The Effect of Links and Excerpts on Internet News Consumption

Jason M.T. Roos* Carl F. Mela† Ron Shachar‡

September 24, 2015

Abstract

Does an Internet news site that excerpts and links to its competition steal their traffic? Or does excerpting increase the linked sites' audience? We develop a structural model to address this question. We show theoretically that the excerpted sites may either benefit (because consumers learn the linked content is suited to their preferences) or be harmed (because excerpting makes the linking site so attractive it steals traffic from the sites it links to). Using data from celebrity news sites, we measure the impact of excerpting on consumers' browsing choices, and find the former effect is dominant—that links are beneficial to both the linking and linked sites, as well as consumers.

Keywords: Structural models, Learning models, Dynamic programming, Bayesian estimation

*Rotterdam School of Management and ERIM, Erasmus University, P.O. Box 1738, 3000 DR Rotterdam, Netherlands; email: roos@rsm.nl; phone: +1 206 317 1713, +31 10 40 82527.

†Fuqua School of Business, Duke University, 100 Fuqua Drive, Durham, North Carolina, 27708; email: mela@duke.edu; phone: +1 919 660 7767.

‡Arison School of Business, Interdisciplinary Center (IDC) Herzliya, Herzliya 46150, Israel; email: ronshachar@idc.ac.il; phone +972 09 960 2408.

The authors would like to thank comScore for the data used in this study as well as Peter Arcidiacono, Andrew Ching, Andres Musalem, Ken Wilbur, Marshall Van Alstyne; and seminar participants at Duke (Fuqua, Dept. of Economics), Erasmus (RSM, ESE), Ohio State, Yale, Wash. U. in St. Louis, Georgia Tech, the 34th ISMS Marketing Science Conference, Tilburg University, UNC Chapel Hill, U. of San Diego, U. of Michigan, 2012 HEC Marketing Camp, U. of Frankfurt, U. of Houston, U. of Pennsylvania, and the 11th ZEW Conference on the Economics of ICT for their thoughtful comments. This paper stems from the first author's dissertation.

1 Introduction

Consuming news is an old habit, but in the last two decades it has experienced a fundamental shift as readers migrate to the Internet. Circulation and ad revenues from print news have been steadily falling for more than a decade, while digital-only news organizations, such as Huffington Post and Vice Media, have grown increasingly prominent (Pew Research Center 2014). A key distinguishing characteristic of this digital medium is the the ability to excerpt content from and hyperlink to other online news sources. Prior to the advent of the commercial Internet, it would have been unthinkable for a (print) newspaper like *The New York Times* to publish a daily summary of news articles available in *The Wall Street Journal*, or for a television network like CBS to promote news programs airing simultaneously on ABC. And yet, thanks to the widespread practice of excerpting and linking, this type of behavior is commonplace in the world of Internet news.[1]

Excerpts and links play an important role in news consumption. Because excerpts provide information about the content available at other sites, which individuals would not otherwise observe, understanding excerpts' influence on consumers' decisions is central to understanding how the Internet has changed news consumption. In a news setting, the information contained in these excerpts can be especially valuable to consumers with limited time and attention. Indeed, the success of news aggregators, such as *Google News*, suggests that consumers find such information to be useful (Athey and Mobius 2012; George and Hogendorn 2013).

Less clear are the circumstances under and degree to which excerpts benefit or harm the sites involved. One perspective is that news aggregators harm the content producers they excerpt because readers who might otherwise visit the content producer's site visit the aggregator instead. That is, the excerpting site benefits by *stealing* traffic from the sites it links to. Two high profile cases involving *Google News* would seem to echo this concern. In separate contract disputes, the Associated Press (AP) and Agence-France Presse (AFP) each demanded that *Google News* pay copyright royalties whenever it excerpted their content (Chiou and Tucker 2015; Isbell 2010). Both disputes were settled with *Google News* agreeing to pay undisclosed royalties to the agencies. Likewise, Athey and Mobius (2012) show that the addition of local news content to *Google News* led consumers to grow more reliant on the aggregator as a starting point when browsing local news sites.

Concern about excerpting's potentially negative effects on news producers has led to legislation in Germany and Spain requiring news aggregators to pay royalties to the sites they

[1] Because excerpts are almost always accompanied by a hyperlink to the excerpted site, and because our empirical study relies on hyperlinks to indicate when excerpting has occurred, we use the terms "links" and "excerpts" interchangeably to refer to excerpts.

excerpt. The outcome of these efforts, however, suggests the opposite—that excerpting may have a *positive* effect on the linked site. In Germany, *Google News* refused to link to any site that didn't waive the right to collect copyright royalties. But confronted with substantial declines in traffic from *Google News*, many German news sites have since opted out of the legislation, and now allow aggregators to excerpt their content for free.[2] In Spain, where the legislation legally prevents news agencies from opting out, *Google News* and a number of other news aggregators have simply ceased operations altogether. The result has been deleterious for Spanish news sites: Traffic dropped by an average of 16% at a time when overall Internet use in Spain increased. Moreover, the biggest declines in traffic occurred among smaller, niche news publishers, raising concerns about the negative impact of this legislation on consumer welfare (Concha et al. 2015). The outcomes of these cases lend support to the perspective that excerpting—at least by large news aggregators—might be beneficial to news publishers.

These examples suggest two competing effects: On the one hand, excerpting may increase the number of visitors going to the excerpted sites. But on the other hand, by making it easier for consumers to find interesting news content, excerpting increases the attractiveness the aggregator, possibly to the point where it steals more traffic than it provides. It remains unclear which effect dominates: the potential increase in traffic from getting excerpted (complementarity), or the potential loss in traffic as the linking site grows more attractive (substitution).

We seek to understand the mechanisms behind these opposing forces, quantify their magnitude, and assess their impact on consumers, with a model that can better guide the policy decisions of news organizations and regulatory bodies. Specifically, we develop and estimate a structural model of forward-looking consumers who visit news sites while simultaneously learning about the content at other sites.

This model introduces a number of novelties relevant to the study of Internet news consumption, the most important of which is our treatment of excerpts as noisy signals of consumers' heterogeneous match with content at the linked sites. Because excerpts provide match signals, observing them can either increase *or decrease* the likelihood of subsequently visiting the linked site, depending on the signal's valence. This model feature stands in contrast with previous theoretical models of linking, which have assumed that the conditional probability of visiting a site never decreases after observing a link or excerpt (Mayzlin and Yoganarasimhan 2012; Dellarocas et al. 2013).

Our approach to modeling excerpts sheds new light on how linking influences the consumption of Internet news. Importantly, by providing information, excerpts improve the

[2] A. Becker, "German publishers vs. Google," *Deutsche Welle*, October 30, 2014, https://web.archive.org/web/20150814100613/http://www.dw.com/en/german-publishers-vs-google/a-18030444.

efficiency and effectiveness of news consumption. Consequently, they increase total news consumption. To be more precise, consider first the linked site: Even though each excerpt can have a positive or negative signaling effect on each consumer's probability of visiting the linked site, on average excerpting will benefit sites that are visited infrequently. This is due to a floor effect on the probability of visiting the excerpted site. Specifically, if the prior probability of visiting a site is already quite low, negative information cannot lower that probability by much, whereas positive information can increase it considerably. As a result, linked sites can benefit from higher traffic originating at the excerpting site. Next consider the linking site: Publishing excerpts makes it more attractive because excerpts provide readers with useful information. As such, forward-looking individuals may prefer to visit such sites early in their browsing sessions so they will have better information when choosing which sites to visit later. Importantly, the higher popularity of the linking site may attract visitors who would have otherwise started their sessions at the linked sites, causing it to steal traffic from the linked sites.

Although ours is a model of costly consumption with learning, it differs from standard models in the vein of Erdem and Keane (1996) in important ways. For example, in the standard setting for these models (e.g., consumer packaged goods), consuming a product (e.g., Danon yogurt) provides a signal to the consumer about the true quality of the chosen good (i.e., Danon yogurt). But in our setting, consuming the news at one site (e.g. *dailykos.com*) also signals the characteristics of the news published *at other sites* (e.g. *politico.com* and *drudgereport.com*).

Because we model consumers choosing which news sites to visit during a browsing session, this study is also related to previous work in marketing and economics that has modeled Internet browsing at both the aggregate (e.g. Danaher 2007; Park and Fader 2004) and individual (e.g., Johnson et al. 2004; Lee et al. 2003) levels. The most similar of these models is that of Goldfarb (2002), which also describes expected utility-maximizing individuals choosing which site to visit next, in consideration of past browsing decisions and any outbound links they may encounter. A key difference though is that in our model, forward-looking consumers may anticipate that excerpts will help them browse more efficiently later in their session, and thus prefer sites with many links early in their browsing session.

Although our model provides new theoretical insights about the effects of excerpting, the magnitude of these effects and whether the linked site is better or worse off remain empirical questions. To answer these, we fit our model to Internet panel data that describe browsing at five celebrity news sites, which we match with data describing content published at those sites. Because these sites format their news articles as blog posts, they provide an ideal envi-

ronment in which to study excerpting (the practice of excerpting and linking which is now common among news sites originated with blogs). Whereas other studies of excerpting have focused exclusively on popular news aggregators, we consider excerpting among sites that publish both original news content and curated links to other sites. These sites may be more representative of how excerpting works among typical news sites.

We estimate our model by combining two recent advances from the econometrics (Imai et al. 2009) and statistics (Girolami and Calderhead 2011) literatures, and our approach provides a template for more efficient Bayesian estimation of single-agent dynamic discrete choice models. The model estimates provide a view into how news sites differentiate from one another by providing niche content or a high news volume, and underscores the importance of links to consumers. In our setting, observing just one excerpt reduces consumers' uncertainty about their match with the excerpted site's content by about 33%.

To measure the overall impact of excerpting on site traffic, and more specifically, to assess the extent to which excerpting benefits or harms linked sites, we conduct counterfactual simulations in which we measure how browsing would have differed had some sites not linked to others (as observed). This procedure quantifies the impact of excerpting in terms of site traffic, consumers' propensity to browse each day, the variety of sites they visit, and other metrics. In one illustrative case, we estimate that eliminating excerpts between two sites in our sample would decrease traffic moving between these two sites between 3% and 5%, and lower their total traffic (and thus revenue and profit for these advertising-driven sites) between 1% and 2%. In other words, we find links to be generally beneficial to linked sites, but more so to the linking site.

This paper offers a number of new insights and contributions. By allowing excerpts to provide match signals, our model provides a theoretical rationale for why excerpting can be positive for the linked site under some conditions and negative under others, potentially explaining the different outcomes in the cases involving *Google News*. Our empirical results reinforce the theoretical findings, while providing specific measurements of the impact of excerpting on site traffic.

These findings lead to another meaningful and important conclusion—the practice of excerpting among news sites is beneficial to consumers. Excerpts increase the consumption of news and encourage consumers to visit a wider range of sites, and do so by improving consumers' choices. By providing information about content at other sites, excerpts help consumers decide whether to continue their browsing session, and if so, which site to visit next. In total, these results suggest that one reason audiences are increasingly switching to the Internet may be because excerpts provide a more efficient way to consume news.

The remainder of this paper is structured as follows. First we present our model of news consumption in the presence of excerpts, and discuss its theoretical implications. We then describe our data and present preliminary analysis indicating excerpts may signal either positive and negative match. After discussing issues pertaining to estimation, we present the structural parameter estimates. Finally, we describe the counterfactual procedure and present its results before concluding with the main insights from this study.

2 Theoretical Model

On the morning of January 20, 2015, the online edition of *The New York Times* featured an article about President Obama's upcoming State of the Union address.[3] The article opened, "With the American job market surging to life..." and highlighted the president's intention to use the prosperous economy to invest in initiatives such as making "community college free for many students." If after reading *The New York Times* one were to have visited *The Wall Street Journal's* site, one would have seen another story covering the president's address.[4] Although this article repeated much of the information already found in *The New York Times* story, it also afforded a different perspective. For example, *The Wall Street Journal* article focused mainly on proposed tax increases and pointed out many aspects missing from the *New York Times* article, such as imposing "capital-gains tax on many inherited assets."

This example illustrates two important aspects of news consumption: First, at any two outlets there might be both unique and redundant information. Second, even when covering the same topic, sites may still differ in their perspectives (i.e., editorial positions). The utility of the individual in our model is centered around these two dimensions: information, and the match between the individual's perspective and the news outlet's editorial position.

One important practice missing from the example above is that Internet news sites frequently excerpt from and link to other sites. For example, if *The New York Times* and *The Wall Street Journal* behaved more like blogs, one might have seen excerpts in *The New York Times* article indicating that *The Wall Street Journal* article focused heavily on the tax implications of the president's proposal. Furthermore, these excerpts would have been accompanied by a link to *The Wall Street Journal's* article. In this way, a reader of *The New York Times* concerned with tax policy might have been alerted to content of interest at *The Wall Street Journal*.

In the rest of this section we describe the two dimensions of consumer utility (i.e. infor-

[3] P. Baker, "In State of the Union Address, Obama Is to Move Past Hardship and Reset Goals," *The New York Times*, January 20, 2015, https://web.archive.org/web/20150120141512/http://www.nytimes.com/2015/01/20/us/politics/ready-to-move-past-hardship-obama-resets-goals.html.

[4] C.E. Lee, et al., "Obama Plan Reignites Tax Fight," *The Wall Street Journal*, January 20, 2015, https://web.archive.org/web/20150120111314/http://www.wsj.com/articles/obama-tax-plan-hits-bumps-1421713523.

mation and match), the impact of excerpts and links, and their theoretical implications for the consumption of news. We begin by detailing nomenclature and the basic structure of the consumer's utility function.

