$$\Sigma(R)^{\frac{\eta}{\eta-1}} = \left(\frac{\mathcal{V}(R)I}{\eta} \right)^{\frac{1}{\eta-1}} \int_{\mathcal{J}} c^{\frac{-1}{\eta-1}} dc \quad (\text{A.7})$$

as the equilibrium equation determining $R(\bar{c})$. Differentiating this equation and rearranging terms gives

$$R'(\bar{c}) = (\eta - 1) \frac{\mathcal{V}(R)^{\frac{1}{\eta-1}} \Sigma(R)}{\eta \mathcal{V}(R)^{\frac{1}{\eta-1}} \Sigma'(R) - \mathcal{V}(R)^{\frac{2-\eta}{\eta-1}} \mathcal{V}'(R) \Sigma(R)} \hat{\pi}_{1+\bar{c}},$$

where the denominator is negative in any stable equilibrium. Using this expression in (A.6) and some algebraic manipulation yield the condition stated in claim (iv).

(v): Note that like Superstar and Long Tail effect, the Matthew effect is defined over all firms, including the entrants. Consider first the case of profits, i.e., $F(c, \bar{c}) = \Pi_c(\bar{c})$. Then, by Definition 2, $M_x(\bar{c}) = \mathcal{V}(R(\bar{c}))I^{\frac{\eta-1}{\eta}}A(\bar{c})$, where

$$A(\bar{c}) \equiv \int_1^{1+x\bar{c}} \hat{\pi}_c dc + \int_1^{1+(1-x)\bar{c}} \hat{\pi}_c dc - 1.$$

As $\frac{\partial \mathcal{V}(R(\bar{c}))}{\partial \bar{c}} > 0$ it suffices for $M'_x(\bar{c}) > 0$ that $A'(\bar{c}) \geq 0$. The proof of Lemma A2 implies that both integrals indeed are strictly increasing in \bar{c} , which completes the proof for profits. \blacksquare

Proof Corollary 3 As $S_c = \hat{\pi}_c w(p)I$, $\hat{\pi}'_c(\bar{c}) < 0$ and $w'(p)p'(R)R'(\bar{c}) \leq 0$ in a stable equilibrium, $S'_c(\bar{c}) < 0$ follows. Further, $\frac{S_c}{S_{c'}} = \frac{\hat{\pi}_c}{\hat{\pi}_{c'}}$ remains constant for any two incumbents. Because $S'_c(\bar{c}), S'_{c'}(\bar{c}) < 0$,

⁵⁰As $R'(\Sigma) < 0$ by Proposition 1, the function $R(\Sigma)$ is invertible.

$\frac{\partial}{\partial \bar{c}}(S_c - S_{c'}) < 0$ follows. ■

Proof Proposition 5 First, note from (10) that the $\Sigma(R)$ -locus in Figure 4 is shifted upwards for $dI > 0$, which implies that $R'(I) < 0$.

(i): Because $\pi_c = \hat{\pi}(c)R(I)$, $R'(I) < 0$ and $\hat{\pi}_c$ is constant, $\pi'_c(I) < 0$ follows. The second claim holds as $\pi_c/\pi_{c'}$ is constant while perception chances decrease.

(ii): As $C_c = \hat{\pi}_c R(I)I$ and $\hat{\pi}_c$ is constant, $C_c(I)$ increases in I iff $R(I)I$ increases.

(iii): The derivation parallels the one of claim (iv) in the proof of Proposition 4. Total market attention $R(I)I$ increases strictly in I iff $\frac{R'(I)I}{R(I)} > -1$. Now, differentiation of (A.7) yields

$$R'(I) = \frac{1}{I} \frac{\mathcal{V}(R)\Sigma(R)}{\eta \mathcal{V}(R)\Sigma'(R) - \mathcal{V}'(R)\Sigma(R)},$$

where the denominator is negative in any stable equilibrium. Using this equation in the condition $\frac{R'(I)I}{R(I)} > -1$ yields the condition in (iii).

(iv): From

$$M_x(I) = \int_1^{1+x\bar{c}} C_c(I)dc + \int_1^{1+(1-x)\bar{c}} C_c(I)dc - 1 = \left(\int_1^{1+x\bar{c}} \hat{\pi}_c dc + \int_1^{1+(1-x)\bar{c}} \hat{\pi}_c dc \right) RI - 1$$

it follows that $dI > 0$ induces a (inverse) Matthew effect iff market attention $R(I)I$ increases (decreases).

(v): As $\Pi_c = \hat{\pi}_c \mathcal{V}(R(I))I^{\frac{\eta-1}{\eta}}$ and $\mathcal{V}'(R), R'(I) < 0$, it follows that each firm's profit increases in I . Further, the same type of argument as in the proof of (iv) shows that $dI > 0$ must results in a Matthew effect, because both the measure of consumers I and the per-consumer attention revenue $\mathcal{V}(R)$ increase. ■

B Supplementary Material (intended for Online Publication)

We begin by proving the claim of the main text about the supposed linearity of the perception chance per unit of attention. In Section B.2 we provide the technology-based and the psycho-physical foundations for the cost function (8) mentioned in the main text. Section B.3 discusses the finite capacity model from Section 2 through the lens of a recent theory about subconscious deliberation. In Section B.4 we briefly explore how to retrace an average attention space from aggregate choice data, and Section B.5 presents a dynamic model of attention. Section B.6 extends the equilibrium analysis to the case where firms differ in their abilities to monetize the attention they attract. Finally, Section B.7 shows that standard models of strategic price competition with horizontally differentiated goods yield the structure of the value function that we assumed in the main text.

B.1 Perception chance per unit of attention

We prove the claim from Section 2.2.2 that the perception chance per unit of attention must be a linear functional, meaning that $\hat{\pi}_j = f_j/\Sigma$.

Proposition B1 *The only C^1 -function $g : \mathbb{R} \rightarrow \mathbb{R}$ that satisfies $\int_0^n g\left(\frac{f_j}{\Sigma}\right) d_j = 1$ and $\Sigma = \int_0^n f_j d_j$ for any $n > 0$ and any integrable, non-negative function $f : [0, n] \rightarrow \mathbb{R}_{++}$ is $g(x) = x$.*

Proof: Let $X_n : [0, n] \rightarrow \mathbb{R}_{++}$, $n > 0$, be an integrable function with $\int_0^n X_n(s) ds = 1$. Let A be the set of all such functions. The proof builds on the following Lemma:

Lemma B1 *Let $g : \mathbb{R} \rightarrow \mathbb{R}$ be a C^1 -function, such that $\int_0^n g(X_n(s)) ds = 1$ whenever $X_n \in A$. Then, $g(x) = x$, i.e., $g(\cdot)$ must be linear.*

Proof: By contradiction, suppose that g is a C^1 -function with $\int_0^n g(X_n(s)) ds = 1$, $\forall X_n \in A$, but $\exists x, x' > 0$ such that $g'(x) \neq g'(x')$, implying that $g(\cdot)$ is not linear.

Then, $\exists n > \alpha > 0$ such that

$$\alpha x + (n - \alpha)x' = 1. \quad (\text{B.1})$$

To see this, let

$$\alpha(n) = \frac{1 - nx'}{x - x'}$$

The pair $(\alpha(n), n)$ satisfies (B.1). $\alpha(n) > 0 \Leftrightarrow n < \frac{1}{x'}$. $n > \alpha(n) \Leftrightarrow n > \frac{1}{x}$. Thus $n > \alpha(n) > 0$ whenever $n \in (\frac{1}{x}, \frac{1}{x'})$. Since $x > x' > 0$, we can always find such n . Let $\bar{n} \in (\frac{1}{x}, \frac{1}{x'})$ and denote $\bar{\alpha} \equiv \alpha(\bar{n})$. Then $(\bar{\alpha}, \bar{n})$ verifies (B.1), and $\bar{n} > \bar{\alpha} > 0$.

Consider the integrable function $X_{\bar{n}} : [0, \bar{n}] \rightarrow \mathbb{R}_+$ defined by $X_{\bar{n}}(s) = x$ for $s \in [0, \bar{\alpha})$, and $X_{\bar{n}}(s) = x'$ for $s \in [\bar{\alpha}, \bar{n}]$. As $\int_0^{\bar{n}} X_{\bar{n}}(s) ds = \bar{\alpha}x + (\bar{n} - \bar{\alpha})x' = 1$ we have $X_{\bar{n}} \in A$. By presumption, we then also must

have $\int_0^n g(X_{\bar{n}}(s))ds = 1$, hence

$$\bar{\alpha}g(x) + (\bar{n} - \bar{\alpha})g(x') = 1. \quad (\text{B.2})$$

Moreover, equation (B.2) must continue to hold for any small change in x, x' as long as (B.1) remains valid for the new values of x and x' . Solving (B.1) for x yields

$$x = \frac{1 - (\bar{n} - \bar{\alpha})x'}{\bar{\alpha}},$$

and plugging this into (B.2) gives

$$\bar{\alpha}g\left(\frac{1 - (\bar{n} - \bar{\alpha})x'}{\bar{\alpha}}\right) + (\bar{n} - \bar{\alpha})g(x') = 1.$$