Every day, the consumer engages in a browsing session, which we index d. By a "browsing session," we refer to the process of sequentially visiting zero or more sites within a day (hence not visiting any sites is an option). During each browsing session, the consumer makes a series of decisions about which (if any) site to visit next. The steps in the browsing section are indexed $t = 1, \ldots, T_d$.[5] We index consumers with i, and the sites they may visit with j.

To simplify matters, we follow the literature on sequential browsing in an online setting (e.g., Kim et al. 2010) by assuming the consumer sees all available content at each site visited, and visits each site no more than once per session. Hence the consumer's choice set, which is initially $\mathcal{J}_{i,d,t}$, becomes $\mathcal{J}_{i,d,t+1} = \mathcal{J}_{i,d,t} \setminus j$ after visiting site j. The consumer's choice can be viewed as a decision of which previously unvisited site to read next in the current session. We denote by $a_{i,d,t}$ the index j of the action taken by consumer i at step t of browsing session d.

Viewing sites is costly in terms of time and effort. We denote this cost by $\gamma_i > 0$ and assume it is known to the consumer and constant over the duration of the browsing session.[6]

The periodic utility individual i gains by visiting site j at step t on day d is:

$$U_{i,j,d,t} = \mu_{i,j,d} + \beta_{i,j,d,t} - \gamma_i + \epsilon_{i,j,d,t} \tag{1}$$

where $\epsilon_{i,j,d,t}$ is an idiosyncratic shock particular to each site at each step of the browsing session. This shock is private information learned just prior to the decision at step t, but not observed by the researcher. Ending the session yields net utility of $\epsilon_{i,0,d,t}$. The random variables μ and β in Equation (1) represent the utility from the two dimensions of media consumption described above: editorial position and information. These two variables are described in the next two sections, as are the processes by which the consumer's beliefs about them update at each step of the browsing session. Because consumers in our model are forward-looking, we then present the value function characterizing their choice problem at each step before concluding with a discussion of the model's implications.

2.1 Match Utility, μ

Match utility, denoted by $\mu_{i,j,d}$ in Equation (1), arises from the match between the site's editorial position and the views of the consumer. For example, a liberal consumer might receive higher match utility from visiting a Democratic-leaning news blog, such as *dailykos.com*, than from a Republican-leaning one, such as *drudgereport.com*. As Internet news sites frequently update

[5] To facilitate exposition, we drop the d subscript and write t instead of d, t when doing so does not lead to ambiguity.

[6] Alternatively, one can view this as the opportunity cost of foregoing the outside alternative.

their content, the level of this match varies with each session. For example, it is possible that site j usually expresses left-wing positions, but on some issues (e.g., healthcare) it expresses middle-of-the-road beliefs. Accordingly, the hot topics of each day will influence the daily value of $\mu_{i,j,d}$.

We model the match utility consumer i receives from visiting site j in session d is a function of 1) the site's long-run editorial position, z_j, 2) a session-specific, idiosyncratic deviation from this average, $\nu_{j,d}$, and 3) the consumer's own perspective or position, v_i. Specifically,

$$\mu_{i,j,d} = \left(z_j + \nu_{j,d}\right) v_i \tag{2}$$

This formulation implies consumers prefer sites satisfying $\operatorname{sign} z_j = \operatorname{sign} v_i$. For instance, a politically conservative consumer with tastes $v_i = -1$ would prefer a conservative site with $z_j = -1$ over a liberal site with $z_j = 1$.

Based on a potentially long history of browsing, the consumer knows site j's long-run editorial position, z_j. But the consumer does not observe $\nu_{j,d}$ until *after* visiting site j on day d. We assume the daily deviations from the long-run position have the following distribution.

$$\nu_{j,d} \sim N\left(0, \tau_\nu^{-1}\right) \tag{3}$$

In the absence of excerpts from other sites, the consumer has no *ex ante* information about these daily deviations, although we assume τ_ν^{-1} is known from prior browsing.

2.1.1 Links and Excerpts

While visiting site ℓ the consumer receives a signal of $\nu_{j,d}$ if site ℓ excerpts and links to site j that day. For example, the excerpt might indicate that the editorial position of site j is more liberal or conservative than average. We assume these signals are noisy, but unbiased reflections of sites' true match positions.

$$s_{j,\ell,d} | \nu_{j,d} \sim N\left(z_j + \nu_{j,d}, \tau_s^{-1}\right) \tag{4}$$

The notation $s_{j,\ell,d}$ indicates that the signal s describing site j (the excerpted site) was observed while visiting site ℓ (the linking site) on day d. The amount of noise in signals, denoted τ_s^{-1}, is constant across sites and known to the consumer.

This setup highlights the informative role of excerpts in helping consumers learn whether the site's daily position is more or less congruent with their preferences. Importantly, links can signal *lower* than average match, making the consumer *less* likely to visit the linked site.

Finally, because sites excerpt from each each other with asymmetric frequencies, we denote by $\omega_{\ell,j}$ the probability that site ℓ excerpts from site j and allow $\omega_{\ell,j}$ and $\omega_{j,\ell}$ to differ. This linking strategy is common knowledge in the model, although consumers do not know a priori which links will appear each day—i.e., the ω's are known, but their realizations are not.

2.1.2 Beliefs About Match Utility

The resulting posterior belief about expected match utility on each day arises from a standard application of conjugate normal distributions in the Bayesian learning literature (West and Harrison 1999):

$$\mathbb{E}\left(\mu_{i,j,d}|I_{i,d,t}\right) = z_j v_i + \left(\frac{\tau_s n_{i,j,d,t}}{\tau_s n_{i,j,d,t} + \tau_v}\right)\left(\bar{s}_{i,j,d,t} - z_j\right) v_i \tag{5}$$

where at step t on day d,

- $n_{i,j,d,t}$ is a state variable indicating the number of sites excerpting site j that were visited prior to step t,
- $\bar{s}_{i,j,d,t}$ is a state variable indicating the average match position signaled by the excerpts, and
- $I_{i,d,t}$ is a state variable representing the information set of the individual at the t^{th} step of the browsing session (e.g., n and $\bar{s}$; a formal definition of $I_{i,d,t}$ is given in Section 2.3).

Expected match utility is thus a weighted average of the long-run match ($z_j v_i$) and the average match signaled by excerpts encountered prior to step t ($\bar{s}_{i,j,d,t} v_i$); with weights determined by the signaling precision of the excerpts (τ_s), the variability in match across days (τ_v^{-1}), and the number of previous linking sites visited ($n_{i,j,d,t}$). Notice that when the individual starts a new browsing session, $n_{i,j,d,t} = 0$ for every site j, and thus expected match utility is exactly equal to the long-run value ($z_j v_i$.) Hence Equation (5) illustrates the value of excerpts to the consumer—on average they shift expectations about match utility away from their long-run average, and toward their actual day-specific values.

2.2 Utility from Information, β

Owing to the coverage of the State of the Union address on January 20, 2015, a person interested in domestic U.S. politics would find that day's news more interesting than someone who cared mostly about foreign affairs. This example illustrates two characteristics of news information: 1) the intensity of news coverage varies day by day, and 2) the relevance of that coverage differs across individuals. In addition, the amount of news information available to consumers can vary across sites (e.g. *nytimes.com* publishes a higher volume of news content than *buffalonews.com*). Both characteristics (the amount of information and its relevance) affect readers' consumption.

Formalizing this idea, Appendix A describes the microfoundations of the informational dimension of utility, which we denote as β. While leaving the details to the appendix, it suffices at this juncture to indicate that β is distributed over $\Re^+$ as follows:

$$\beta_{i,j,d,t} \sim \mathscr{F}_d\left(I_{i,d,t}, \lambda_i, \alpha_j\right) \tag{6}$$

where 1) the parameter $\alpha_j \in (0,1)$ reflects the typical amount of information on site j such that sites with more extensive coverage have higher values of α_j; 2) the parameter λ_i represents consumer i's ex ante expected utility from news information (i.e., the average relevance of, or value placed on each unit of news content); and 3) the distribution $\mathscr{F}$ is indexed by d to reflect variation in news information over days (i.e. some days are richer with news than others). $I_{i,d,t}$, as mentioned before, denotes the consumer's information state, and includes the cumulative exposure to information prior to visiting the t^{th} site on day d. Because news information typically overlaps across sites (as in our State of the Union example), the information utility from visiting site j depends not only on the information available at that site, but also on whether some of that information was already seen at other sites. In this way, the benefit from visiting new sites decreases as the stock of previously unseen information is depleted.

Because the amount and relevance of a site's information on any given day is only learned after visiting that site, the consumer must form beliefs about each site's information. As sites publish overlapping information, readers can engage in a learning process about information at unvisited sites based on content they have already seen. For example, after visiting the *New York Times* on the morning of January 20th, a reader would have learned that there is a considerable amount of news concerning the State of the Union address, as well as the degree to which that day's news was personally relevant. The reader could therefore update beliefs about the likelihood of finding news of interest at other sites that day (e.g. at *The Wall Street Journal)*.

Appendix A.2 builds upon the microfoundation of Equation (6) to show how consumer i's ex ante beliefs about information utility update through a Bayesian learning process such that the expected information utility from site j at step t on day d is

$$\mathbb{E}\left(\beta_{i,j,d,t}|I_{i,d,t}\right) = \underbrace{\left[\left(\frac{\alpha_j}{1+A_{i,d,t}+\alpha_j}\right)\left(N-K_{i,d,t}\right)\right]}_{\text{(a) Expected number of new units of information at site } j} \times \underbrace{\left[\lambda_i + \left(\frac{K_{i,d,t}}{\kappa_0 + K_{i,d,t}}\right)\left(\overline{u}_{i,d,t}-\lambda_i\right)\right]}_{\text{(b) Expected utility from a unit of information}} \tag{7}$$

and where at step t on day d,

- $A_{i,d,t}$ is a state variable equal to the sum of α_j's for whichever sites have already been visited,
- $K_{i,d,t}$ is a state variable reflecting the number of units of novel information already accumulated,
- $\overline{u}_{i,d,t}$ is a state variable reflecting the cumulative average utility from each unit of information,
- N is a parameter representing the theoretical upper limit on the amount of information available each day, and

- κ_0 is a parameter representing the consumer's prior beliefs about the variability of interesting news topics each day.

Although a detailed characterization of Equation (7) is provided in Appendix A.2, the intuition is straightforward. Equation (7) factors the expected utility from information into two terms. The first term (7a) captures the expected amount of new (i.e., non-redundant) information at site j, whereas the second term (7b) represents the expected relevance of that information. More loosely, the first term can be thought of as "how much new information might still be available at site j,"and the second as "how relevant will that information be to me."

More formally, the term $\alpha_j/(1 + A_{i,d,t} + \alpha_j)$ in (7a) describes the probability that a previously unseen unit of news information will be at site j. Hence, when α_j is large relative to $A_{i,d,t}$, such as at the start of the session, the consumer expects to find a greater amount of novel information at site j (accordingly, sites with high α_j are more attractive). At the same time, accumulating information decreases $\left(N - K_{i,d,t}\right)$, the amount of novel information that might yet be seen, and this in turn decreases the expected amount of new information at site j. Thus, there are two mechanisms by which visiting sites and accumulating information leads to changes in expectations about the amount of new information remaining at other sites. Finally, as more information is accumulated by visiting sites, the expected utility provided by any novel information (7b) is shifted away from the consumer's prior belief, λ_i, moving closer to the average utility from the information that was already seen, $\overline{u}_{i,d,t}$.

2.3 Value Function

When consumers read a site's content, they not only gain current period utility, they also update their beliefs about μ and β at other sites. Forward-looking consumers anticipate this updating and therefore face the standard exploitation-exploration trade off when deciding which site to visit next. For example, a consumer might choose to visit a site with many excerpts from other sites, such as *Google News*, in the expectation that excerpts will increase (decrease) the chance of subsequently visiting a site with high (low) match. By choosing sites that are informative about other sites (i.e., those that contain excerpts or have more extensive news coverage), consumers can increase the value of the rest of their browsing session. Dropping the i and d subscripts for clarity, the following value function corresponds with the consumer's utility function and beliefs about μ and β:

$$V(I_t, \epsilon_t) = \max\left(\epsilon_{0,t},\ \max_{j \in \mathcal{J}_t \setminus 0}\left\{\mathbb{E}\left(\beta_{j,t} \mid I_t\right) + \mathbb{E}\left(\mu_j \mid I_t\right) - \gamma + \epsilon_{j,t} + \int V(I', \epsilon') f\left(I' \mid I_t, j\right) g\left(\epsilon'\right) dI' d\epsilon'\right\}\right) \quad (8)$$

$$I_t \equiv \left\{n_t, \bar{s}_t, K_t, h_t, \overline{u}_t\right\} \quad (9)$$

where at step t,

- $f\left(I' \mid I_t, j\right)$ is the distribution of the next information set given the current information set I_t and choice j,
- h_t indicates which sites were already visited (hence, $A_t \equiv \sum_j h_{j,t}\alpha_j$), and
- $g(\epsilon)$ is the distribution of ϵ.

Although the value function (8) is quite standard, a brief description of its specific elements is useful. The first two elements in the information set, n_t and $\bar{s}_t$, are the state variables involved in updating beliefs about match utility from Equation (5); the other three elements, K_t, h_t, and $\overline{u}_t$, are the state variables involved in updating beliefs about information utility from Equation (7). Some of these state variables evolve in a deterministic way, whereas others evolve stochastically (see Appendix B for a complete characterization of the state transitions). For example, h_{t+1} is always h_t with the addition of an indicator for site j—i.e. h evolves deterministically conditional on the choice of site j. In contrast, $\bar{s}_t$ (the average signal values for sites' match positions) evolves according to a stochastic process. An individual choosing site ℓ is uncertain about which sites it excerpts (if any), and instead knows only the probabilities (the $\omega_{\ell,j}$'s defined in Section 2.1.1). By visiting site ℓ, the reader learns the realization of this link probability: if ℓ links to j, then $n_{t,j} = n_{t-1,j} + 1$ and a new value of $\bar{s}_{t,j}$ is obtained; but if no link is observed, then $n_{t,j} = n_{t-1,j}$ and $\bar{s}_{t,j} = \bar{s}_{t-1,j}$.

The state transition function, f (detailed in Appendix B), reflects the effect of choosing site j on the rest of the browsing session. Such potential effects come from two sources, corresponding with the two components of utility (match and information). First, the excerpts at site j can improve the precision of an individual's prediction about match utility at sites yet to be visited. For this reason, sites with many outbound links, such as *Google News*, are especially attractive early in the browsing session when the information set of the individual is quite empty. But later in the session, excerpts are informative predominantly when they point to sites that previous sites have not already linked to.

Second, the choice of site j affects the rest of the browsing session via its influence on expectations about the utility from information. Sites with more information (higher α_j) allow users to better assess how much information is available across all sites on a given day (i.e., to learn whether it is a "big news day"). Hence, forward-looking consumers have an incentive to visit information-heavy sites first—not only for the utility of their information—but also because such sites allow consumers to update their beliefs about the utility from content at other sites more expediently.

Finally, we note that consumers do not discount the future value of browsing in our model. Choices within a single browsing session are all made on the same day, hence any discounting within a session should be negligible. Although consumers might discount the value of future

browsing sessions, the present discounted value of those future sessions is constant across all options, and therefore does not influence choices. Put another way, browsing decisions on one day do not affect decisions on future days.

To summarize, several aspects of the consumer problem underpin the order and number of news sites visited in a session:

1. **Match:** individuals find sites whose editorial positions are congruent with their preferences to be more appealing.
2. **Links:** sites with many outbound links provide value because excerpts improve consumers' choices in the rest of the session.
3. **Information:** individuals' utility is higher when visiting sites with greater amounts of news information; moreover, by visiting such sites, consumers learn about the prospect of finding additional information over the remainder of the session.

These conclusions have implications not only for the overall attractiveness of sites, but also for order in which they are visited. For example, the value of a site with many excerpts or extensive news coverage (high ω_j or α_j) is greatest earlier in the session, whereas a site with high average match utility but no outbound links and little news information is equally attractive at every stage of the session.

2.4 Implications

In this section, we discuss the implications of excerpting for consumer behavior by way of a stylized example, in which a politically liberal consumer is limited to visiting up to two political news sites each day. The results we report are based on numerical simulations, and further details and additional results can be found in Section F of the Online Appendix.

In this example, one of the two sites (site L) regularly links to the other (site R), but the reverse never happens (and recall that according to our model, site L's excerpts signal higher than average match at site R 50% of the time). To further simplify the discussion and isolate the effect of links, we assume that both sites provide the same average level of match utility (e.g., their coverage is equally liberal on average) and provide negligible amounts of information utility (i.e., $\beta = 0$). We further assume the consumer's browsing cost is high enough that each site has less than 50% chance of being visited each day. Even under this highly stylized setup, excerpting plays an important role in the consumer's choice of which site to visit next, and can be either beneficial or detrimental to the linked site. Below we highlight three key results from this analysis:

For sites that are visited infrequently, getting excerpted increases the number of visitors to the excerpting site who subsequently visit the linked site. If our politically liberal consumer

visits site L and sees an excerpt indicating site R's content is more liberal than usual, then the chance of visiting R might increase substantially (remember the baseline probability of visiting site R is already low). If the excerpt signals R's content is more conservative than usual, then the chance of visiting site R might be lower, but not by much (it is already low to begin with). In the real world, most sites are not visited very often, hence there is generally a "floor effect" that allows excerpts to have a positive effect on the linked site, even though half of those excerpts will typically signal lower than average match.

The increase in traffic at the excerpted site (R) comes from visitors to the linking site (L) who, in the absence of seeing an excerpt, might have chosen instead to end their session. For this reason, excerpting increases the average number of sites consumers visit in each browsing session—that is, total media consumption is higher when sites excerpt. This theoretical result is consistent with the declines in local news traffic experienced in Germany and Spain after aggregators were prohibited from excerpting for free.

Because excerpts allow forward-looking consumers to browse more efficiently, providing them increases a site's popularity at the start of the browsing session. Consider the politically liberal consumer's beliefs before visiting any sites. A visit to site L will reveal a signal about R's content. If the excerpt signals site R is more liberal then usual, then the consumer can benefit by subsequently visiting R; if instead it signals R's content is more conservative than usual, then the consumer can benefit by avoiding R and ending the session. Hence, even though both sites provide the same amount of match utility in expectation (by assumption in this example), starting the session at site L leads to higher total expected utility from the entire browsing session.

The increased attractiveness of site L has two effects on browsing. First, it expands the number of consumers who browse, because those who might otherwise abstain from browsing now have a reason to visit site L instead (complementarity). This effect may explain empirical results reported in Athey and Mobius (2012) and George and Hogendorn (2013), whereby people who typically started their sessions at *Google News* browsed more often after the site expanded its news coverage. Second, the increased attractiveness of site L causes it to steal some traffic from R, because those who might otherwise start their sessions at R now have reason to start at L instead (substitution). This is the typical claim made by those seeking to curtail or monetize excerpting (e.g., in the AP and AFP contract disputes with *Google News*).

The overall effect of excerpting on traffic at excerpted sites can be positive or negative. The preceding discussion implies there are two ways excerpting can increase traffic at the excerpted site: First, excerpting can increase the flow of traffic from the linking site to the excerpted site. Second, it can increase the popularity of the linking site, which, by expanding

the total number of consumers who browse each day, further amplifies the flow of traffic to the excerpted site. There is a countervailing effect, however, which can lead to an overall decrease in traffic at the excerpted site: If excerpting makes the linking site popular enough, then it may end up stealing more traffic from the sites it links to than it provides.

In light of these results, it is evident that the impact of excerpting on linked sites and consumers is an empirical question that depends on a variety of factors, including: 1) the linking frequency among sites, 2) the informativeness of match signals, 3) the relative level of match utility provided by each site, and 4) the average frequency of visits to the sites. In an empirical setting, any of these forces may come to dominate. In other words, the overall effects of linking is a measurement issue ideally suited for a structural model.

3 Data

We estimate our model using data that describe reader browsing and content at five celebrity news sites between October 1, 2009 and December 31, 2009, a period of 92 days. We assemble these data from two sources: 1) comScore panel data describing consumers' browsing, and 2) links and content scraped from the sites via an automated web crawling procedure. We describe both of these before concluding with preliminary evidence that links can either encourage or discourage visits to linked sites.

3.1 Consumer Data

The browsing data were provided by comScore as part of a larger panel data set describing visits by 2.5 million U.S. consumer to more than 3,000 sites (all of which are members of the same blog-oriented advertising network). We focus on celebrity news sites in this study because 1) these sites cover a limited range of news items each day, and 2) they frequently excerpt from each other. We limit our attention to the five most visited celebrity news sites in our panel: *celebuzz.com*, *dlisted.com*, *egotastic.com*, *perezhilton.com*, and *thesuperficial.com*.

3.1.1 Sample Selection and Consumer Characteristics

Most panelists visit only a fraction of the total available sites. We therefore limit attention to the most active readers, which we define as anyone who 1) visited one or more of the 3,000 sites on at least 16 occasions in Q4 2009, 2) had at least 5 of those visits occur in each of the 3 calendar months, and 3) visited at least 2 of the 5 sites used for this study. Browsing and demographic data for the 127 consumers who fit this profile make up the estimation panel.

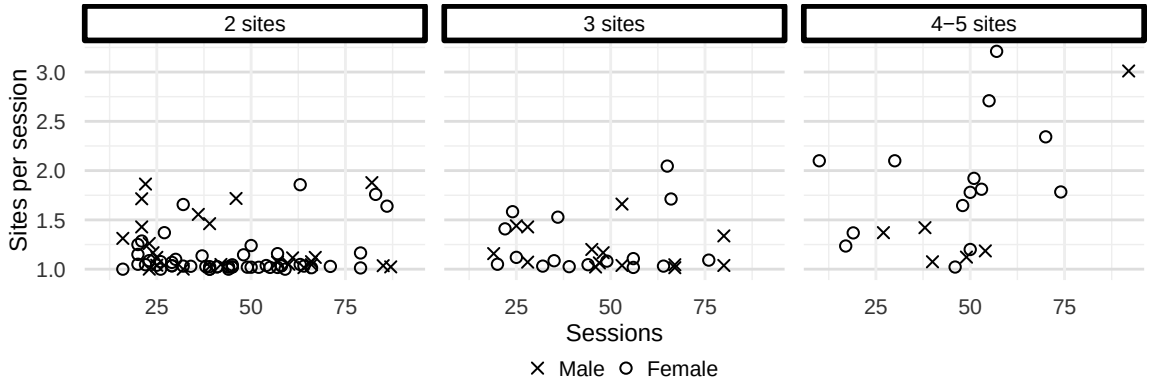
Most consumers in the estimation panel are female (65%). Most panelists (60%) are aged between between 25 and 55, with 35% younger and 5% older. Income is reported categorically, with a median in the range of $55–65k per year. Most panelists have children living with them

Table 1: Browsing Behavior by Site and Gender

	Visitors per Day			Step in Session		
Site	Male	Female	All	Male	Female	All
celebuzz	2.5	7.9	10.3	1.37	1.46	1.44
dlisted	3.4	9.0	12.4	1.42	1.28	1.32
egotastic	6.8	2.9	9.5	1.23	1.57	1.33
perezhilton	12.3	28.5	40.8	1.19	1.14	1.15
thesuperficial	4.6	3.6	8.0	1.38	1.94	1.62

NOTES: Based on 19,130 observed choices over the course of 5,757 browsing sessions. There are 127 consumers in the estimation panel (45 male and 82 female). "Visitors per Day" indicates the average number of male or female panelists visiting each site per day. "Step in Session" indicates the average time index t across visits, hence lower values indicate visits that occurred earlier in the browsing session.

Figure 1: Number of Sessions and Average Number of Sites Visited per Session, by Variety of Sites Visited and Gender



NOTES. Each panel shows a subset of consumers according to the variety of sites visited across all sessions in Q4 2009. For example, the left panel plots the number of sessions (x-axis) and average number of sites visited per session (y-axis) among the subset of consumers who (during Q4 2009) only ever visited 2 of the 5 sites included in our estimation data.

(57%), and the average household size is 2.7. Five panelists listed their race as African American. We code binary variables as {−.5, .5}, scale the 7 income categories between 0 and 1 using the center of the category range, and scale household size by subtracting the median (2) and dividing by two standard deviations (2.89). We denote by D_i the row vector of demographic variables for consumer i.

3.1.2 Browsing Data

As mentioned in Section 2, we define the length of a browsing session to be one day, since celebrity news sites operate under the same 24-hour news cycle as other media (Leskovec et al. 2009). For each panelist, we observe the order $t = 1, \ldots, T_d$ in which any of the five sites were visited each day (the choices $a_{i,d,t}$ in our model). During Q4 2009, the 127 panelists in our estimation sample made 19,130 such choices over the course of 5,757 browsing sessions.

Table 2: Link Frequencies (%)

	Link Target				
Linking Site	celebuzz	dlisted	egotastic	perezhilton	thesuperficial
celebuzz	-	6.5	0	1.1	9.8
dlisted	69.6	-	68.5	2.2	2.2
egotastic	0	65.2	-	0	0
perezhilton	7.6	0	0	-	0
thesuperficial	63	0	0	0	-

NOTES: Links were embedded in posts; we ignore "static" or "sidebar" links, as well as links to a site's own content.

Panelists varied considerably in the subset of sites visited, as well as the order in which they were typically visited. Table 1 shows that perezhilton was by far the most popular site among both male and female consumers, and was visited earliest on average. Preference for visiting the other sites differs by gender: male panelists with relatively higher preference for egotastic and thesuperficial, and female panelists with relatively higher preference for dlisted and celebuzz.

Figure 1 depicts the extent of variation in the number of days with browsing (x-axis), average number of sites visited per session (y-axis), and variety of sites visited (panels). Males comprise just 35% of the panel, but browsed more often than females (median male: 46 days averaging 1.12 sites; median female: 44.5 days averaging 1.05 sites). Figure 1 also shows that even at the lowest and highest extents of intensity, there is considerable heterogeneity in browsing habits. In total, the variability in browsing behavior suggests there are differences across individuals and browsing sessions that can be explained by our model.

3.2 Web Site Data

We created an automated web crawler to collect the full text from all news posts published at each of the five sites in Q4 2009. For each of those days, we use the text scraped from each site to determine 1) which other sites it linked to, and 2) how many words it published. We describe each of these next.