Differentiation with respect to x' yields

$$-\bar{\alpha}g'(x)\frac{(\bar{n} - \bar{\alpha})}{\bar{\alpha}} + (\bar{n} - \bar{\alpha})g'(x') = 0,$$

which further implies that $g'(x) = g'(x')$, contradiction. \square

We are now ready to prove Proposition B1. Let $X_n \in A$. Then the function $f_n : [0, n] \rightarrow \mathbb{R}_{++}$ defined by $f_n(j) = X_n(j)$ has the property that $\int_0^n f_j dj = 1 = \Sigma$, and $f_n(j)/\Sigma = X_n(j)$. Thus, for any $X_n \in A$, a corresponding integrable function f_n can be found in this manner. In particular, the presumption in Proposition B1 requires that the conditions $\int_0^n g(X_n(j))dj = 1$ and $1 = \int_0^n X_n(j)dj$ must hold for any $X_n \in A$. Thus, the claim follows from Lemma B1. \blacksquare

B.2 Cost function: Main foundations

Steven's Law Building on empirical evidence dating back as far as the 1930's, Steven's Law of perception describes the relation between the (objective) intensity of stimuli, quantified by $e \geq 0$, and the magnitude of the (subjective) mental sensation $\psi(e)$ they invoke. The key observation summarized by Steven's Law is that "equal stimulus ratios generate equal sensation ratios", which requires that $\psi(e')/\psi(e) = (e'/e)^\alpha$, or equivalently

$$\psi(e) = ke^\alpha, \quad (\text{B.3})$$

where $k, \alpha > 0$ are constants.⁵¹ To see that (B.3) provides a foundation for (9), let $e_j \geq 0$ denote the (objective) intensity of the attention-seeking stimulus emitted by j , and $f_j = \psi_j(e_j) \geq 0$ is the (subjective) intensity of the mental impression on a receiver's mind triggered by e_j . Let producing a stimulus of intensity e be equally costly for all firms, with a price normalized to

⁵¹See Kello et al., 2010 for a survey on Steven's Law.

one. The impressions evoked on a consumer's mind compete for attention according to relative salience, such that $\hat{\pi}_j = \frac{\psi_j(e_j)}{\int_s \psi_s(e_s) ds}$. Thus $\Pi_j = \frac{\psi_j(e_j)}{\int_s \psi_s(e_s) ds} RV_j(R)I - e_j$. Assuming that $\psi_j(\cdot)$ is strictly increasing, we can equivalently state the payoff function in terms of the impression intensity f_j that firms seek to impose on consumers:

$$\Pi_j = \frac{f_j}{\Sigma} RV_j(R)I - \varphi_j(f_j),$$

where $\varphi_j(f_j) \equiv \psi^{-1}(f_j)$ and $\Sigma = \int_s f_s ds$. It is now easy to see that if $\psi_j(\cdot)$ follows Steven's Law, such that $\psi_j(e_j) = k_j e_j^{1/\eta}$, then $\varphi_j(f_j) = c_j f_j^\eta$ with $c_j \equiv k_j^{-\eta}$, which leads to (9).

Regarding interpretation, we first note from (B.3) that heterogeneity in k_j captures that some firms may be more efficient in generating mental impressions than others. As the previous derivation shows, such efficiency advantages translate directly into cost advantages. Furthermore, the cost parameter η captures how sensitive the intensity of the mental impression depends on the stimulus intensity. In particular, a large value of η means that a large change of stimulus intensity is needed to trigger a small change in mental impressions. In our model, this is further reflected in the relative perception chances as $\frac{\pi_j}{\pi_s} = \frac{f_j}{f_s} = \frac{k_j}{k_s} \left(\frac{e_j}{e_s} \right)^{1/\eta}$. Thus,

$$\frac{d(\pi_j/\pi_s)}{\pi_j/\pi_s} = \frac{1}{\eta} \frac{de_j}{e_j} \quad \forall j \neq s,$$

showing that η fully determines how elastic relative perception chances respond to a 1%-change in effort.

Attention Production Function We now provide an alternative, more traditional foundation for (9). Let $\pi_j = \frac{f_j}{\Sigma} R$ be the aspired perception chance given consideration set size R and total volume Σ . Suppose that f_j , the intensity of the attention-seeking activities of firm j , is the output from a multi-unit input process with production function $H_j : \mathbb{R}_+^m \rightarrow \mathbb{R}_+$, $f_j = H(a_j x_j)$, where $x_j \in \mathbb{R}_+^m$ is a vector of inputs acquired on competitive markets, $a_j \leq 1$ is an efficiency parameter, and $H(\cdot)$ is a strictly increasing production function homogeneous of degree $1/\eta < 1$.⁵² If $w \in \mathbb{R}_{++}^m$ denotes the factor prices, and

$$C(w, f, a) \equiv \min_{x \geq 0} w \cdot x \quad s.t. \quad H(ax) \geq f,$$

it follows from standard optimization theory that $C(w, f, a) = 1/af^\eta C(w, 1, 1)$. Normalizing $C(w, 1, 1) = 1$ and $1/a_j \equiv c_j \geq 1$, it follows for a firm with efficiency a_j that attaining a desired

⁵²A similar result is obtained from assuming that $f_j = a_j H(x_j)$.

intensity of f_j comes at a cost of $C_j(f_j) = c_j f_j^\eta$.

B.3 On the optimal consideration set size

In this section, we revisit the notion of optimal consideration set size from Section 2.1. We first show that the key prediction $R'(\Sigma) < 0$ also arises for more general objective functions, e.g., reflecting the indirect utility associated with R . In Section B.3.2 we connect our general approach to the recent literature on the “power of the subconscious”.

B.3.1 Value-based consideration set size

It is conceivable that the cost function $H(\cdot)$ in (1) depends on the processing rate τ . One could, for example, imagine that a lower processing rate τ , ceteris paribus, could increase the (opportunity) costs of devoting mental resources to the decision task. Another reason is that such a cost function can capture the case, where consideration set size $R = \tau T$ enters (1) via a general function $z(R)$, such that T maximizes $z(\tau T) - H(T)$. Once we allow for $H(T; \tau)$, the assumption from Section 2.1 that R enters the objective function linearly is not restrictive. To see this, note that the optimality condition is $\tau = \frac{H'(T)}{z'(\tau T)} \equiv H_T(T; \tau)$ for any given τ, T . Then, the appropriate cost function $H(T; \tau)$ is the anti-derivative of $H_T(T; \tau)$ with respect to T , and maximizing $\tau T - H(T; \tau)$ is equivalent to maximizing $z(\tau T) - H(T)$.

It is easy to see that the key condition $R'(\Sigma) < 0$ whenever overload occurs continues to hold if $z(R)$ – the value of consideration set size R – is given by an arbitrary, strictly increasing C^2 -function, and $z(\tau T) - H(T)$ is strongly quasi-concave in T . Due to overload, an interior solution must occur with $R^* < n$. This solution is uniquely determined by the first-order condition $\tau = \frac{H'(T)}{z'(\tau T)}$, and as $R = \tau T$ it follows that R^* must solve $\tau z'(R) = H'(R/\tau)$. The Implicit Function Theorem then reveals that $R'(\tau) > 0$, from which $R'(\Sigma) < 0$ follows.

As an illustration, suppose that a consumer has symmetric CES-preferences over all existing goods. If all goods are priced at p , the consumer buys an equal amount $x = \frac{w}{Rp}$ of all goods in her consideration set $A \in \mathcal{A}(R)$, and the expected utility from a consideration set size R is

$$z(R) = \sum_{A \in \mathcal{A}} P(A) \left(\sum_{j \in A} x_j^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} = \frac{w R^{\frac{1}{\sigma-1}}}{p},$$

where $\sigma > 1$ is the elasticity of substitution. If $h(\Sigma) = \Sigma$ and $H(T) = \delta T^\mu$, then for $\gamma \equiv \frac{1}{\sigma-1}$

we obtain $H(T; \tau) = \frac{w}{p(\mu-\gamma+1)} \tau^{1-\gamma} T^{\mu-\gamma+1}$, and the solution $R(\Sigma)$ in case of overload is

$$R(\Sigma) = \alpha \Sigma^{-\beta}, \quad \beta \equiv \frac{\mu}{\mu - \gamma} \in (1, \infty)$$

for $\mu > \gamma$, which formally coincides with (4). This derivation suggests an additional interpretation of β in (4). In the CES context, a larger value of β occurs, the weaker substitutes the products are (smaller σ). Accordingly, R responds more sensitively to changes in Σ , the smaller σ is. The intuition is that the size of the consideration set becomes less relevant if the products are rather substitutes than complements.

A similar procedure yields $z(R)$ -functions for discrete choice models. Let the value of option j to a consumer be $X_j = u - p_j + \varepsilon_j$, where u is a constant, p_j is the price, and ε_j is an iid random variable with density $f(e_j)$. Then $z(R)$ is given by the expected first order statistic $z(R) = u - p + R \int z f(z) F(z)^{R-1} dz$. For example, if ε_j is uniformly distributed on $[0, 1]$, then $z(R) = u - p + \frac{R}{R+1}$.

B.3.2 Consciousness

Whether the linear formulation in (1) or an indirect utility $z(R)$ is more appropriate may depend on how experienced a decision-maker is with the choice options. In particular, with lack of experience, or with low-involvement products, it seems particularly reasonable to assume a “guess” in thus that the value of consideration is roughly proportional to its size, rather than a perfect forward-looking derivation of the true expected utility, which requires to correctly anticipate all possible consideration sets that could arise.