3.2.1 Link Data

Links that appear within the text of posts are typically accompanied by an excerpt from the linked site or a brief description of the linked content. Hence, even though we use the shorter term "link" to refer to both the link and excerpt, it is the excerpted content, and not the link per se, that signals consumers' match with the linked site (for this reason, we ignore so-called "sidebar" or "static" links that may appear as part of a site's navigation, but are never accompanied by an excerpt, and we do not consider so-called "around the web" display ads, as these

Table 3: Number of Words Published Each Day

Site	Min	25%	Median	Mean	75%	Max
celebuzz	0	1,140	1,923	1,912	2,873	4,076
dlisted	1,746	6,627	11,013	11,072	14,137	33,461
egotastic	0	0	463	604	727	2,872
perezhilton	0	2,113	4,906	4,482	6,336	9,002
thesuperficial	0	280	928	755	1,068	1,769

NOTES: Includes all words in the headline and body text of every post published on a given day.

were not used by the sites in our sample). We extract any links that appeared in the body of a news post. Then, using the sequence of sites visited by consumer i and the set of observed links between sites on each day d, we infer the number of match signals to each site that were seen at each step of the browsing session ($n_{i,j,d,t}$ from Equation (5)).

The frequencies with which sites linked to each other (ω in our model) are shown in Table 2. As many sites never linked to each other, half of the entries in Table 2 contain zeros. By contrast, dlisted and egotastic linked to each other about 67% of the time during Q4 2009. This variation, both within and across sites, allows us to measure the impact of links and excerpts on browsing.

3.2.2 Word Count Data

Recall that one of the dimensions of utility in our model comes from obtaining news information. As we discuss in Section 4.2, the daily word counts at each site (summarized in Table 3) provide an indirect measure of the amount of news information available to consumers on each day. To account for diminishing marginal information in the number of words published at each site we transform each site's daily word count: $w_{j,d} \propto \log\left(1 + \text{words}_{j,d}\right)$. We provide further details about the relationship between the state variable $K_{i,d,1}$ and the transformed word counts $w_{j,d}$ in Section 4.2.

3.3 Preliminary Analysis

Recall from our model that an excerpt can signal either higher or lower match, thereby increasing the likelihood of visiting the linked site. Because this aspect of our model runs counter to the standard assumption in the theoretical literature, wherein observing a link to another site never makes the reader less likely to visit the linked site (Dellarocas et al. 2013; Mayzlin and Yoganarasimhan 2012), we conduct preliminary analysis with the goal of understanding whether consumers in our estimation sample were more or less likely to visit linked sites. Accordingly, we conduct this analysis at the level of individual consumers. We define two empirical choice probabilities for each consumer i at each site j. The first is the probability

that consumer i visits site j after seeing one or more links to j:

$$\widehat{\Pr}_i\left(a = j | n_{i,j} > 0\right) = \frac{\sum_d \sum_t 1\left(a_{i,d,t} = j \text{ and } n_{i,j,d,t} > 0\right)}{\sum_d \sum_t 1\left(n_{i,j,d,t} > 0\right)} \tag{10}$$

The second is the probability consumer i visits j without previously seeing a link:

$$\widehat{\Pr}_i\left(a = j | n_{i,j} = 0\right) = \frac{\sum_d \sum_t 1\left(a_{i,d,t} = j \text{ and } n_{i,j,d,t} = 0\right)}{\sum_d \sum_t 1\left(n_{i,j,d,t} = 0\right)} \tag{11}$$

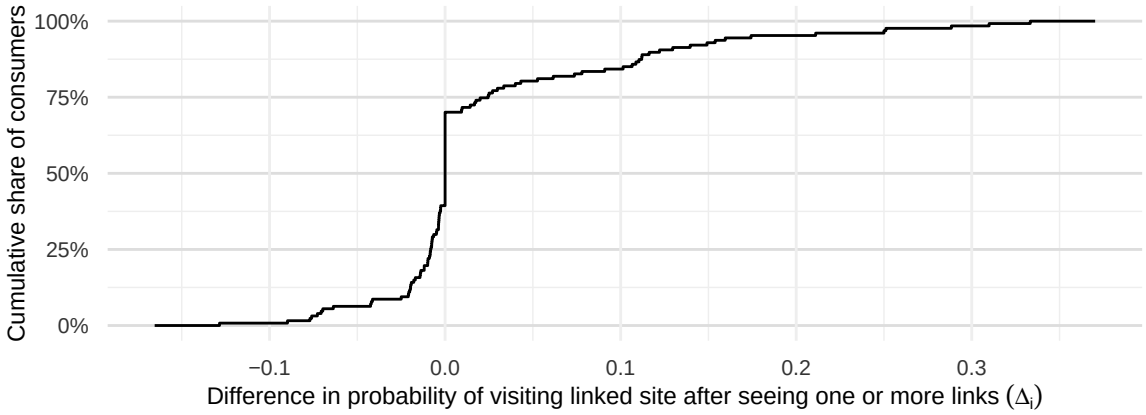
Next, we calculate for each consumer i the frequency-weighted average of each of these probabilities (i.e., averaging across all 5 sites). Thus $\widehat{\Pr}_i(a > 0 | n_a > 0)$ and $\widehat{\Pr}_i(a > 0 | n_a = 0)$ denote the probability consumer i visits *any* site a, given prior exposure to $n_a > 0$ links to that specific site. Finally, we calculate the difference between these two probabilities: $\Delta_i = \widehat{\Pr}_i(a > 0 | n_a > 0) - \widehat{\Pr}_i(a > 0 | n_a = 0)$. If links tend to encourage consumer i to visit (avoid) the linked site, then we expect $\Delta_i > 0$ ($\Delta_i < 0$); if they have no effect, then we expect $\Delta_i \approx 0$.

Figure 2 plots the empirical cumulative distribution of the difference between the two choice probabilities (Δ_i) across consumers. The left tail corresponds with the 39% of consumers who were less likely on average to visit the linked site after seeing links ($\Delta_i < 0$), the vertical line at 0 corresponds with the 31% showing no difference in their visit probabilities after seeing links ($\Delta_i = 0$), and the right tail corresponds with the remaining 30% who were more likely on average to visit sites after seeing links to them ($\Delta_i > 0$).[7] This evidence provides preliminary support for our modeling approach, whereby links can either increase or decrease traffic to the linked site. Specifically, although previous studies have not taken into account the possibility that links might discourage individuals from visiting the linked site, the data demonstrate that this is indeed possible, and may occur in quite meaningful numbers (i.e., almost 40% of consumers).

Although this result indicates that excerpting might be detrimental to the linked site (for at least some of its audience), recall that in Section 2.4 we explained how the average effect across all consumers might still be positive under these conditions (due to a floor effect when the probability of visiting the excerpted site is already low). We see evidence for this in Figure 2, as the magnitudes of increases in choice probability (the right tail) are greater than the magnitudes of decreases (the left tail). To confirm our intuition that excerpts might have an overall positive effect in this setting, we also calculate frequency-weighted averages of the probabilities in Equations (10) and (11) for each site (i.e., $\widehat{\Pr}\left(a = j | n_j > 0\right)$ and $\widehat{\Pr}\left(a = j | n_j = 0\right)$), and find that the average effects are indeed positive for all five sites (ranging from a 2.0%

[7] To verify the numerical robustness of this analysis, we repeat it for each subset of consumers who saw a total of at least ℓ links, for $\ell = 1, \ldots, 50$. The share of consumers with $\Delta_i < 0$ ranges between 22.6% and 40.3%, the share with $\Delta_i = 0$ ranges between 16% and 30.7%, and the share with $\Delta_i > 0$ ranges between 30.6% and 52.5%.

Figure 2: Effect of Observing Links on Consumers' Probability of Visiting the Linked Site



NOTES. The difference in probability (x-axis) describes a consumer's frequency-weighted average probability of visiting a site after seeing a link, minus the probability of visiting that same site in the absence of a link, denoted Δ_i in the text.

increase at perezhilton to 5.6% at dlisted).

4 Estimation

Here we discuss details related to our full empirical model, alternative specifications, model identification, and our MCMC sampling procedure.

4.1 Consumer Parameters

Consumers are heterogeneous with respect to their values of match preference (v_i), the utility they receive from each unit of news information (λ_i), and their browsing costs (γ_i). We model this heterogeneity using consumers' observed demographic variables (D_i) via the following prior distributions:

$$v_i \sim N\left(\eta_v + D_i\phi_v, \zeta_v^2\right), \qquad \log\lambda_i \sim N\left(\eta_\lambda + D_i\phi_\lambda, \zeta_\lambda^2\right), \qquad \log\gamma_i \sim N\left(\eta_\gamma + D_i\phi_\gamma, \zeta_\gamma^2\right) \tag{12}$$

Note that although these prior distributions assume independence among these parameters, this does not rule out any dependencies among their posterior distributions.

We anticipate the incentive to browse could be different on weekends and U.S. Federal holidays (Columbus, Veterans, Thanksgiving, and Christmas Days) due to differences in the value of consumers' time. The following specification allows consumers' γ_i's to differ systematically on these days:

$$\gamma_{i,d} = \begin{cases} \gamma_i \exp\left(\gamma_w\right) & \text{if } d \text{ is a weekend or holiday} \\ \gamma_i & \text{otherwise} \end{cases} \tag{13}$$

All else equal, a value of $\gamma_w > 0$ will lead to less browsing on weekends and holidays.

4.2 Word Counts

Although we do not observe sites' average information quantities (α_j) directly, we do observe a related quantity: the number of words published at each site each day ($w_{j,d}$). Hence, we assume sites that publish more information on average also publish more words each day. Below we provide an overview of our empirical approach to linking these two variables; technical details can be found in Appendix C.

Because consumer i obtains all available information from each site visited, the realization of the state variable $K_{i,d,1}$—the quantity of information obtained from the first site visited on day d—is also equal to the *total* quantity of information available from that site (note that this is not the case when visiting sites at later steps of the session). We therefore relate consumer i's realization of $K_{i,d,1}$ with the word count $w_{j,d}$ at the first site visited on day j.

More specifically, recall that our theoretical model stipulates that both prior beliefs and realized values of $K_{i,d,t}$ follow a common distribution that depends only on the α_j's, the previous value of $K_{i,d,t-1}$, and the set of sites already visited. In our empirical specification, we make one change: We assume that realizations of $K_{i,d,1}$ are drawn from a random distribution that is conditioned on the number of words published at the first site. This relationship *only* pertains to the realized values of $K_{i,d,1}$ (i.e., at step $t = 1$)—subsequent realizations of $K_{i,d,t}$ at steps $t > 1$, as well as consumer's expectations at every step of the browsing session, still depend on the α_j's exactly as formulated in the theoretical model. Because consumers have rational expectations in our model, this approach leads to estimates of the α_j's that partially rationalize the average number of words published at each site each day.

4.3 Model Likelihood and Bayesian Posterior Distribution

We now present the likelihood and posterior distribution of the model parameters. Following the literature on single agent, dynamic discrete choice models (Aguirregabiria and Mira 2010), we assume that the unobserved utility shocks ($\epsilon_{i,d,j,t}$'s) follow an i.i.d $EV(0,1)$ distribution. Accordingly, we can express the value of visiting site j, conditional on the state variables $I_{i,d,t}$, as $V_j\left(I_{i,d,t}\right) + \epsilon_{i,d,j,t}$, where the function $V_j\left(I_{i,d,t}\right)$ denotes the choice-specific value function:

$$V_j\left(I_{i,d,t}\right) = \underbrace{\mathbb{E}\left(\beta_{i,d,j,t}|I_{i,d,t}\right) + \mathbb{E}\left(\mu_{i,d,j}|I_{i,d,t}\right) - \gamma_{i,d}}_{\text{Expected period utility}} + \underbrace{\int \log \sum_{j' \in \mathcal{J}_{i,d,t} \setminus j} \exp\left[V_{j'}\left(I'\right)\right] f\left(I'|I_{i,d,t}, j'\right) dI'}_{\text{Emax function}} \tag{14}$$

The choice-specific value function comprises two parts: 1) the expected "period" utility from visiting site j at step t, and 2) the expected maximum utility from the remainder of the session, after visiting site j (the "emax" function). Integrating over the unobserved utility shocks (the $\epsilon_{i,d,j,t}$'s) leads to the conditional likelihood of the model parameters, θ, given the observed

Table 4: Estimated Parameters

Parameter	Dimension	Description
$\left(z_j, \alpha_j\right)$	5×2	Match location and information quantity for each site
$\left(\phi_v, \phi_\lambda, \phi_\gamma\right)$	7×3	Demographic coefficients for match preferences (v_i), information utility (λ_i) and browsing cost (γ_i)
$\left(\eta_\lambda, \eta_\gamma\right)$	1×2	Intercepts for information utility and browsing cost
$\left(\zeta_\lambda, \zeta_\gamma\right)$	1×2	Prior scales for information utility and browsing cost
γ^w	1×1	Incremental browsing cost on weekends and holidays
τ_s	1×1	Precision of link signals

NOTES: Parameters listed do not include those that are integrated out of the posterior distribution via data augmentation.

browsing choices, $a = \left\{a_{i,d,t}\right\}$, and the state variables, $I = \left\{I_{i,d,t}\right\}$:

$$L\left(\theta|a, I\right) \propto \prod_i \prod_d \prod_t^{T_{i,d}} \prod_{j \in \mathcal{J}_{i,d,t}} \left\{ \frac{\exp\left[V_j\left(I_{i,d,t}|\theta\right)\right]}{1 + \sum_{j' \in \mathcal{J}_{i,d,t}} \exp\left[V_{j'}\left(I_{i,d,t}|\theta\right)\right]} \right\}^{1\left(a_{i,d,t}=j\right)} \tag{15}$$

The likelihood function depends on the state variables (I) of which $\overline{u}$, K, and $\overline{s}$ are unobserved by the researcher. To obtain the marginal likelihood $L\left(\theta|a, n, w, h\right)$—where n indicates the observed links, w the word counts, and h the set of sites previously visited within the current session—we integrate over the distribution of the unobserved state variables $\overline{u}$, K, and $\overline{s}$. We use the standard Bayesian approach of data augmentation—treating $\overline{u}$, K, and $\overline{s}$ as latent parameters, estimating the joint distribution of (θ, I), and then numerically integrating over the unobserved states (Tanner and Wong 1987; Rossi et al. 2005). Further details are available in Appendix D, which also lists prior distributions for the remaining model parameters. Denoting this joint prior distribution $p\left(\theta|D\right)$, the full posterior distribution of θ is

$$p\left(\theta|a, n, h, w, D\right) \propto \int \left\{L\left(\theta|a, I\right) f\left(I|n, h, w\right) p\left(\theta|D\right)\right\} d\overline{s}\, d\overline{u}\, dK \tag{16}$$

As we cannot evaluate this distribution directly, we sample from it using MCMC, as explained in Section 4.6.