More generally, the cognitive process guiding the pre-selection of options to pay attention to may be substantially different from the one determining choice in a given consideration set. Intuitively, the evaluation of a given (small) choice set is much more of a conscious deliberation process compared to the choice of how much information to consider. The latter is a far more complex problem, that may easily come up against the limit to information processing. In such a situation, intuition and unconscious deliberation may become decisive. Dijksterhuis et al. (2006); Dijksterhuis and Nordgren (2006); Dijksterhuis (2004); Mukherjee and Srinivasan (2013) argue that such integrated problems are solved by the human brain with “power of the unconscious”. In the “Unconscious Thought Theory”, conscious thought (“deliberation with attention to the problem”) is a very precise and rule-based type of thinking (arithmetic, analytic, comparative,...), that can only parse a small information set Dijksterhuis (2004). In terms of our framework, this conscious thought process is suitable to describe the process of choosing an

option from a given and sufficiently small consideration set. By contrast, unconscious thought addresses a form of deliberation, where by the sheer complexity of the problem, the brain decides “in the background”, relying on intuitive principles and memory. In particular, this suggests to uncouple the precise valuation of R by the brain during the consideration set formation process from a possible expected utility that requires, in principle, to work through all possible consideration consequences in a fully analytical way.

B.4 Attention data: Retracing the allocation of attention

A recent literature has argued that it may be possible to elicit attention and preferences from sufficiently rich choice data Masatlioglu et al. (2012); Manzini and Mariotti (2014). For example, Dardanoni et al. (2018) demonstrate that if all agents have the same strict linear preference order and the likelihood of any possible consideration set is uniformly distributed, then the probability distribution over the various consideration set sizes can be calculated from aggregate choices.⁵³ By contrast, our attention framework allows us to retrace the attention space $\mathbb{A} = (\mathcal{A}, \mathcal{P}, P)$ that best matches a given aggregate attention data on average, independent from preferences. That is, we can learn the average attention allocation $P(\cdot)$ over the possible consideration sets, which Dardanoni et al. (2018) assume to be uniformly distributed.

The key observation we use is the simple connection between the consideration set size R and the perception probabilities π_j identified by Lemma 1. Suppose that a number of I consumers generates data that yield some dispersive measure of of attention, e.g., online clicks frequencies. Say that we observe the following dispersion of clicks on $n = 10$ different items produced by $I = 10'000$ people:

$$\{8620, 4926, 2815, 1608, 919, 525, 300, 172, 98, 56\}.$$

This data is consistent with an average user having a consideration set size of (approximately) $R = 2$. To see why, denote the observed clicks in the above list by \hat{C}_j , $j = 1, \dots, n$. Then, Lemma 1 assures that $\sum_s \hat{C}_s / I = R$, because $\hat{C}_j / I = \pi_j$, $\forall j$, gives the average perception chance of option j , and $\sum_s \pi_s = R$.

Once R is known, the set of all consideration sets $\mathcal{A}(R)$ is determined. It then is possible to derive the allocation of attention $P : \mathcal{A}(R) \rightarrow [0, 1]$ that best describes the data on average by putting slightly more structure on the attention allocation process. An intuitive possibility is to assume an “order-independent” attention allocation rule. Let $\mathcal{N} = \{1, \dots, n\}$ and \tilde{P} be a

⁵³This result extends to heterogeneous preferences, but requires i) that preferences and attention are independently distributed, and ii) to observe the decisions by the same agents for multiple choice situations.

probability measure on $(\mathcal{N}, \mathcal{P}(\mathcal{N}))$ with

$$P(A) \equiv \tilde{P}(A|\mathcal{A}) = \frac{\tilde{P}(A)}{\tilde{P}(\mathcal{A})} = \frac{\prod_{j \in A} \hat{\pi}_j}{\sum_{A \in \mathcal{A}} \prod_{j \in A} \hat{\pi}_j}, \quad A \in \mathcal{A}, \quad (\text{B.4})$$

where $\hat{\pi}_j$ is the unitary sampling chance of option j , i.e., $\hat{\pi}_j = \tilde{P}(\{j\})$. Intuitively, $\hat{\pi}_j$ is the probability to “draw” item j from the urn containing all n items, and (B.4) constructs the probability of a consideration set $P(A)$ by assuming that the order of the draws is irrelevant.⁵⁴ Because $\pi_j = \sum_{A \in \mathcal{B}_j} P(A)$ for each $j \in \mathcal{N}$, the unitary sampling chances $\hat{\pi}_1, \dots, \hat{\pi}_n$ can be derived as the solution of the $n - 1$ equations

$$\pi_j = \frac{\sum_{A \in \mathcal{B}_j} \prod_{j \in A} \hat{\pi}_j}{\sum_{A \in \mathcal{A}} \prod_{j \in A} \hat{\pi}_j}, \quad j = 1, \dots, n - 1$$

and $\sum \hat{\pi}_s = 1$ for a given dispersion of perception chances π_1, \dots, π_n . These perception chances corresponds to the relative frequencies of the clicks on the various items in our example. With the above data, the unique numerical solution to the problem is

$$(\hat{\pi}_1, \dots, \hat{\pi}_{10}) \cong (0.69, 0.14, 0.08, 0.042, 0.024, 0.013, 0.008, 0.004, 0.002, 0.001)$$

Once the $\hat{\pi}_j$ are known, (B.4) can again be used to calculate $P(A) \forall A \in \mathcal{A}$. For example, our data implies that $P(1, 2) = 0.395$, while $P(1, 8) = 0.01$. Put differently, a representative consumer has $R = 2$, and if all consumers had $R = 2$, then about 40% of them would only compare options 1 and 2, while only 1% would compare options 1 and 8.

B.5 Dynamic Attention

In this section, we study a dynamic version of the model, where firms compete in attention in each period, but the attention they manage to attract depends on past attention as well. We show that the equilibrium of our static model corresponds to the steady-state equilibrium of its dynamic version. Moreover, we find that the path-dependence of the attention process has a similar effect on the dispersion of attention or profits as consumer sensitivity η .

Indexing firms with cost types $c \in [1, 1 + \bar{c}] \equiv \mathcal{J}$, suppose that the firms’ perception chances

⁵⁴For example, if $A = \{1, 2\}$ then, assuming sampling with replacement, $\tilde{P}(A) = \hat{\pi}_1 \hat{\pi}_2$, meaning that the first draw was $j = 1$ and the second draw was $j = 2$. We thus ignore that the order could also have been reversed.

per unit of attention evolve according to a discrete-time dynamic process of the form

$$\hat{\pi}_c(t) = \frac{\tilde{f}_c(t)}{\Sigma(t)}, \quad \tilde{f}_c(t) \equiv f_c(t)h(\hat{\pi}_c(t-1)), \quad \Sigma(t) \equiv \int \tilde{f}_s(t)ds, \quad (\text{B.5})$$

where $h(\cdot)$ is an increasing function. Thus, each firm's current attention $\pi_c(t) = \hat{\pi}_c(t)R$ increases in its current attention-seeking efforts, $f_j(t)$, as well as in the attention $\hat{\pi}_j(t-1)$ it received in the past. Intuitively, $\tilde{f}_c(t)$ can be viewed as the effective attention stimulus produced by the firm in period t . Such “attention-generates-attention” is a property associated, e.g., with collaborative filters, such as with popularity-based recommender systems (Fleder and Hosanagar, 2009). Then, firm profit in period t is of the form

$$\Pi_c(t) = \frac{f_c(t)h(\hat{\pi}_c(t-1))}{\Sigma(t)}RV(R)I - cf_c(t)^\eta. \quad (\text{B.6})$$

Let the density function $\hat{\pi}_c(0)$ be an exogenous initial condition. We suppose that for $t = 1, 2, \dots$ each firm myopically chooses her period attention effort $f_c(t)$ to maximize (B.6), taking $\Sigma(t)$ and $\hat{\pi}_c(t-1)$ as given.⁵⁵ This leads to a sequence of effort functions $\{f(t)\}_{t=1}^\infty$, $f(t) : \mathcal{J} \rightarrow \mathbb{R}_{++}$, where for any given $c \in \mathcal{J}$, $f_c(t)$ maximizes (B.6). The equilibrium dynamics of attention thus are given by the sequence $\{\hat{\pi}_c(t)\}_{t=0}^\infty$ of density functions $\hat{\pi}(t) : \mathcal{J} \rightarrow \mathbb{R}_{++}$, where for each $c \in \mathcal{J}$ and each $t \geq 1$, $\hat{\pi}_c(t)$ is determined by $\{f(t)\}_{t=1}^\infty$ and (B.5).⁵⁶

Replacing $f_c(t)$ by $\hat{\pi}_c(t)$ from (B.5) into the optimality condition $\frac{\partial}{\partial f_c(t)}\Pi_c(t) = 0$, the period t equilibrium value of $\hat{\pi}_c(t)$ solves

$$\frac{RV(R)}{\eta}I = \frac{c}{h(\hat{\pi}_c(t-1))^\eta} \hat{\pi}_c(t)^{\eta-1} \Sigma(t)^\eta \quad (\text{B.7})$$

As (B.7) shows, the effective attention costs $\frac{c}{h(\hat{\pi}_c(t-1))^\eta}$ borne by the firms depend dynamically on attention. A firm managing to attract a lot of attention in the past faces lower costs, *ceteris paribus*, of maintaining its attention share in the present.

We explore the equilibrium dynamics for the tractable case $h(x) = x^\delta$, where a larger value of $\delta \in [0, 1)$ means that the allocation of attention exerts a stronger path dependence.⁵⁷

Proposition B2 *For $\eta > \frac{1}{1-\delta}$, the above dynamic model has a unique steady-state. In the*

⁵⁵Alternatively, one could seek to solve the full dynamic programming problem, where firms discount their future payoffs and optimally choose their sequence of effort $\{f_c(t)\}_t$ over time. Beside substantial formal complications, the main difference is that forward-looking firms seek to smooth their attention expenditures over time, taking into account that today's attention expenditures affect tomorrow's attention costs.

⁵⁶For simplicity we take $R < n$ as constant and given by its steady-state value.

⁵⁷A steady-state of the model can be shown to exist more generally.

steady-state, the perception chances $\hat{\pi}_c^*$, and profit levels Π_c^* are given by

$$\hat{\pi}_c^* = \frac{\left(\frac{1}{c}\right)^{\frac{1}{\eta(1-\delta)-1}}}{\int_{\mathcal{J}} \left(\frac{1}{s}\right)^{\frac{1}{\eta(1-\delta)-1}} ds}, \quad \Pi_c^* = \hat{\pi}_c^* RV(R) I \frac{\eta-1}{\eta}. \quad (\text{B.8})$$

Proof: Using $h(x) = x^\delta$ in (B.7) and rearranging yields

$$\frac{\hat{\pi}_c(t)^{\eta-1}}{\hat{\pi}_c(t-1)^{\delta\eta}} = \frac{RV(R) I}{\eta \Sigma(t)^\eta} \frac{1}{c} \quad (\text{B.9})$$

From (B.9) it follows that a unique steady-state $(\hat{\pi}_c^*, \Sigma^*)$ exists if $\eta > \frac{1}{1-\delta}$ where, using $\int_{\mathcal{J}} \hat{\pi}_s ds = 1$,

$$\Sigma^* = \left(\frac{RV(R) I}{\eta} \right)^{\frac{1}{\eta}} \left(\int_{\mathcal{J}} c^{\frac{-1}{\eta(1-\delta)-1}} dc \right)^{\frac{\eta(1-\delta)-1}{\eta}},$$

and $\hat{\pi}_c^*$ is determined by (B.8). The expression for steady-state profit levels follows from plugging (B.5) in (B.6) and evaluating this expression at $\hat{\pi}_c^*, \Sigma^*$. ■

Expression (B.8) is identical, up to δ , to its static counterpart in Proposition 2. In particular, a Power Law dispersion of attention and profits results in the steady-state, and the comparative statics of the steady-state with respect to R , n or I adopt the same pattern as in the main text.

In addition, (B.8) shows that a stronger path dependence in the allocation of attention (larger δ) magnifies the long-run differences in attention and profits between the firms. The intuition is that a stronger path dependence decreases the equilibrium costs borne by firms which attracted more attention in the past.

To study the transitional dynamics of the model, we assume an equal initial dispersion of attention, i.e., $\hat{\pi}_c(0) = 1/\bar{c}$, $\forall c \in \mathcal{J}$. Figure 7 illustrates the dynamics for the discrete case of 10 firms with cost parameters $c \in \{1, 2, \dots, 10\}$, $\eta = 2$ and $\delta = 1/4$. The transition dynamics are such that $\hat{\pi}(t)$ features a gradually increasing inequality as $t \rightarrow \infty$, where the steady-state distribution $\hat{\pi}_c$ is approximated by a sequence of clockwise rotations of $\hat{\pi}_c(t)$. This pattern holds more generally.

Corollary B1 *Let $\eta > \frac{1}{1-\delta}$ and $\hat{\pi}_c(0) = 1/\bar{c}$, $\forall c \in \mathcal{J}$. Then $\hat{\pi}(t')$ is a clockwise rotation of $\hat{\pi}_j(t)$ whenever $t' > t$ as in Figure 7.*

Proof: Let $c < c'$ and note from (B.9) that

$$\frac{\hat{\pi}_c(t)}{\hat{\pi}_{c'}(t)} = \left(\frac{c'}{c} \right)^{\frac{1}{\eta-1}} \left(\frac{\hat{\pi}_c(t-1)}{\hat{\pi}_{c'}(t-1)} \right)^{\frac{\delta\eta}{\eta-1}}.$$

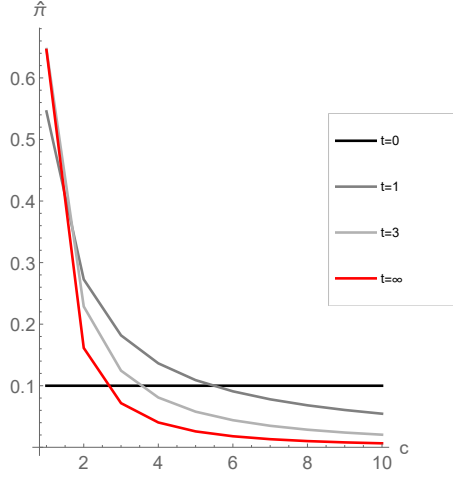


Figure 7: Equilibrium dynamics

Because $\hat{\pi}_c(0) = \hat{\pi}_{c'}(0)$ it follows that $\frac{\hat{\pi}_c(t)}{\hat{\pi}_{c'}(t)}$ must be strictly increasing in t . As such an increasing ratio condition must hold for any $c < c'$ it follows that $\pi(t')$ must be a clockwise rotation of $\pi(t)$ for $t' > t$ (Hefti and Teichgräber, 2020). ■

B.6 Heterogeneous Abilities to Monetize Attention

In this section, we consider what happens if firms differ in their abilities to monetize the attention they receive by assuming that $V_j(R) > V_k(R)$ whenever $j < k$; as before, firms are ordered left-to-right. We restrict attention to equilibria with choice overload ($R < n$). Such heterogeneity could, e.g., reflect the differential revenues earned from quality-differentiated products (Rosen, 1981). In this spirit, we interpret firm j as offering a higher-quality product than firm k , while other interpretations are possible.

The equilibrium shares many key properties with its counterpart from Section 3, such as the Power Law dispersions and various distributional comparative-statics. However, there are some differences, because changes in R now can have differential effects due to the different abilities of monetizing attention.

Section B.6.1 derives the equilibrium properties and the effects of a change in the consideration set size, similar to its counterpart in Section 3. Section B.6.2 considers the distributional comparative-static effects of a growing market size and a growing tail similar to Section 4. All proofs are in Section B.6.3.

B.6.1 Equilibrium Analysis

To make the effects of a differential ability to monetize the attention received most salient, we abstract from heterogeneity in attention-seeking abilities, and normalize $c_j = c_k = 1 \forall j, k$. The payoff of a firm $j \in [0, n]$ is

$$\Pi_j = \frac{f_j}{\Sigma} R V_j(R) I - f_j^\eta. \quad (\text{B.10})$$

Compared to the main text, this version of the model is formally more challenging because there is firm heterogeneity in terms of different *functions* $V_j(\cdot)$, rather than coefficients c_j . Formally, we assume that $V : \mathbb{R}_+ \times \mathbb{R}_{++} \rightarrow \mathbb{R}_{++}$, $(j, R) \mapsto V_j(R)$ is continuous, strictly increasing in j and strictly decreasing and continuously differentiable in R . Thus, for any firm $j \in [0, n]$, there is a unique revenue function $V_j(R)$, which we think of as representing its quality type, just as different cost coefficients identified different cost types. Because in the current model, the “type” is a function rather than a number, it is more convenient to index firms by j rather than by their types. Using $f_j = \hat{\pi}_j \Sigma$, the equilibrium FOC can be stated as

$$\frac{RI}{\eta} = \frac{1}{V_j(R)} \hat{\pi}_j^{\eta-1} \Sigma^\eta, \quad (\text{B.11})$$

where the RHS of (B.11) is the marginal cost of attention per unit of revenue. To make the connection to the model from the main text explicit, recall from Section 3 that the equilibrium optimality condition for a firm with cost type c is

$$\frac{RI}{\eta} = \frac{c}{V(R)} \hat{\pi}_c^{\eta-1} \Sigma^\eta. \quad (\text{B.12})$$

For a given R , we thus can map (B.11) into (B.12) by defining $c \equiv \frac{1}{V_j(R)}$. That is, for given R a stronger quality type is as if this type had lower attention costs. The only meaningful difference thus is that in the current model a change in R may have a differential impact for different quality types.

Proposition B3 *A unique attention equilibrium (f^*, Σ^*) exists, and total information load Σ^* , perception chances π_j^* , profits Π_j^* and attention expenditures E_j^* are*

$$\Sigma^* = \left(\frac{RI}{\eta} \right)^{\frac{1}{\eta}} K^{\frac{\eta-1}{\eta}} \quad K \equiv \int_0^n V_s(R)^{\frac{1}{\eta-1}} ds \quad (\text{B.13})$$

$$\hat{\pi}_j^* = \frac{f_j^*}{\Sigma^*} = \frac{V_j(R)^{\frac{1}{\eta-1}}}{K}, \quad \pi_j^* = \hat{\pi}_j^* R, \quad (\text{B.14})$$

$$\Pi_j^* = \hat{\pi}_j^* R V_j(R) I \frac{\eta-1}{\eta}, \quad E_j^* = \hat{\pi}_j^* \frac{R V_j(R)}{\eta} I, \quad C_c^* = \pi_j^* I. \quad (\text{B.15})$$

The equilibrium structure in Propositions 2 and B3 are similar. In particular, competitive attention leads to a power law dispersion of attention (clicks), profits and expenditures over quality types for a given R . To see this, fix R , and let $v \equiv v_j(R) \in \mathbb{R}_{++}$ be firm j 's quality type, such that $[\underline{v}, \bar{v}] = [V_0(R), V_n(R)]$ is the set of quality types. Then, expressing Proposition B3 in terms of quality types v , the analogue to (B.14) is $\hat{\pi}_v = \frac{v^{\frac{1}{\eta-1}}}{K}$, $K = \int_{\underline{v}}^{\bar{v}} s^{\frac{1}{\eta-1}} ds$ showing the power law shape over v .