4.4 Identification and Parameter Normalizations

Identification of the model parameters is straightforward, as the 19,130 browsing decisions in our estimation sample contain a variety of moments that vary by individual, site, and day. First is a set of moments related to consumers' browsing decisions, including: 1) *Session frequency*—the number of days each consumer visited one or more sites; 2) *Session length*—the average number of sites she visited per session; 3) *Unconditional site visits*—the number of times she visited each site across all sessions; 4) *Conditional site visits*—the number of times she visited each site after an earlier visit to every other site; and 5) *Site order*—the average step within a session at which each site was visited. Second is a set of moments related to

sites' content decisions, including: 6) *Link frequency*—the number of links between each pair of sites; and 7) *Word counts*—the number of words published at each site. Third is a set of moments related to interactions between site content and individual browsing, such as the share of consumers who visit a site after seeing the same link, or end their browsing session after encountering the same number of words, etc. And fourth are the moments of the consumer demographic data.

Table 4 lists the 37 parameters that we estimate. Since the number of parameters we estimate is rather small, the large number of moments is more than enough for identification. Following is a brief discussion of the parameters and the moments that identify them.

Sites' average match locations, z_j, and the distribution of consumer's match preferences, ϕ_v, play similar roles to brand fixed- and consumer random-effects in traditional marketing choice models. Identification of these parameters is immediate given consumer heterogeneity in unconditional site visits and the demographic data. The parameter for the informativeness of excerpts as signals of sites' daily match locations, τ_s, is identified from the relationship between link data and individuals' behavior: If average conditional site visits from one site to another are systematically different on days when one site links to another, then the informativeness of links must be high.

As noted in Section 4.2, sites' average information quantities (α_j) are identified by 1) their average word counts, with sites that publish more words having higher values of α_j; and 2) consumers' browsing at later stages of their session—sites with high α_j lead to lower expected information utility at the next site, and therefore a greater chance of subsequently ending the browsing session. Accordingly, the parameters describing the distribution of preferences for news information (η_λ, ϕ_λ, and ζ_λ from Equation (12)) are identified by consumers' heterogeneous preferences for sites with higher or lower values of α_j. Specifically, the more a consumer is attracted to sites with high α_j, the greater the utility from news information. Hence, the average attraction for sites with high α_j identifies η_λ (the "constant" in the distribution of this utility), whereas heterogeneity in preference for sites with high α_j interacted with the demographic data identifies ϕ_λ (the "coefficients" of the demographics) and ζ_λ (the "variance").

In the same way, consumers' cost parameters (η_γ, ϕ_γ, and ζ_γ) are identified by the interaction of heterogeneity in session frequency and length with the demographic data. Specifically, consumers who browse more often and visit more sites per session will have lower browsing costs. And of course, if sessions are less frequent and/or shorter on weekends and holidays, then γ_w (which is the same for all individuals) will be higher.

Finally, in Appendix D, we discuss in detail the following parameter normalizations. First, we set $N = 30$ and $\kappa_0 = 4$ (related to news utility). Second, we set $\tau_\nu = 1$, $\sum_j z_j = 0$, $\eta_v = 0$, and

$\zeta_v = 1$ (related to match utility).

4.5 Alternative Models

To assess the effectiveness of link data in explaining consumer browsing, we estimate alternative specifications that restrict or remove entirely the link data. We compare the fit of these (and our full model) with the observed browsing, as described in Section 5 (and Appendix E). In one alternative model, consumers are not forward-looking, hence links only affect decisions *after* consumers see them (i.e., consumers do not seek out sites for their excerpts). In another, we ignore the link data altogether. To assess how word counts influence the α_j's, we also estimate a version of the full model (i.e., with links and forward-looking consumers) without the word count data.

4.6 MCMC Sampling Procedure

We conclude this section with a discussion of our estimation approach (further details can be found in the Online Appendix). We use the method of Imai, Jain, and Ching (2009, hereafter IJC) to sample from the data-augmented posterior distribution of the model parameters. The IJC procedure is based on a standard Metropolis-Hastings (M-H) sampler augmented with a method for calculating the emax function (Equation (14)). Compared to the standard nested fixed point algorithm for approximating the emax function (Aguirregabiria and Mira 2010), IJC's method requires significantly fewer computational resources (see Imai et al. 2009 and Ching et al. 2012 for further discussion of IJC's advantages).

But even though the computational gains from IJC are great, they come at a cost: The procedure can produce sample chains that are highly autocorrelated (compared to the same model without forward-looking consumers). To alleviate this autocorrelation, we use Girolami and Calderhead's (2011) MMALA procedure to construct high-quality proposal distributions for the M-H accept/reject steps in IJC. These proposal distributions have two important qualities: First, the deterministic component of the proposal distribution usually lies in the direction of higher density regions of the parameter space (relative to the current parameter vector). Second, the covariance of the random component is adjusted at each step to approximate the curvature of the posterior distribution. Together, these features greatly improve the rate of convergence and reduce autocorrelation.

To construct the MMALA proposal distribution, one must know the values of the first, second, and third partial derivatives of the target log-density function. For single-agent DDC's, these derivatives are not available in convenient closed forms, so we obtain their values through a technique known as automatic differentiation (also referred to as AD; Griewank et al. 1996; Su and Judd 2012). AD is a procedure for automatically augmenting computer code such that

while evaluating the value of an arbitrary function $f(x)$, the augmented program also evaluates $f'(x)$, $f''(x)$, etc. by algorithmically applying the chain rule corresponding with the basic operations (addition, multiplication, etc.) comprising the original function. The M-H proposal distributions we construct are based on derivatives of the model posterior distribution while ignoring IJC's numerical approximation to the emax function (in our case, the increased numerical efficiency from performing AD on IJC's approximation to the emax function does not offset the higher computational expense).[8]

5 Results

As noted in Section 4.5, we estimate three alternative models in addition to our full specification and compare their posterior predictive fit with the observed browsing. Specifically, we calculate and report in Appendix E the mean absolute percent error (MAPE) of four key summary statistics: 1) total traffic per site, 2) daily traffic per site, 3) daily traffic between each pair of sites, and 4) total site visits per consumer. As expected, the full model has the best overall fit with the data. On 3 out of 4 measures, the full model (with or without word count data) outperforms the alternatives which limit the use of link data. And the full model with word count data outperforms the version without (on 3 out of 4 measures).[9] In the remainder of the paper, we report results based on 15,000 posterior draws (after a burn-in of 5,000) sampled from the full model with the word count data.

Recall that sites in our model provide two types of utility to consumers, match and information. We first present the parameters related to match utility and the informativeness of links, then present the parameters related to information utility and browsing cost. We conclude with a discussion of how the parameter estimates yield insights for understanding differentiation among news sites.

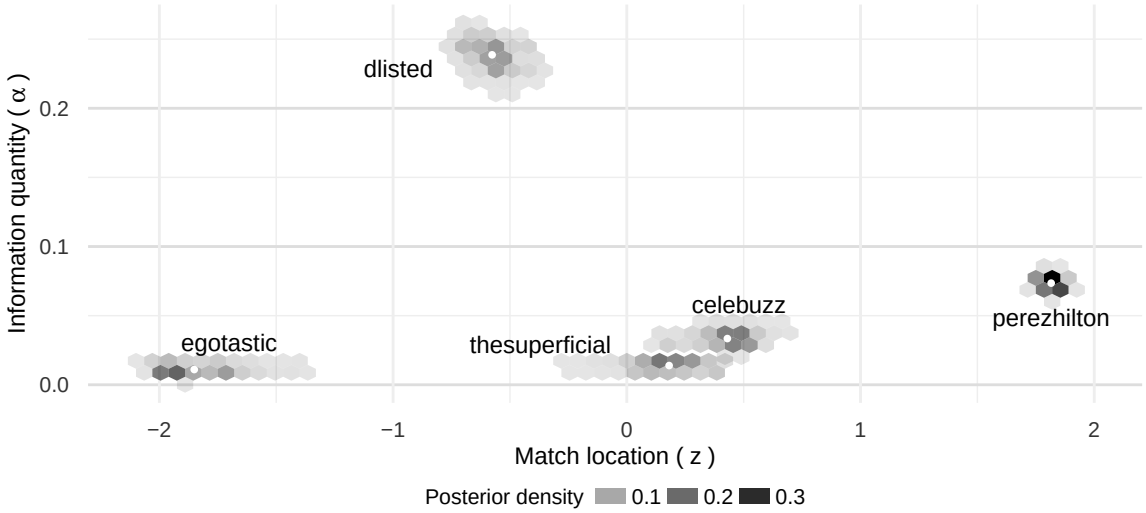
5.1 Match Utility and Link Informativeness

The average match utility consumer i receives from site j has two components, a site-specific match location, z_j, and a consumer-specific preference for this location, v_i (c.f. Equation (2)). Here we discuss estimates for both sets of parameters, before turning to the informativeness of links, τ_s.

[8] Markov chains sampled from the model with myopic consumers using: 1) MMALA proposals, and 2) random walk proposals (with the same target M-H acceptance rate) indicate that the MMALA chain has lag-1, -5, and -50 autocorrelations that are 19%, 36%, and 56% lower, and effective sample sizes that are 13 times higher. In other words, 1/13 of the draws are needed to obtain the same efficiency. Additional detail about the sampling algorithm is provided in Section G of the Online Appendix.

[9] The full model does a better job matching the moments related to aggregate browsing and traffic flows between sites, but a worse job matching the moments related to each individual's total browsing.

Figure 3: Joint Distribution of Sites' Information Quantities (α_j) and Match Locations (z_j)



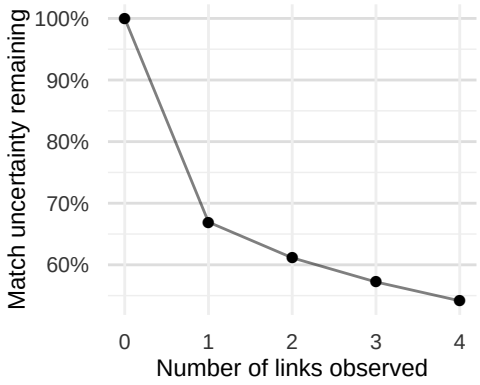
NOTES. White points indicate locations of posterior means.

Table 5: Consumer Heterogeneity Parameter Estimates

	Match Location (v)	Information ($\log \lambda$)	Cost ($\log \gamma$)
Observed factors			
Female	1.14*	0.81*	0.14
	(0.19)	(0.25)	(0.08)
Age<25	−0.13	−0.60*	0.00
	(0.19)	(0.25)	(0.08)
Age>55	0.27	−1.35*	−0.22
	(0.32)	(0.52)	(0.17)
Income	0.34	0.41	0.08
	(0.24)	(0.37)	(0.12)
Children	0.21	0.17	−0.06
	(0.24)	(0.33)	(0.11)
Household Size	−0.02	−0.25	0.02
	(0.24)	(0.31)	(0.11)
African American	−1.59*	0.91	0.40
	(0.37)	(0.57)	(0.21)
Intercept (η)	0.00	−2.36	1.52
	-	(0.46)	(0.15)
Unobserved factors			
Prior variance (ζ^2)	1.00	1.38	0.18
	-	(0.24)	(0.03)
Posterior variance	3.25	1.18	0.16
Total heterogeneity			
Posterior variance	3.80	1.49	0.17
Explained by observed factors	12.5%	21.0%	7.0%

NOTES: Estimates are posterior means with standard deviations in parentheses. For observed heterogeneity parameters, asterisks indicate estimates with 95% CI's excluding zero.

Figure 4: Links Reduce Uncertainty about Match Utility



NOTES. Match uncertainty remaining (y-axis) is the ratio of the posterior and prior variance of match utility after observing $n = 0, \ldots, 4$ links (x-axis), $\text{var}\left(\mu_{j,d}|n_{j,d} = 0, \ldots, 4\right)/\text{var}\left(\mu_{j,d}|n_{j,d} = 0\right)$, estimated as the posterior mean of $(\tau_s n + 1)^{-1}$.

Posterior densities for sites' average match locations (z_j) are depicted along the x-axis in Figure 3. The posterior distributions of match locations are negative for egotastic (−1.85) and dlisted (−0.58), close to zero for thesuperficial (0.18), and positive for celebuzz (0.43) and perezhilton (1.82). Qualitatively, this ordering of sites is consistent with the relatively high amount of sexually-oriented content (specifically, pictorials of attractive female entertainers and models) published by egotastic, dlisted, and thesuperficial; and, to a lesser extent, these sites' greater reliance on humor and sarcasm when reporting on celebrities. Although celebuzz and perezhilton publish sexually-oriented content, they do so less frequently and feature male celebrities much of the time. And although reporting at celebuzz and perezhilton includes humor and sarcasm, posts at these sites align more closely with traditional tabloid celebrity gossip compared to the other three. In light of these differences, we interpret the z_j's as points along a continuum ranging between content that is more "sexy" at one extreme ($z_j < 0$) and more "gossipy" at the other ($z_j > 0$).