Firms with a stronger ability to monetize their attention can also afford to maintain higher attention expenditures, resulting in a larger market share. The reason for the power law shape of the market shares parallels the one from the main text. Competitive forces equate the marginal attention costs per unit of revenue $V_j(R)$, and the functional form of the attention costs, as implied by Steven's Law, yields

$$\frac{\hat{\pi}_v}{\hat{\pi}_{v'}} = \left(\frac{v}{v'} \right)^{\frac{1}{\eta-1}} \quad (\text{B.16})$$

for any two quality types $v > v'$, from which it directly follows that $\hat{\pi}(\cdot)$ must obey a power law. Further, (B.16) shows that the sensitivity of consumer to the attention stimuli, η , continues to play a key role for market inequality and market response. As before, small differences in quality can lead to large differences in attention, revenues and profits if consumers are highly sensitive to (changes of) the attention stimuli (low η), *ceteris paribus*.

The two essential differences to the case where firms differ only in their abilities to attract attention are that i) revenues $V_j(R)I$ conditional on attracting attention now are heterogeneous, and ii) the equilibrium market share, $\hat{\pi}_j(R)$, now depends directly on R through its effect on $V_j(R)$. These two differences imply some disparities to the equilibrium in Section 3.

First, the profit share distribution now must be a clockwise rotation of the distribution of attention shares, as sketched in Figure 8 (a), whereas the two coincide with symmetric revenues. To see this, note that $\hat{\pi}_j^*$ is j 's market share of total attention RI , while $s_j^* \equiv \Pi_j^* / \int_0^n \Pi_s^* ds$ is j 's market share of total profits. By Proposition B3, we have $\frac{s_j^*}{s_k^*} = \frac{\hat{\pi}_j^* V_j(R)}{\hat{\pi}_k^* V_k(R)}$, and hence $\frac{s_j^*}{s_k^*} > \frac{\hat{\pi}_j^*}{\hat{\pi}_k^*}$ for any $j < k$, which implies that the dispersion of s_j^* must be a clockwise rotation of $\hat{\pi}_j^*$ (Hefti and Teichgräber, 2020). The reason is that a comparably higher quality allows a firm to afford a larger attention share *and* extract more money from the attention it attracts.

Second, the dispersions $\hat{\pi}_j^*$ and s_j^* can possibly rotate if R changes as depicted in Figure 8 (b), while R had no effect on these shares in the model of the main text. The reason is that the marginal attention costs per unit of revenue may now depend differentially on R via $V_j(\cdot)$.

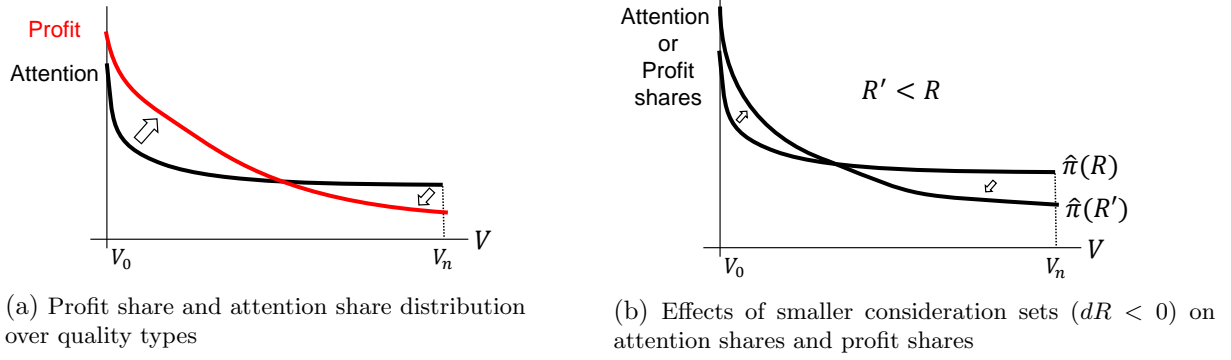


Figure 8: Heterogeneous quality: Attention and Profits

Consideration Set Size Effects We now explore the equilibrium consequences of an *exogenous* change to the consideration set size R . Understanding these effects matters for the general comparative statics, as R changes according to the finite capacity model if the information load changes. Consistent with assumption (13) from the main text, we suppose that consumer inattention is profitable, i.e., $-V'_j(R) > \frac{V_j(R)}{R}$, $\forall j$.

Expression (B.14) reveals that the market shares of attention, profits or revenues are invariant to R (similar to the main text) if and only if $\frac{V'_j(R)}{V_j(R)} = \frac{V'_k(R)}{V_k(R)}$, $\forall j, k$. The last condition further is equivalent to $V_j(R)$ being multiplicatively separable in (j, R) , i.e., to $V_j(R) = \alpha(j)g(R)$. Intuitively, this amounts to a situation where the average consumption expenditure shares per firm are independent of R . In this case, the equilibrium analysis and the comparative-static predictions coincide qualitatively with those from Section 3. In particular, the results in Propositions 3 - 5 apply here as well.

For R to have an effect on the equilibrium dispersions, it is necessary and sufficient that some agents are able to increase their attention revenues $V_j(R)$ by a greater proportion than others if R decreases. We exemplify this in the following proposition by assuming $\left| \frac{V'_j(R)}{V_j(R)} \right| > \left| \frac{V'_k(R)}{V_k(R)} \right|$, $\forall j < k$, i.e., that the revenues of those firms that better manage to monetize the attention they receive are also more exposed to a change in R .

Proposition B4 *If $\left| \frac{V'_j(R)}{V_j(R)} \right| > \left| \frac{V'_k(R)}{V_k(R)} \right|$, $\forall j < k$ and $R > 1$, then a decrease in the consideration set size ($dR < 0$) leads to a clockwise rotation of the attention and profit share functions. Moreover, $dR < 0$ induces a Superstar effect in attention, revenues and profits but never a Long tail effect.*

B.6.2 Market Size and Entry Effects

Proposition B4 further shows that attention and profit shares become less equally dispersed if R declines whenever firms with stronger abilities to monetize their attention also have revenues that react more volatile to changes in consumer R . Intuitively, if high-quality firms disproportionately benefit from consumer inattention, this allows them to disproportionately increase their attention expenditures, which cannot be matched by optimally behaving low-quality firms.

Compared to the equilibrium analysis of Section 4, this implies that increases in market size I or product diversity n can have additional effects through their impact on R . We now repeat the comparative-static analysis of Section 4, assuming that $\left| \frac{V'_j(R)}{V_j(R)} \right| > \left| \frac{V'_k(R)}{V_k(R)} \right|$, $\forall j < k$, such that a reduction in R causes a rotation as outlined by Proposition B4. Such a rotation means that there must be winners and losers among the incumbent forms in terms of market shares. In the following, we say that a firm j is a winner with respect to a measure W_j (e.g., profits, revenue or attention) if $dW_j \geq 0$, and a losers otherwise.

A growing Market Size Our first result considers the effects of consumer entry ($dI > 0$).

Proposition B5 *An increase in the market size ($dI > 0$) has the following effects.*

- (i) *For any $x \in (0, 1/2)$ there is a Superstar effect in attention, revenue or profits, but never a Long Tail effect.*
- (ii) *Regarding the distribution of attention, revenue or profits (each measure considered separately), the following patterns are possible: (a) all firms are winners, (b) all firms are losers, (c) there is a $j_0 \in \mathcal{J}$, such that all firms in $[0, j_0]$ are winners while all firms in $(j_0, n]$ are losers.*
- (iii) *The gap in attention, revenues or profits between any two winners and between any winner and any loser widens, while the gap between any two losers narrows.*
- (iv) *For a given $x \in (0, 1/2)$, there is a Matthew effect in attention, revenues or profits (for each separately) if either all firms are winners or if there is a $j_0 \in [xn, (1-x)n]$ such that all firms with $j \leq j_0$ are winners while all firms with $j > j_0$ are losers. If all firms are losers, then there is an inverse Matthew effect.*

A central difference to Section 4 is that $dI > 0$ now causes a Superstar effect. The reason is that the reduction of R due to choice overload now itself causes a Superstar effect. In addition, the fact that the attention, revenue and profit share distribution rotates implies that $dI > 0$ now

can cause winners and losers as measured by the absolute changes in the respective quantities. Should winners and losers coexist, then there must be a firm j_0 , such that all firms on its left must be winners, while all firms on its right must be losers.

Nevertheless, the essential insights are similar to Proposition 5: A Matthew effect in attention, revenues or profits arises whenever all firms end up as winners in the respective quantities.⁵⁸ Moreover, a Matthew effect results if winners and losers coexist. The only difference is that if all firms end up as losers, e.g., in profits, then $dI > 0$ causes an inverse Matthew in profits.

A growing Tail Compared to Section 4, the comparative-statics of a growing tail become more cumbersome, because n now also affects the revenue $V_j(R)$ via its effect on R . By (B.13), it follows that $\Sigma'(n) > 0$ and thus also $R'(n) < 0$ in a stable equilibrium, as in the main text. Our first result summarizes the effects of a growing tail on the incumbent firms, similar to Proposition 4.

Proposition B6 *A growing tail ($dn > 0$) has the following effects on the incumbents:*

- (i) *Regarding the distribution of attention, revenue or profits (each separately), one of the following patterns must occur: (a) all incumbents are winners, (b) all incumbents are losers, (c) there is a $j_0 \in \mathcal{J}$, such that all incumbents in $[0, j_0]$ are winners while all incumbents in $(j_0, n]$ are losers.*
- (ii) *The gap in attention, revenues or profits between any two winners and between any winner and any loser widens, while the gap between any two losers narrows.*

Proposition B6 shows that all incumbents may be winners or losers due to a growing tail. Moreover, the gap between any two incumbents widens (narrows) if all incumbents are winners (losers). These results are identical to what Proposition 4 found. The only novel aspect is that now only some incumbents may win while others lose.

Finally, we argue that a growing tail also is conducive to a Superstar effect, but not to a Long Tail effect. Formally, this is more complicated than in the main text, because the type space now is a functional space, rather than an real interval, and the new entrants come with new functions $V_j(R)$, rather than with new costs coefficients. Abstaining from a full analysis, the following argument demonstrates that a Superstar (Long Tail) effect is likely to arise, in

⁵⁸The actual determination of the winners and losers is more intricate compared to Section 4.1 and depends on details of $V_j(\cdot)$. However, the following can be shown: A firm's revenue increases iff its profit increases, and a firm's revenue increases if its attention does not decrease (the converse needs not hold in general).

particular because now $dR < 0$ already yields a Superstar effect. For a similar reason, a Long Tail effect does not occur.

Let $x \in (0, 1)$, and note that $Z_T(n)$ is of the form

$$Z_T(n) = \frac{\int_0^{xn} G_j(R) dj}{\int_0^n G_j(R) dj}, \quad G_j(R) \equiv V_j(R)^m$$

in case of profits, attention or revenues. By definition, a Superstar effect emerges if $Z'_T(n) > 0$, i.e., if

$$\frac{G_{xn}(R) \int_0^n G_j(R) dj - G_n(R) \int_0^{xn} G_j(R) dj}{\int_0^n G_j(R) dj^2} + \frac{\partial Z_T(n)}{\partial R} R'(n) > 0 \quad (\text{B.17})$$

In a stable equilibrium, $R'(n) < 0$, and $\frac{\partial Z_T(n)}{\partial R} < 0$ by Proposition B4, which is the Superstar effect in R . Thus, (B.17) implies that

$$\frac{G_j(R)j}{\int_0^j G_s(R) ds} > \frac{G_n(R)n}{\int_0^n G_s(R) ds} \quad (\text{B.18})$$

is a sufficient condition for $Z'_T(n) > 0$. This condition simply requires that the elasticity of $H(j) \equiv \int_0^j G(s) ds$ is strictly decreasing $\forall j \in (0, n)$, which formally is exactly the same condition that was identified by Lemma A1 to assure a Superstar effect upon a growing tail. For the same reason, a Long Tail effect is not likely to occur. A Long Tail effect occurs if

$$Z_B(n) = \frac{\int_{(1-x)n}^n G_j(R) dj}{\int_0^n G_j(R) dj} = 1 - \frac{\int_0^{(1-x)n} G_j(R) dj}{\int_0^n G_j(R) dj}$$

satisfies $Z'_B(n) > 0$. If (B.18) holds, then the above analysis directly excludes the occurrence of a Long Tail effect for the same reason why a Superstar effect occurs.

B.6.3 Proofs

Proof Proposition B3 Proceed as in the proof of Proposition 2. ■

Proof Proposition B4 The attention share distribution $\hat{\pi}_j$ is determined by (B.14). Thus

$$\frac{\hat{\pi}_j(R)}{\hat{\pi}_k(R)} = \left(\frac{V_j(R)}{V_k(R)} \right)^m, \quad (\text{B.19})$$

where $m = \frac{1}{\eta-1} > 0$. If $\left| \frac{V'_j(R)}{V_j(R)} \right| > \left| \frac{V'_k(R)}{V_k(R)} \right| \forall j < k$, then $\frac{\hat{\pi}_j(R)}{\hat{\pi}_k(R)}$ is strictly decreasing in R (as $V'_j(R) < 0 \forall j$). This implies that for $R' < R$ the distribution $\hat{\pi}_j(R')$ is a clockwise rotation of $\hat{\pi}_j(R)$.⁵⁹ Let

⁵⁹See Hefti and Teichgräber (2020).

$s_j(R) \equiv \frac{\Pi_j(R)}{\int_{\mathcal{J}} \Pi_s(R) ds}$ denote the profit share distribution. Then

$$s_j(R) = \frac{V_j(R)^{\frac{\eta}{\eta-1}}}{\int_{\mathcal{J}} V_s(R)^{\frac{\eta}{\eta-1}}}$$

by (B.19) and (B.15). This shows that $s_j(R)/s_k(R)$ is of type (B.19), and the claim follows. The claim about the Superstar and Long Tail effects is more involved, and needs the following Lemma.

Lemma B2 *Let $\varphi : [a, b] \rightarrow \mathbb{R}_{++}$ and $\varphi' : [a, b] \rightarrow \mathbb{R}_+$ be two integrable functions with*

$$\frac{\varphi'(x)}{\varphi(x)} > \frac{\varphi'(y)}{\varphi(y)} \quad \Leftrightarrow \quad x < y \quad (\text{B.20})$$

Then for $x, j \in (a, b)$:

$$\frac{\int_j^x \varphi'(s) ds}{\int_j^x \varphi(s) ds} < \frac{\varphi'(j)}{\varphi(j)}, \quad x > j, \quad (\text{B.21})$$

$$\frac{\int_x^j \varphi'(s) ds}{\int_x^j \varphi(s) ds} > \frac{\varphi'(j)}{\varphi(j)}, \quad j > x. \quad (\text{B.22})$$

Proof We only prove (B.21); (B.22) is proven similarly. Let $x > j$. Then (B.20) implies that $\varphi'(j)\varphi(s) > \varphi'(s)\varphi(j)$ for any $s \in (j, x]$. Thus also $\varphi'(j) \int_j^x \varphi(s) ds > \varphi(j) \int_j^x \varphi'(s) ds$, which implies (B.21). \square

Let $x \in (0, 1/2)$ and define $j \equiv nx \in (0, n)$. We show that the distribution of attention $C_j = \pi_j I$ exhibits a Superstar effect for $dR < 0$. Then, by (14)

$$Z_T(R) = \frac{\int_0^j \hat{\pi}_s(R) R I ds}{\int_0^n \hat{\pi}_s(R) R I ds} = \frac{\int_0^j \hat{\pi}_s(R) ds}{\int_0^n \hat{\pi}_s(R) ds} = \frac{\int_0^j V_s(R)^m ds}{\int_0^n V_s(R)^m ds}. \quad (\text{B.23})$$

Define $G_s(R) \equiv V_s(R)^m$. A Superstar effect exists for $dR < 0$ if $Z'_T(R) < 0$, hence if

$$\frac{\int_0^j -G'_s(R) ds}{\int_0^j G_s(R) ds} > \frac{\int_0^n -G'_s(R) ds}{\int_0^n G_s(R) ds}. \quad (\text{B.24})$$

Because for any $c < d$ we have $\frac{-V'_c(R)}{V_c(R)} > \frac{-V'_d(R)}{V_d(R)}$ by presumption, it follows that

$$\frac{-G'_c(R)}{G_c(R)} = \frac{-mV'_c(R)}{V_c(R)} > \frac{-mV'_d(R)}{V_d(R)} = \frac{-G'_d(R)}{G_d(R)}.$$

Then Lemma B2 assures that

$$\frac{\int_0^j -G'_s(R) ds}{\int_0^j G_s(R) ds} > \frac{-G'_j(R)}{G_j(R)} > \frac{\int_j^n -G'_s(R) ds}{\int_j^n G_s(R) ds} \quad (\text{B.25})$$

The claim now directly follows from the fact that

$$\frac{a}{b} > (<) \frac{c}{d} \quad \Leftrightarrow \quad \frac{a}{b} > (<) \frac{a+c}{b+d}. \quad (\text{B.26})$$

for positive numbers.

We now turn to the Long Tail effect, and again consider the distribution of attention. Fix $x \in (0, 1/2)$ and define $j \equiv (1-x)n$. By (14), a Long Tail effect fails to exist in consumer inattention if $Z'_B(R) > 0$, where

$$Z_B(R) = \frac{\int_j^n \hat{\pi}_s(R) R I ds}{\int_0^n \hat{\pi}_s(R) R I ds} = \frac{\int_j^n V_s(R)^m ds}{\int_0^n V_s(R)^m ds}. \quad (\text{B.27})$$

Define $G_s(R) \equiv V_s(R)^m$. Then $Z'_B(R) > 0$ if

$$\frac{\int_j^n -G'_s(R) ds}{\int_j^n G_s(R) ds} < \frac{\int_0^n -G'_s(R) ds}{\int_0^n G_s(R) ds} = \frac{\int_0^j -G'_s(R) ds + \int_j^n -G'_s(R) ds}{\int_0^j G_s(R) ds + \int_j^n G_s(R) ds}. \quad (\text{B.28})$$

By (B.26), it suffices to show that

$$\frac{\int_j^n -G'_s(R) ds}{\int_j^n G_s(R) ds} < \frac{\int_0^j -G'_s(R) ds}{\int_0^j G_s(R) ds}$$

in order to verify (B.28). The previous inequality indeed holds by (B.25) as a consequence of the presumption on $V_j(\cdot)$.