Consumers' preferences for sites' match locations (v_i) are highly heterogeneous, as shown in column 1 of Table 5. This heterogeneity is partly explained by two demographic variables. The most important variable is gender: Males have v_i's that are on average negative, meaning they receive higher match utility on average from sites with $z_j < 0$ (i.e. the "sexy" content of egotastic and dlisted), and lower match utility from sites with $z_j > 0$ (the "gossipy" content of celebuzz and perezhilton). The other demographic variable is African American: these consumers receive higher match utility at egotastic and dlisted, although we note that this estimate reflects the preferences of just 5 panelists. Altogether, demographic variables account for 12.5% of the total heterogeneity in consumers' preferences for match location.

The true location of each site, and hence the actual amount of match utility received, devi-

ates each day, and links provide consumers with information about these daily deviations (c.f. Equation (4)). The amount of information contained in links and excerpts is reflected in the parameter τ_s, which formally represents the precision of link signals around sites' true match locations (relative to the daily variation in match location). The marginal posterior distribution of τ_s is highly right-skewed, with a posterior median of 0.10 and a mean of 217, indicating links are indeed informative in this setting. Furthermore, in Section 6, we demonstrate that this level of informativeness has a meaningful impact on browsing. Figure 4 further illustrates the informativeness of links by showing the reduction in uncertainty about a site's match utility after observing increasingly more links. Observing one link reduces uncertainty about match utility by about 33%; seeing a second link reduces uncertainty by another 6%. Overall, we find compelling evidence that links provide informative signals about match utility at other sites.

5.2 Information Utility

Just as with match utility, we also find site differentiation and heterogeneous preferences for information utility. The expected quantity of information for each site (α_j) is depicted along the y-axis in Figure 3, with dlisted estimated to provide the greatest and egotastic the least. These estimates reflect both consumers' browsing habits and differences in the number of words published at each site.[10]

Consumers are heterogeneous in their expected utility received from news information (λ_i), as shown in column 2 of Table 5. Demographic variables explain 21% of this heterogeneity, with female consumers and those aged 25–55 receiving the most utility from news information. In total, the demographic variables do a better job explaining preferences for news information than match location.

Sites with higher average amounts of news information (α_j) are most attractive when visited early in the browsing session, because they provide useful information about the intensity of news coverage each day. This is especially true for individuals who value news information the most. For this reason, we expect individuals with high values of λ_i (based on a median split) to prefer visiting sites with high α_j's at the start of their session. Consistent with this prediction, we find those with higher λ_i's visited dlisted 10 times as often at the start of their session than those with lower λ_i's. But after visiting two sites, both groups were equally likely to visit dlisted. This suggests that the α_j and λ_i parameters reflect preferences for news information that evolve over the course of a browsing session, as expected by the model.

[10]When we estimate the full model without the word count data, we find that estimates of α_j are similar for all sites except celebuzz, which is estimated to have a higher value of α_j compared to the model described here.

5.3 Browsing Cost

We turn next to the parameters for browsing costs. Column 3 of Table 5 shows that none of the demographic variables are significantly related to consumers' browsing costs (γ_i), and collectively explain just 7% of the variation in $\log \gamma_i$. As predicted by our theoretical model, consumers with the highest costs visit the fewest number of sites. Moreover, this group was most likely to visit egotastic and perezhilton (both sources of high match utility) at the start of their sessions. By contrast, consumers who initially visited celebuzz and thesuperficial tended to have the lowest costs, and visited more sites per session. Finally, as expected, the estimate for γ^w indicates browsing costs are about 8.8% (SD 1.4%) higher on weekends.

5.4 Site Differentiation and Consumer Benefits

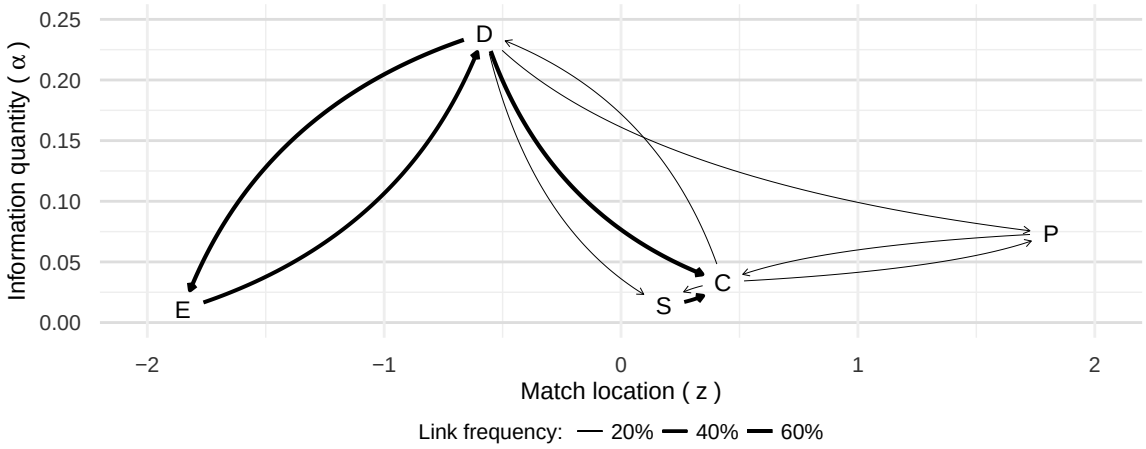
Sites are differentiated by their match locations (z_j), information quantities (α_j), and linking frequencies (ω_j). Figure 5 depicts these spatially, with match locations along the x-axis and information quantities along the y-axis; link frequencies are overlaid as arcs of varying widths. Here we comment briefly on how these three sources of differentiation affect consumers' valuations for these sites.

First, all consumers value the news information available at sites, but some place a higher value on it than others. Hence, by providing greater amounts of news information, dlisted achieves a degree of vertical differentiation from its competitors. Second, consumers' match preferences may be positive or negative, hence egotastic (with a negative z_j) and perezhilton (with a positive z_j) appeal to different audiences and, as such, are horizontally differentiated from each other (by contrast, thesuperficial, with a z_j close to 0 is relatively undifferentiated from either). Third, sites tend to link to competitors with similar values of z_j (i.e., their closest neighbors along the x-axis). Because links provide information about daily match locations, sites that frequently link to their closest competitors provide value by informing their audiences about sites with similar levels of match utility. If instead, excerpts tended to come from sites with very different match locations (e.g., if egotastic were to link to perezhilton), then consumers would find excerpts to be far less useful, even though the excerpt might be highly informative. As we show next via counterfactual experimentation, a significant portion of dlisted's value to consumers stems from its tendency to excerpt from a wide variety of sites.

6 Counterfactual Analysis

Does it ever make sense for a site to prevent competitors from linking to it, as when *Google News* was legally blocked from linking to sites in Spain and Germany? Although the theoretical implications of our model indicate that excerpting can be either beneficial or detrimental

Figure 5: Link Frequencies and Site Heterogeneity



NOTES. Link frequency indicates the empirical distribution of links as observed in the data. Sites are located at their posterior means. C = celebuzz; D = dlisted; E = egotastic; P = perezhilton; S = thesuperficial.

to news sites, these implications also indicate the importance of other factors determining the relative attractiveness of the sites involved. That is, the theoretical results by themselves do not provide an unambiguous answer to this question. Nor do the empirical results directly measure the impact of excerpts on browsing. Thus, to understanding how links affect behavior, we conduct counterfactual simulations in which we exogenously manipulate the linking behavior of particular sites and simulate the impact of these changes on browsing. In the remainder of this section, we describe this approach and discuss the various linking scenarios; we conclude with a general discussion of the value of excerpts in this setting.

6.1 Procedure

The objective of this analysis is to understand how excerpting affects the number of consumer browsing sessions, the flow of traffic between sites, and the number of visitors to each site. The empirical distribution of links between sites, listed in Table 2 and depicted in Figure 5, provides the baseline for these comparisons. Our counterfactual simulations entail removing subsets of these links. For each of the 5 sites (the "focal site" for the simulation), we evaluate 3 counterfactual scenarios. First, we consider what happens when the focal site unilaterally prevents all other sites from linking to it, as happened during the contract disputes between *Google News* and AP and AFP. Second, we consider what happens when the focal site unilaterally ceases linking to other sites, as in Germany and Spain. And third, to measure the total value of excerpting for each site, we consider what happens when all links to and from the

focal site are eliminated.[11]

Our general procedure is to simulate browsing S times for every consumer under each of 16 scenarios ($3 \times 5 = 15$ counterfactuals, plus the baseline), with each of the S simulations corresponding to a sample drawn from the data-augmented posterior distribution of the model parameters. We average the results from each scenario over the S simulations (i.e., integrate over the posterior distribution of the parameters). The counterfactual scenarios entail setting certain values of ω (link probability) and n (number of links observed) to zero; because consumers' expectations about the links they will encounter at certain sites depend on ω, we re-estimate the value function prior to each simulation. Because this is computationally expensive, we set $S = 200$. To account for any simulation error introduced by this decision, we calculate bootstrap confidence intervals for all estimates and focus attention on scenarios with measured effects that are reliably different from zero.

6.2 Results

Table 6 presents the main results of the counterfactual simulations. Each row describes one of 15 counterfactual scenarios as the percent change in three quantities: 1) total traffic at the focal site, 2) total traffic at the other four sites, and 3) number of browsing sessions initiated by consumers. Because sites link to each other with varying frequencies in the baseline scenario, the counterfactual's impact on browsing is greater in magnitude for some focal sites (e.g., dlisted, which excerpts a lot) than others (e.g., perezhilton, which does not). We focus on cases involving sites that excerpt most often in the discussion that follows.

6.2.1 Inbound Links

We first consider what happens when the focal site prevents other sites from linking to it (as in the AP and AFP cases). Recall that our theoretical results indicate excerpting may be positive or negative for the focal (excerpted) site. Table 6 shows that preventing inbound links has an insignificant effect on the focal site's traffic. For egotastic and dlisted, preventing inbound links leads to a small loss in traffic, whereas thesuperficial seems to benefit by preventing inbound links. None of these measured changes have bootstrap CI's that exclude zero, however,

[11] For several reasons, we do not model potentially endogenous responses to exogenous changes in sites' linking behaviors. First, we anticipate these indirect effects would be insubstantial, and thus would not change the qualitative nature of our findings. Second, a complete accounting of endogenous responses to changes in linking would entail solving a game played by 5 sites, each with 4 link probabilities to choose. Computing a Nash equilibrium over linking profiles would result in a system of up to 20 implicit functions that could only be solved with numerical methods, if at all (it is unclear whether such a solution exists, and if so, whether it is unique). Third, it is not clear that firms are actually playing such a game. Owing to the uncertain benefits from such an exercise weighed against the costs, we leave consideration of this linking game for future research. In this regard, one can interpret our findings in light of the demand side implications of an exogenous change in links, which would need to be solved prior to solving the endogenous linking game.

Table 6: Impact of 15 Counterfactual Simulations on Sites and Consumers

Counterfactual		Percent Change		
Type of Links Removed	Focal Site	Focal Site Traffic	Other Sites' Traffic	Browsing Sessions
Inbound	dlisted	−0.34	−0.25*	−0.26*
	perezhilton	0.03	−0.20	−0.03
	celebuzz	−0.06	−0.04	−0.10
	egotastic	−0.34[a]	0.02	−0.06
	thesuperficial	0.99	0.03	0.03
Outbound	dlisted	−0.50*	0.06	−0.11
	perezhilton	−0.02	−0.18	−0.08
	celebuzz	−0.06	−0.06	−0.09
	egotastic	−2.11*, [a]	0.05	−0.25*
	thesuperficial	−0.20	−0.04	−0.07
Both	dlisted	−0.76*	−0.26*	−0.36*
	perezhilton	0.00	0.22	0.04
	celebuzz	0.12	−0.01	−0.02
	egotastic	−2.00*, [a]	−0.18	−0.37*
	thesuperficial	0.02	−0.05	−0.13

NOTES: In each simulation, the focal site prevents other sites from linking to it (inbound), stops linking to other sites (outbound), or both. Percent changes in traffic refer to the changes in total visitors at either the focal site, or the other four sites combined. Browsing sessions indicates the total number consumer/day combinations visiting one or more sites. * indicates 95% bootstrap CI around the estimate excludes 0. [a] indicates results also reported in Table 7.

Table 7: Impact of Counterfactuals Involving dlisted and egotastic

Outcome		Percent change in outcome when...		
		dlisted stops linking to egotastic	egotastic stops linking to dlisted	neither site links to the other
Direct traffic from...	dlisted to egotastic	−3.20*	0.43	−3.03*
	egotastic to dlisted	1.35	−5.54*	−5.42*
Sessions starting at...	dlisted	−0.67*	0.00	−0.80*
	egotastic	−0.17	−2.53*	−2.26*
Total traffic at...	dlisted	−0.47*	−0.07	−0.78*
	egotastic	−0.34[a]	−2.11*, [a]	−2.00*, [a]

NOTES: Direct traffic from X to Y is the number of times consumers visited site Y immediately after site X with no intervening visits. Sessions starting at X is the number of times consumers visited X at step $t = 1$. * indicates 95% bootstrap CI around the estimate excludes 0. [a] indicates results also reported in Table 6.

and the average total effect on the excerpted site appears to be neutral.