The proof is completed by noting that $Z_T(R)$ and $Z_B(R)$ are also of types (B.23) and (B.27), respectively, in case of revenues or profits. ■

Proof Proposition B5 We prove (i) in case of the attention distribution; the proof for revenues and profits is entirely similar. By (B.13), we must have $R'(I) < 0$ in any stable equilibrium. Let $x \in (0, 1)$ and $j \equiv xn$. A Superstar effect in I occurs if

$$Z_T(I) = \frac{\int_0^j \hat{\pi}_s(R(I)) R(I) I ds}{\int_0^n \hat{\pi}_s(R(I)) R(I) I ds} = \frac{\int_0^j V_s(R(I))^m ds}{\int_0^n V_s(R(I))^m ds}$$

verifies $Z'_T(I) > 0$. The claim now follows from Proposition B4 because I affects $Z_T(I)$ only via $R(I)$, and an increase in I causes a decrease of R . Likewise, Proposition B4 that a Long Tail effect cannot arise.

We now turn to claim (ii) of Proposition B5 in case of the attention distribution. By Proposition B4, $dR < 0$ leads to a clockwise rotation of the attention share distribution $\hat{\pi}_j$; thus there are winners and losers in terms of attention shares. As the total attention of firm j is $\hat{\pi}_j(R) R I$ the following situations can arise. (a) All firms lose attention. (b) There is $j_0 \in \mathcal{J}$ such that all firms with $j \in [0, j_0]$ win, while all firms with $j \in (j_0, n]$ lose. (c) All firms win attention. To see this, note that $\frac{C_j(R)}{C_k(R)} = \left(\frac{V_j(R)}{V_k(R)} \right)^m$ is strictly *increasing* in *inattention* ($dR < 0$) for any $j < k$. Now, if firm j is a winner, i.e., $dC_j(R) > 0$ as

$dR < 0$, then necessarily also $dC_{j'}(R) > 0$ for any $j' < j$. Likewise, if j is a loser, i.e., $dC_j(R) < 0$, then any firm with $j' > j$ must also be a loser. The same type of argument applies to revenues as well as to profits.

We next prove claim (iii). Consider two firms with $j < k$. If j is a winner and k a loser, the claim is trivial. Hence, suppose first that both firms are winners. Expression (B.19) shows that relative attention shares, $\frac{\hat{\pi}_j(R)}{\hat{\pi}_k(R)}$, $j < k$, is strictly increasing as $dR < 0$ as a consequence of the assumed property regarding $V_j(\cdot)$. Because both firms are winners, it must therefore be that firm j gains more attention than k ; the gap widens. By contrast, if both firms are losers, the fact that, again by (B.19), relative attention shares also increase, necessarily implies that the attention gap between j and k narrows down.

We finally prove claim (iv). Let $Q_j(I)$ denote the distribution of attention, i.e.,

$$Q_j(I) = C_j(I) = \hat{\pi}_j R I = \frac{V_j(R)^m}{\int_0^n V_s(R)^m} R I, \quad (\text{B.29})$$

where $m \equiv \frac{1}{\eta-1}$. A Matthew effect in attention exists if

$$M(I) \equiv \int_0^{xn} Q_j(I) dj - \int_{(1-x)n}^n Q_j(I) dj = \int_0^{xn} (Q_j(I) - Q_{j+(1-x)n}(I)) dj$$

verifies $M'(I) > 0$. If there is a $j_0 \in [xn, (1-x)n]$ such that all firms with $j \leq j_0$ are winners while all firms with $j > j_0$ are losers, then obviously $M'(I) > 0$. For the remaining part of claim (iv), note that the ratio $\frac{Q_j(I)}{Q_k(I)}$ is strictly increasing in I for any $j < k$ by the presumption of the proposition. Thus, if all firms are winners, (iii) implies that the gap between any two winners widens, which implies that also $M'(I) > 0$ in this case. If all firms are losers instead, then by (iii) the gap narrows, meaning that also $M'(I) < 0$; an inverse Matthew effect results. To complete the proof, suppose now that $Q_j(I)$ either denotes revenues ($Q_j(I) = \hat{\pi}_j R V_j(R) I$) or profits ($Q_j(I) = \hat{\pi}_j R V_j(R) I^{\frac{\eta-1}{\eta}}$). In both cases, the claims in (iv) follow from the previous arguments by noting that the ratio $\frac{Q_j(I)}{Q_k(I)}$ is strictly increasing in I for any $j < k$ also in these cases. ■

Proof Proposition B6 : (i) Consider the distribution of attention $C_j(n) = \hat{\pi}_j R I$ (the other cases are proven similarly). Note that for any two incumbents $j < k$, the ratio $\frac{C_j(n)}{C_k(n)} = \left(\frac{V_j(R(n))}{V_k(R(n))} \right)^m$ is strictly increasing in n because $R'(n) < 0$. Thus, if j is a winner ($C'_j(n) \geq 0$), then this must also hold for any $j' < j$. Likewise, the increasing ratio condition implies that if j is a loser ($C'_j(n) < 0$), then this must also hold for any $j' > j$. This implies that one of the cases (a), (b) or (c) must apply. (ii) All claims follow from the fact that the ratio $\frac{V_j(R(n))}{V_k(R(n))}$ is strictly increasing in n . ■

B.7 Price Competition

In this section, we provide a foundation for the value function $V^j(\cdot)$ based on strategic price competition with differentiated products. The goal is to show that conventional models of imperfect price competition yield the main properties we assumed in our analysis. In particular,

we vindicate that the value functions $V^j(\cdot)$ arising from such models verify the symmetry property $V^j(A) = V^j(A')$, $j \in A \cap A'$, whenever the distribution of preferences is unbiased towards any particular firm. As an intermediate step, we establish that a symmetric price equilibrium $p^* = p(R)$ always exists under standard regularity conditions, where this equilibrium is invariant to the allocation of attention, and all firms earn the same average rent $V(R)$ from the consumers paying attention to them. Finally, we show that assumption (13), which played a central role in our comparative-static analysis, is equivalent to the requirement that $p'(R) < 0$ in the price equilibrium. The latter inequality captures that a larger (perceived) market intensifies equilibrium price competition, which is a property common to conventional models of imperfect price competition.

B.7.1 Imperfect Price Competition

Consider a number of $n > R$ firms offering horizontally differentiated products. Let $p = (p_1, \dots, p_n) \in \mathbb{R}_{++}^n$ be the vector of product prices. All consumers are endowed with the same attention space $(\mathcal{A}, \mathcal{P}(\mathcal{A}), P)$.⁶⁰ Consumer i 's demand for firm j is specified by a random variable $d_i^j : \mathcal{A} \times \mathbb{R}_{++}^n \rightarrow \mathbb{R}_+$, where $d_i^j(A, p)$ is the quantity i demands from j if consideration set $A \in \mathcal{A}$ occurs and prices are given by p . Accordingly, the average demand for firm j from attention set $A \in \mathcal{A}$ is $d^j(A, p) = E[d_i^j(A, p)]$. If all firms face the same constant unit production cost $c \geq 0$, the value of attention then is $V^j(A, p) = (p_j - c)d^j(A, p)$. Recalling that $B_j \equiv \{A \in \mathcal{A} : j \in A\}$, we assume $d^j(A, \cdot) = 0$ whenever $A \notin B_j$, and that prices of non-considered products have no effects on a consumer's choice from A . In addition, we impose the symmetry assumption that the precise composition of the consideration set plays no role for the average demand $d^j(A, p)$ as long as the prices of the competing products in A remain the same.⁶¹

The above properties arise naturally in conventional settings, including discrete choice models or models with an additively separable utility function, if the underlying preferences are not systematically biased towards any particular firm in the consumer population.

To see this, consider first the case where consumer demand is the outcome of maximizing an additively separable utility function given a standard budget constraint and a fixed consideration

⁶⁰This assumption is only for ex simplicity. Our results also hold, e.g., if consumer demand d_i and attention P_i are uncorrelated.

⁶¹Formally, for any $A \in B_j$, let $p_A^j \in \mathbb{R}_{++}^{R-1}$ denote the subvector of prices of firms other than j that also belong to A . Then, $d^j(A, p) = d^j(A', q)$ for any $j, A, A' \in B_j$ and any price vectors p, q , where $p_j = q_j$, and p_A^j and $q_{A'}^j$ are permutations of each other.

set $A \in \mathcal{A}$. Consumer demand then is

$$d_i(A, p) = \arg \max_{x_j \geq 0} \sum_{j \in A} u_i(x_j) \quad s.t. \quad \sum_{j \in A} p_j x_j = w_i.$$

Conventional optimization methods reveal that a demand system with the above properties of $d^j(A, p)$ emerges if for any $A \in \mathcal{A}$ and any i , $u_i(\cdot)$ is strictly increasing and $\sum_{j \in A} u_i(x_j)$ is strictly quasi-concave. Further, it is easy to see that if all firms set the same price \bar{p} , then $d_i^j(A, \bar{p}) = \frac{w_i}{\bar{p}R}$. Accordingly, the aggregate attention rent of firm j is $V^j(A, \bar{p}) = \int V_i^j(A, \bar{p}) di = (\bar{p} - c) \frac{w}{\bar{p}R} I$, with $w \equiv \frac{1}{I} \int w_i di$.