Turning next to the effect of dropping inbound links on the other four sites, we again see little change in traffic and number of sessions, with the exception of when dlisted prohibits inbound links (traffic at the other four sites and the number of browsing sessions both decrease by about 0.25%). As indicated in Section 6.2.3, these decreases are almost entirely due to a large (2%) decline in traffic at egotastic. However, on average, prohibiting other sites from excerpting appears to have no effect on the focal (excerpted) site, and only a small negative impact on the other four.

6.2.2 Outbound Links

We next examine the scenarios whereby the focal site ceases excerpting (as in Germany and Spain). As Table 6 shows, eliminating outbound links reduces traffic at the focal (excerpting) site, with the greatest decreases occurring when the focal site is egotastic or dlisted. Total traffic at egotastic drops by 2% when it stops excerpting, a loss that is due entirely to the removal of excerpts from dlisted (recall from Figure 5 that all of egotastic's outbound links go to dlisted; we discuss this unique case in Section 6.2.3). From an economic standpoint, this 2% loss is meaningful because it would likely translate directly into a 2% loss in advertising revenue (on a CPM basis) and profit (as the marginal cost of serving ads is zero). Table 6 also shows a substantial loss in traffic when dlisted (which provides the most outbound links) stops excerpting. dlisted's links provide substantial benefits to its audience, and their removal would decrease its traffic by 0.5%.

Theoretically, removing outbound links could be beneficial or detrimental to traffic at the linked sites. But as Table 6 shows (and echoing the results from Section 6.2.1), there is almost no impact on the other four sites when dlisted stops linking. Excerpting thus appears to be more beneficial to the linking site than the sites it links to.

Eliminating outbound links also has a consistently negative impact on consumers, as it reduces the likelihood that consumers initiate browsing sessions each day. The greatest decrease in browsing occurs when either egotastic or dlisted stops linking: In both cases, the number of browsing sessions is about 0.35% lower. Note however that because we consider a subset of celebrity news sites, the total impact on consumer browsing across all sites is unclear. For example, consumers might compensate for this decrease in celebrity news consumption by increasing their consumption of other types of news.

6.2.3 Link Exchange

We have so far looked at cases in which a site benefits from *all* of the links it gives or receives. However, because some sites never excerpt each other, these average effects overlook some

meaningful dyadic effects on traffic. Hence, we now take a closer look at a pair of sites with an interesting relationship: dlisted and egotastic. These sites link to each other often and with approximately the same frequency. Because egotastic links exclusively to dlisted, we use the counterfactual scenarios in which egotastic is the focal site to understand the importance of this "link exchange" with dlisted. Specifically, the three counterfactuals involving egotastic correspond directly to three ways the link exchange could break down: 1) dlisted no longer links to egotastic, 2) egotastic no longer links to dlisted, and 3) neither site links to the other. Following Table 7, we discuss each of these scenarios next.

First, we consider how the link exchange affects consumers who visit either dlisted or egotastic at step t, followed by the other at step $t + 1$. As Table 7 shows, and consistent with our theory, if dlisted were to stop linking to egotastic, the share of dlisted's audience going directly to egotastic would drop by 3.2%. Similarly, if egotastic were to stop linking to dlisted, traffic from egotastic to dlisted would drop by 5.5%. In the absence of links, readers at one site do not learn about high match content at the other, reducing the incentive to continue browsing. Note however that the magnitude of these decreases on total traffic at each site is on the order of 0.2% to 0.3%, as only 9% of egotastic's and 6% of dlisted's traffic comes from people who previously visited the other site.

Next, we consider how the link exchange affects consumers who start their sessions at either site. When dlisted stops linking to egotastic, sessions starting at dlisted are 0.67% lower, and when egotastic stops linking to dlisted, sessions starting at egotastic are 2.5% lower. In the absence of links, consumers do not anticipate the potential benefits from observing useful match signals about content at the other site. Hence these sites become relatively less attractive to consumers at the start of their session. Note that the loss in traffic is smaller at dlisted than at egotastic in part because dlisted links to a variety of sites, whereas egotastic only links to dlisted.

Finally, we consider how the link exchange affects total traffic at each of the two sites. Recall that about 9% of egotastic's and 6% of dlisted's traffic comes from people who previously visited the other site. In contrast, 79% of egotastic's and 71% of dlisted's traffic comes from people visiting at the start of a session. Consequently, the overall effect of eliminating the link exchange is primarily determined by how its removal affects consumers starting their sessions at either site. Specifically, when dlisted stops linking to egotastic, traffic at both sites is lower, but dlisted suffers more (−0.47% versus −0.07%). Similarly, when egotastic stops linking to dlisted, egotastic suffers more (−2.11% versus −0.34%). These scenarios show that excerpting is beneficial to both sites, but the greater benefit accrues to the site that does the excerpting.

7 Conclusion

Understanding how excerpting among news sites affects consumers is relevant to 1) content producers, who need to know whether excerpting will be more beneficial to their own sites or the competition's; 2) policy makers, who need to understand whether excerpting generates value for consumers or creates an unfair competitive advantage for content aggregators; and 3) advertisers, who need to know how changes in linking affect the reach and frequency of ads running on multiple sites. In this paper, we present a theory that distinguishes the effects of excerpting on the linking and linked site, and thus generates new insights into the question of why excerpts can be beneficial to the excerpted site in some circumstances and detrimental in others. Moreover, we quantify the magnitude of these effects in an empirical setting in order to assess how excerpting influences consumers browsing for Internet news. These efforts advance our understanding of excerpting, and more generally, the consumption of Internet news.

A novel aspect of this research is that excerpts are modeled as signals of consumers' heterogeneous match with content at the excerpted site. This signaling mechanism allows excerpts to either increase or decrease the likelihood of subsequently visiting the linked site, depending on the valence of the signal. Yet even though our model allows any given link to have a negative effect at the individual level, it also provides a theoretical rationale for why the practice of excerpting may still be generally positive for both the linking and linked sites at the aggregate level.

Specifically, our theoretical results indicate that when the prior probability of visiting an excerpted site is already low, any decreases in the probability of visiting the excerpted site due to lower expected match will be smaller in magnitude than any increases due to higher expected match. For this reason, excerpting should generally have a positive direct effect on the linked site. However, because consumers value excerpts, sites that offer them become more popular, and if this increase in popularity is large enough, excerpting can also have a negative indirect effect on the excerpted sites.

Our empirical results reinforce these insights. Excerpting benefits the excerpted site by increasing the share of traffic originating at the linking site, and benefits the linking site by making it more popular at the start of consumers' browsing sessions. Although the overall impact is positive for both sites, the excerpting site benefits more than the sites it links to. Our results indicate that excerpting is economically important to news sites, as it leads to increased traffic at both the excerpting and linked sites. For example, we find that removing a link exchange between two of the sites in our sample leads to an overall traffic loss on the

order of 1–2%. Owing to the nature of digital advertising, such decreases in traffic translate directly to lower ad revenue and profit.

Findings suggest that excerpting increases the consumption of news, and does so by improving consumers' choices. Excerpts signal to consumers whether they will like the excerpted site more or less than usual, which helps them make better choices. Forward-looking consumers anticipate this benefit, seek out sites offering many excerpts at the start of their browsing sessions, and consume news more efficiently. Consistent with previous studies (Athey and Mobius 2012; George and Hogendorn 2013), we find empirical evidence that excerpting increases news consumption, leading consumers to browse more frequently and visit a wider range of sites.

In addition to generating new theoretical and substantive insights about the consumption of news on the Internet, this study also provides a number of methodological advances. First, our model of excerpts as match signals can be easily applied to other settings where consuming one product leads to learning about another, or where a firm's advertising contains information about its competitors. Second, we formulate a model of news consumption with learning that can be directly applied to the study of other (non-Internet) news media. And third, our estimation procedure, which is based on a combination of two recent advances from the econometrics and statistics literatures (Imai et al. 2009; Girolami and Calderhead 2011), provides a template for more efficient Bayesian estimation of single-agent dynamic discrete choice models.

There are a number of limitations to this study that may provide the basis for future extensions. First, we do not model the strategic decision of whether to link to another site. The decision of which sites to link to may depend, for example, on how similar sites are, or on their relative market power, as well as the distribution of consumer preferences. An empirical study that accounts for these factors might provide new insights into why excerpting is so prevalent among blogs and news sites. Another limitation of this study is that the match locations and link signals are unobserved. An interesting extension would be to model the site's match location as a function of its content, and the match signal as a function of the text immediately surrounding the link. Such insights would guide the design and content of links. Finally, this study has limited its focus to the practice of excerpting among Internet news sites. But excerpting is far more widespread than the specific context of news sites. Thus, it would be interesting to understand how the effects of excerpting differ in other contexts, such as Twitter, Internet discussion boards, and other social media.

References

Aguirregabiria, V., and P. Mira. 2010. "Dynamic discrete choice structural models: A survey." *Journal of Econometrics* 156 (1): 38–67.

Allen, B. 1983. "Neighboring information and distributions of agents' characteristics under uncertainty." *Journal of Mathematical Economics* 12 (1): 63–101.

Allen, B. 1986. "The demand for (differentiated) information." *The Review of Economic Studies* 53 (3): 311.

Allen, B. 1990. "Information as an economic commodity." In *Papers and Proceedings of the Hundred and Second Annual Meeting of the American Economic Association,* edited by R. L. Oaxaca and W. St. John, 268–273. Pittsburgh: American Economic Association, May.

Athey, S., and M. Mobius. 2012. "The impact of news aggregators on Internet news consumption: The case of localization." Working paper.

Ching, A. T., S. Imai, M. Ishihara, and N. Jain. 2012. "A practitioner's guide to Bayesian estimation of discrete choice dynamic programming models." *Quantitative Marketing and Economics* 10 (2): 151–196.

Chiou, L., and C. Tucker. 2015. "Content aggregation by platforms: The case of the news media." NBER Working Paper No. 21404.

Concha, P. P. de la, A. G. García, and H. H. Cobos. 2015. *Impacto del Nuevo Artículo 32.2 de la Ley de Propiedad Intelectual.* Technical report. NERA Economic Consulting. Summarized in https://web.archive.org/web/20150814111804/http://www.aeepp.com/noticia/2272/actividades/informe-economico-del-impacto-del-nuevo-articulo-32.2-de-la-lpi-nera-para-la-aeepp.html as accessed via Google Translate.

Danaher, P. J. 2007. "Modeling Page Views Across Multiple Websites with an Application to Internet Reach and Frequency Prediction." *Marketing Science* 26 (3): 422–437.

Dellarocas, C., Z. Katona, and W. Rand. 2013. "Media, aggregators, and the link economy: Strategic hyperlink formation in content networks." *Management Science* 59 (10): 2360–2379.

Erdem, T., and M. P. Keane. 1996. "Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets." *Marketing Science* 15 (1): 1–20.

George, L. M., and C. Hogendorn. 2013. "Local news online: Aggregators, geo-targeting and the market for local news." Working paper.

Girolami, M., and B. Calderhead. 2011. "Riemann manifold Langevin and Hamiltonian Monte Carlo methods." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73 (2): 123–214.

Goldfarb, A. 2002. "Analyzing website choice using clickstream data." *Advances in Applied Microeconomics* 11:209–230.

Griewank, A., D. Juedes, and J. Utke. 1996. "Algorithm 755: ADOL-C: A Package for the Automatic Differentiation of Algorithms Written in C/C++." *ACM Transactions on Mathematical Software* 22 (2): 131–167.

Imai, S., N. Jain, and A. Ching. 2009. "Bayesian estimation of dynamic discrete choice models." *Econometrica* 77 (6): 1865–1899.

Isbell, K. 2010. "The Rise of the News Aggregator: Legal Implications and Best Practices." The Berkman Center for Internet & Society, Research Publication No. 2010-10.

Johnson, E. J., W. W. Moe, P. S. Fader, S. Bellman, and G. L. Lohse. 2004. "On the Depth and Dynamics of Online Search Behavior." *Management Science* 50 (3): 299–308.

Kim, J. B., P. Albuquerque, and B. J. Bronnenberg. 2010. "Online demand under limited consumer search." *Marketing Science* 29 (6): 1001–1023.

Lee, S., F. Zufryden, and X. Drèze. 2003. "A Study of Consumer Switching Behavior Across Internet Portal Web Sites." *International Journal of Electronic Commerce* 7 (3): 39–63.

Leskovec, J., L. Backstrom, and J. Kleinberg. 2009. "Meme-tracking and the dynamics of the news cycle." In *Proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining*, 497–506. New York: ACM.

Mayzlin, D., and H. Yoganarasimhan. 2012. "Link to success: How blogs build an audience by promoting rivals." *Management Science* 58 (9): 1651–1668.

Musalem, A., E. T. Bradlow, and J. S. Raju. 2009. "Bayesian estimation of random-coefficients choice models using aggregate data." *Journal of Applied Econometrics* 24 (3): 490–516.

Park, Y.-H., and P. S. Fader. 2004. "Modeling browsing behavior at multiple websites." *Marketing Science:* 280–303.

Pew Research Center. 2014. "State of the News Media 2014: Paying for News: The Revenue Picture for American Journalism, and How It Is Changing." March. http://web.archive.org/web/20150429160713/http://www.journalism.org/files/2014/03/Revnue-Picture-for-American-Journalism.pdf.

Roos, J. M. T., and R. Shachar. 2014. "When Kerry met Sally: Politics and perceptions in the demand for movies." *Management Science* 60 (7): 1617–1631.