Likewise, a demand system with the above properties arises in case of random utility models. The (indirect) utility of a consumer i of option j is specified by $U_i(j) = m - p_j + \varepsilon_j$, where the match values $(\varepsilon_1, \dots, \varepsilon_n)$ are distributed with a density function $f(e_1, \dots, e_n)$ over the consumer population of measure $I > 0$. Then, for a given $A \in B_j$, the market demand of firm j corresponds to $\Pr(U_i(j) = \max\{U_i(k) : k \in A\}) I$, the measure of consumers that find j to be their first-best choice within a given attention set. Let $f_A(e_1^A, \dots, e_R^A)$ denote the marginal density obtained from $f(\cdot)$ by integrating out those $n - R$ options that do not belong to A . Denoting the fraction of consumers who perceive firm j to be their first-best choice in consideration set $A \in B_j$ by $d^j(A, p) \equiv \Pr(U_i(j) = \max\{U_i(k) : k \in A\})$, we obtain

$$d^j(A, p) = \int_{-\infty}^{\infty} \int_{-\infty}^{p_2^A - p_j + e_j} \dots \int_{-\infty}^{p_R^A - p_j + e_j} f_A(e_2^A, \dots, e_j, \dots, e_R^A) de_R^A \dots de_2^A de_j, \quad (\text{B.30})$$

where p_2^A, \dots, p_R^A are the prices of the $R - 1$ options in A other than j . It is easy to verify from (B.30) that if $f(e_1, \dots, e_n) = f(e_{s(1)}, \dots, e_{s(n)})$ for any permutation s of $\{1, \dots, n\}$, capturing a symmetric dispersion of tastes in the consumer population, a demand system with the respective properties emerges, and the value function is $V^j(A, p) = (p - c)d^j(A, p)$. Further, we observe that if all firms set the same price \bar{p} , then $d^j(A, \bar{p}) = \frac{1}{R}$, such that $V^j(A, \bar{p}) = (\bar{p} - c) \frac{1}{R}$.

Applied work often considers the iid case, where $f(e_1, \dots, e_n) = g(e_1) \dots g(e_n)$ for a univariate density function $g(\cdot)$. If $U_i(j) = m - p_j + t\varepsilon_j$, where $t > 0$ is a “preference-intensity” parameter, all ε_j are iid uniform on $[0, 1]$, and $p_{-j} = \bar{p}$, (B.30) evaluates to

$$d^j(A, p) = \begin{cases} \frac{1 - \left(\frac{\bar{p} - p_j}{t}\right)^R}{R} + \frac{\bar{p} - p_j}{t}, & \frac{\bar{p} - p_j}{t} \geq 0 \\ \frac{\left(1 + \frac{\bar{p} - p_j}{t}\right)^R}{R}, & \frac{\bar{p} - p_j}{t} < 0 \end{cases}, \quad (\text{B.31})$$

for $p_j \in (\bar{p} - t, \bar{p} + t)$.

B.7.2 Equilibrium Pricing

We now analyze the pricing equilibrium for given attention imposing the above symmetry assumption on demand. Thus, the value function is of the form $V^j(A, p) = (p_j - c)d^j(A, p)$, where $V^j(A, p) = V^j(A', p)$ for any $A, A' \in B_j$ and equal prices $p_1 = \dots = p_n$.

As a first result we establish that a single symmetric price equilibrium $p_j^* = p(R) \forall j$ exists, and the equilibrium price is independent from attention $P(\cdot)$. That is, despite that firms may attract a substantially different amount of attention, and gain different profits, the Law of One Price applies. Thus, identical prices need not indicate a competitive outcome in the sense that all firms have the same attention shares.

Let $\Pi^j(p_j) = \sum_{A \in B_j} P(A)V^j(A, p)I$. We assume that no firm goes completely unnoticed: $\forall j \exists A \in B_j : P(A) > 0$ or, equivalently, $\pi_j > 0, \forall j = 1, \dots, n$. Each firm chooses its price p_j to maximize $\Pi^j(p_j)$, taking the other prices and $P(\cdot)$ as given. A price equilibrium is a price vector $p^* = (p_1^*, \dots, p_n^*) \in \mathbb{R}_{++}^n$ that is a solution to the first-order conditions

$$\sum_{A \in B_j} P(A) \frac{\partial V^j(A, p)}{\partial p_j} I = 0, \quad \forall j = 1, \dots, n. \quad (\text{B.32})$$

We claim that (B.32) has a single symmetric solution $p_1^* = \dots = p_n^*$.⁶² To derive any symmetric price equilibrium, it suffices to consider the maximization problem of an arbitrary firm j under the presumption that all other firms choose the same price $\bar{p} \in \mathbb{R}_{++}$, and evaluate its solution at $p_j = \bar{p}$ (Hefti, 2017). Consider firm j and any price vector of the form $p = (p_j, \bar{p})$ where all firm other than j choose the same price $\bar{p} > 0$. The symmetry in demand implies that $V^j(A, p) = V^j(A', p)$ for any such p , and we shall therefore use the notation $\bar{V}(p_j, \bar{p}, R) \equiv V^j(A, p) \forall A \in B_j$ and $p = (p_j, \bar{p})$. We impose the following regularity assumption on the function $\bar{V}(p_j, \bar{p}, R)$:

Assumption 1 $\bar{V}(p_j, \bar{p}, R)$ is twice continuously differentiable, and strictly quasi-concave in p_j . Further, the derivative function $\bar{V}_1(p, p, R) \equiv \frac{\partial \bar{V}(p, p, R)}{\partial p_j}$ satisfies

- (i) $\bar{V}_1(c, c, R) > 0$ and $\exists \hat{p} > c : \bar{V}_1(\hat{p}, \hat{p}, R) < 0$
- (ii) $\bar{V}_1(p, p, R) = 0 \Rightarrow \frac{\partial \bar{V}_1(p, p, R)}{\partial p} < 0, \bar{V}(p, p, R) > 0$
- (iii) $\bar{V}_1(p, p, R) = 0 \Rightarrow \frac{\partial \bar{V}_1(p, p, R)}{\partial R} < 0$.

These assumptions amount to standard Inada and monotonicity conditions that would assure equilibrium existence and uniqueness in conventional models of price competition with a full

⁶²We do not consider the possible existence of asymmetric equilibria.

consideration set of size n . Common examples, such as CES-demand or the above iid random utility model, satisfy Assumption 1.

As the following result shows, Assumption 1 implies the existence of a single symmetric price equilibrium $p^* = p(R)$, where p^* is independent of the allocation of attention $P(\cdot)$. With respect to our main analysis this means that, in the symmetric price equilibrium $p(R)$, the attention rent of a firm j is $\pi_j \bar{V}(p(R), p(R), R)I \equiv \pi_j V(R)I$, and thus of the type that we assumed throughout the main text.⁶³

Proposition B7 *Under Assumption 1, a single symmetric price equilibrium with $p_j = p(R) \in (c, \infty), \forall j$, exists. Further, $p(R)$ satisfies $p'(R) < 0$ and is invariant to $P(\cdot)$.*

The proof is presented at the end of this section. The main intuition is that the firms price to the market as perceived by consumers. While the various consideration sets may have very different probabilities of occurring, reflecting the possible heterogeneity in $P(\cdot)$, the fact that tastes are unbiased towards any particular firm implies that tastes are also unbiased across all consideration sets. As a consequence, a symmetric price equilibrium exists.

Finally, we show that condition (13) applies in the current setting. Both the random utility and the additive utility framework imply that the equilibrium attention rents are of the simple form $V(R) = \frac{v(p(R))}{R}$, where $v(p) = (p - c)$ in the former and $v(p) = \frac{(p-c)w}{p}$ in the latter case. It follows that $v'(p) > 0$ in both cases, capturing that a higher equilibrium price assures a higher equilibrium revenue of attention from those consumers to whom the firm manages to sell. It is now easy to see that condition (13) holds if and only if the equilibrium price satisfies $p'(R) < 0$, where the latter condition is assured to hold in any symmetric equilibrium by Proposition B7.

Corollary B2 *Let $V(R) = \frac{v(p(R))}{R}$ and $v'(p) > 0$. Then $-V'(R) > \frac{V(R)}{R}$ holds in any symmetric price equilibrium iff $p'(R) < 0$.*

Proof: The claim follows from $V'(R) = R = \frac{v'(p(R))p'(R)R - v(p(R))}{R^2}$. ■

Proof Proposition B7 Note that any interior, symmetric price equilibrium, by (B.32), is characterized by $\pi_j \bar{V}_1^j(p, p, R)I = 0$ or, equivalently, by $\bar{V}_1^j(p, p, R) = 0$. The strong quasi-concavity of $\bar{V}^j(p_j, \bar{p}, R)$ assures that any solution $p_j = p^*$ to this equation is the unique, global maximizer of $\Pi^j(p_j)$ given that $\bar{p} = p^*$. Further, (i) are (Inada) conditions, assuring that $V_1(p, p, R)$, as a function of p , crosses zero at least once, and (ii) assures that such a crossing is unique, and occurs for a price p^* that is strictly profitable for each firm. Finally, (iii) assures that $p'(R) < 0$ in the symmetric price equilibrium. To

⁶³For example, for (B.31) we obtain $p(R) = c + \frac{t}{R}$ in equilibrium, which yields $V(R) = \frac{t}{R^2}$.

see this, denote $\bar{V}_{1p}^j(p, p, R) \equiv \frac{\partial \bar{V}_1^j(p, p, R)}{\partial p}$ and $\bar{V}_{1z}^j(p, p, R) \equiv \frac{\partial \bar{V}_1^j(p, p, R)}{\partial R}$. Then, Assumption (ii) and the Implicit Function Theorem assure that $p'(R) = -\frac{\bar{V}_{1R}^j(p, p, R)}{\bar{V}_{1p}^j(p, p, R)} < 0$. ■