Rossi, P. E., G. M. Allenby, and R. McCulloch. 2005. *Bayesian statistics and marketing.* John Wiley & Sons, Ltd.

Su, C.-L., and K. L. Judd. 2012. "Constrained optimization approaches to estimation of structural models." *Econometrica* 80 (5): 2213–2230.

Tanner, M. A., and W. H. Wong. 1987. "The calculation of posterior distributions by data augmentation." *Journal of the American Statistical Association* 82 (398): 528–540.

West, M., and J. Harrison. 1999. *Bayesian forecasting and dynamic models.* 2nd. Springer-Verlag.

A Information Utility

We represent information in the following way. On day d there exists a finite maximum amount of news information that could be published. Following Allen (1983; 1986; 1990), we represent this news information as N unique and indivisible "bits" representing the smallest unit of news information that can be relevant (i.e. provide utility) to the consumer. Every day, bits are distributed heterogeneously across sites, and any bit could appear at more than one site. Because N represents a theoretical upper limit on the production of news information, some bits might not appear at any site.

We assume that when the consumer encounters a bit of news information for the first time, it provides an amount of utility and then becomes prior knowledge. Once a bit has become prior knowledge, further encounters with that bit at other sites provide no additional utility. We also assume that knowledge is superior to ignorance, and normalize the utility from the latter to zero. Accordingly, the utility from each bit is positive.

At step t of a browsing session, the utility from seeing the news information at site j depends on 1) which bits were already seen, with the consumer's prior knowledge denoted $k_t \in \{0,1\}^N$ (and where at the start of the session $k_{b,0} = 0 \,\forall$ bits b); 2) the set of bits available at site j, denoted $\iota_j \in \{0,1\}^N$; and 3) the utility provided by each bit, denoted $u \in \Re_+^N$. (The process repeats each day, hence we drop the d and i subscripts for clarity.) The utility from the news information obtained from site j at step t is

$$\beta_{j,t} = \sum_{b=1}^{N} u_b \iota_{j,b} \left(1 - k_{b,t}\right) \tag{A.1}$$

The consumer knows the utility from each bit, as well as which bits were already seen, but does not know which bits will be news that day, nor which ones will appear at each site. Hence u and ι are random variables from the consumer's perspective. In the following subsections, we discuss the consumer's prior and updated beliefs about these variables. First, however, note that conditional on the observing the bits at the next site j, an equivalent specification for (A.1) is

$$\beta_{j,t} = K'\overline{u}' - K_t\overline{u}_t \tag{A.2}$$

where $K_t = \sum_{b=1}^{N} k_{b,t}$ is the cumulative number of bits observed prior to step t, $\overline{u}_t = \frac{1}{K_t}\sum_{b=1}^{N} u_b k_{b,t}$ is their average utility, and K' and $\overline{u}'$ represent their updated values *after* obtaining the new information at site j. Note however that the consumer knows K_t and $\overline{u}_t$ but does not learn K' or $\overline{u}'$ until *after* visiting site j. In the remainder of this section, we derive the consumers expectations about K' and $\overline{u}'$.

A.1 Quantity and Quality of Information

We now present the data-generating process for the ι_j's and u. We refer to the vector ι_j as the "bit availability" or "information quantity" at site j, and the vector u as the "bit utility" or "information quality" of each bit.

Bit availability. The probability of any bit b appearing at site j is decomposed into two factors. The first pertains to the availability of the bit in the environment, the second to its availability at site j. The first factor is the random variable $\pi_b \in [0,1]$ and is common across all sites; it represents the probability of a bit appearing at a site that can publish all of the day's news. The second factor is the site-specific parameter $\alpha_j \in (0,1)$ and is particular to site j, but common across all bits. The two factors jointly define the probability that any bit b appears at site j, that is, $\Pr[\iota_{j,b} = 1|\pi_b]$. We denote this probability $\rho_{j,b}$:

$$\rho_{j,b} \equiv \Pr\left[\iota_{j,b} = 1|\pi_b\right] = 1 - (1 - \pi_b)^{\alpha_j} \tag{A.3}$$

This decomposition is such that when site j publishes more information on average ($\alpha_j \to 1$), then $\rho_{j,b} \to \pi_b$; and when site j publishes less information on average ($\alpha_j \to 0$), then $\pi_b > \rho_{j,b} \to 0$. The parameter α_j thus attenuates the probability of finding information at site j relative to the overall news environment.

Bit utility. The consumer's uncertainty about which bits appear each day leads to uncertainty about their average value. We assume the true distribution of the u_b's available each day is exponential distribution with scale σ (hence σ is the unobserved expected utility per bit):

$$u_b|\sigma \sim Expo\left(\sigma^{-1}\right) \tag{A.4}$$

Although we omit the d subscript here, we emphasize that the consumer receives a different value of σ and a new vector of u_b's each day.

A.2 Prior Beliefs and Updating

The preceding discussion described the distributions determining which bits appear at each site, and the amount of utility they provide. We now describe how observing some of the ι_j's and u_b's leads to updated beliefs about σ and the π_b's, and hence updated forecasts about the ι_j's and u_b's at the remaining sites.

Bit availability. The consumer's prior beliefs about bit availability are assumed to be

$$\tilde{\pi}_{b,0} \overset{i.i.d.}{\sim} Beta\left(\alpha_0, 1\right), \quad \alpha_0 > 0 \tag{A.5}$$

where the tilde (~) indicates variables pertaining to the consumer's beliefs.[12] From the consumer's perspective, it is not necessary to predict the entire set of information at each site, but instead just the subset of bits that were not already seen (i.e., those for which $k_{b,t} = 0$). Hence, after visiting one or more sites, the consumer's updated beliefs about the distribution of the remaining unseen bits is

$$\tilde{\pi}_{b,t}|k_{b,t} = 0, h_t \sim Beta\left(\alpha_0, 1 + A_t\right) \tag{A.6}$$

where h_t is a vector indicating which sites were visited prior to step t, and $A_t \equiv \sum_{j=1}^{J} h_{t,j}\alpha_j$ is the sum of the α_j's for those sites (see Claim 1 in Section H of the Online Appendix for a proof of this result).

It follows that the distribution of the total number of unseen bits at the next site j (of the $N - K_t$ that remain) is binomial with expected value

$$\mathbb{E}\left[K' - K_t|I_t, j\right] = (N - K_t)\left(1 - \frac{B\left(\alpha_0, 1+A_t+\alpha_j\right)}{B(\alpha_0, 1+A_t)}\right) \tag{A.7}$$

where $B(\cdot,\cdot)$ indicates the beta function, and I_t represents the set of state variables at step t, including K_t and h_t, and thus A_t (Claim 2 in Section H of the Online Appendix). The first term $(N - K_t)$ represents the number of unseen bits remaining, and the second term represents the expected probability of observing any of those bits at site j.

In our empirical application, the parameter α_0 is not separately identified from the α_j's. Hence in the remainder of this section we set $\alpha_0 = 1$, whereby Equation (A.7) simplifies to

$$\mathbb{E}\left[K' - K_t|I_t, j\right] = (N - K_t)\left(\frac{\alpha_j}{1+A_t+\alpha_j}\right) \tag{A.8}$$

Bit utility. The consumer's prior beliefs about the average utility from bits each day is assumed to be

$$\tilde{\sigma}_0 \overset{i.i.d.}{\sim} Inv\text{-}Ga\left(\kappa_0 + 1, \kappa_0\lambda\right), \quad \kappa_0 > 0, \lambda > 0 \tag{A.9}$$

The value of λ is consumer i's prior expected utility from any bit of information, and κ_0 indicates the dispersion of this belief. After observing K_t bits with an average utility of $\overline{u}_t$, the consumer's updated belief about the average utility from information is

$$\tilde{\sigma}_t|I_t \sim Inv\text{-}Ga\left(\kappa_0 + K_t + 1, \kappa_0\lambda + K_t\overline{u}_t\right) \tag{A.10}$$

where I_t includes the state variables K_t and $\overline{u}_t$ (Claim 3). Finally, the expected information utility from the content at site j is the expected average utility from each remaining bit times

[12] The i.i.d. assumption here is not restrictive. Consumers only care about the bits they haven't seen yet, which by definition did not appear at any previously visited sites. This means the (lack of) appearance of those unseen bits is uncorrelated with the appearance of the ones that were already encountered, and any violation of the i.i.d. assumption would be due to consumers' prior beliefs. Because consumers do not know which bits will appear each day, we assume independence in their prior beliefs.

the expected number of bits at site j (Claim 4):

$$\mathbb{E}\left[\beta_{j,t}|I_t\right] = \left[\left(\frac{\alpha_j}{1+A_t+\alpha_j}\right)(N-K_t)\right]\left[\lambda + \left(\frac{K_t}{\kappa_0+K_t}\right)(\overline{u}_t - \lambda)\right] \tag{A.11}$$

Distribution of state variables. The value function (8) depends on the joint conditional distribution of the consumer's beliefs about the state variables K' and $\overline{u}'$, given I_t. This joint conditional distribution can be factored as the product of distributions of $\overline{u}'|K', I_t$ and $K'|I_t$. The latter is binomial with expected value given by (A.8). The p.d.f. of the conditional distribution $\overline{u}'|K', I_t$ is (Claim 5):

$$p\left(\overline{u}'|K', I_t\right) = \begin{cases} \frac{K'\left(\frac{K'\overline{u}'-K_t\overline{u}_t}{\kappa_0\lambda+K'\overline{u}'}\right)^{K'-K_t}\left(\frac{\kappa_0\lambda+K_t\overline{u}_t}{\kappa_0\lambda+K'\overline{u}'}\right)^{\kappa_0+K_t+1}}{\left(K'\overline{u}'-K_t\overline{u}_t\right)B(\kappa_0+K_t+1,K'-K_t)}, & K' > K_t \\ \delta_{\overline{u}_t}\left(\overline{u}'\right) & K' = K_t \end{cases} \tag{A.12}$$

When new bits are observed ($K' > K_t$), the p.d.f. of their updated average utility is given by the expression on the top line (this distribution is similar to an inverse-gamma distribution, which is appropriate since $\overline{u}'$ represents an average of gamma-distributed variables). If no new bits are observed, then $\overline{u}' = \overline{u}_t$ with probability 1. Finally, it follows from the preceding derivations that the distribution of $\beta_{j,t}$ indicated in Equation (6) in the main text is

$$\mathscr{F}\left(\beta_{j,t}|I_t, \lambda, \alpha_j\right) = \sum_{K'=K_t}^{N} Ga\left(\beta_{j,t}\middle|K'-K_t, \lambda + \left(\frac{K_t}{\kappa_0+K_t}\right)(\overline{u}_t-\lambda)\right) Binom\left(K'-K_t\middle|N-K_t, \frac{\alpha_j}{1+A_t+\alpha_j}\right) \tag{A.13}$$

B State Variables and Transition Probabilities

Here we provide a full specification of the state variables and their transition probabilities. The consumer's information state is the set $I_t \equiv \left\{n_t, \overline{s}_t, K_t, \overline{u}_t, h_t\right\}$, and we note that $A_t \equiv \sum_j h_{t,j}\alpha_j$. The probability of the next I' conditional on the previous I_t and the decision to visit site j is denoted

$$f\left(I'|I_t, j\right) = f\left(n', \overline{s}', K', \overline{u}', h'|n_t, \overline{s}_t, K_t, \overline{u}_t, h_t, j\right) \tag{B.1}$$

We decompose this distribution in the following way:

$$f\left(I'|I_t, j\right) = p\left(\overline{s}'|n', n_t, \overline{s}_t\right) p\left(n'|n_t, j\right) p\left(\overline{u}'|K', K_t, \overline{u}_t\right) p\left(K'|K_t, h_t, j\right) p\left(h'|h_t, j\right) \tag{B.2}$$

The distribution $p\left(\overline{u}'|K', K_t, \overline{u}_t\right)$ is specified in (A.12) and the distribution of $p\left(K'|K_t, h_t, j\right)$ in (A.7). The distribution of h' is deterministic conditional on the choice j,

$$p\left(h'|h_t, j\right) = \delta_1\left(h_{t,j}\right),$$

i.e., the next h' equals the previous h_t, but with h'_j set to 1. The evolution of n and $\overline{s}$ are specified as follows. First, the distribution of n' is discrete: for any site $j' \neq j$ that has not yet been visited, $n'_{j'}$ will equal $n_{t,j'} + 1$ with probability $\omega_{j,j'}$, and $n_{t,j'}$ with probability $1 - \omega_{j,j'}$. Formally,

$$p\left(n'_{j'}|n_t, j\right) = \omega_{j,j'}\delta_{n_{t,j'}+1}\left(n'_{j'}\right) + \left(1-\omega_{j,j'}\right)\delta_{n_{t,j'}}\left(n'_{j'}\right) \tag{B.3}$$

When a new link to site j' is observed, the value of $\bar{s}'_{j'}$ evolves according to the rules for Bayesian updating of standard Normal conjugate distributions, and when no new link is observed, then $\bar{s}'_{j'} = \bar{s}_{t,j'}$. Formally,

$$p\left(\bar{s}'_j | n'_j, n_{t,j}, \bar{s}_{t,j}\right) = \begin{cases} N\left(z_j + \frac{\tau_s n_{t,j}\left[\bar{s}_{t,j} - z_j\right]}{\tau_s n_{t,j} + \tau_v}, \tau_s^{-1} + \left[n_{t,j}\tau_s + \tau_v\right]^{-1}\right), & n'_j = n_{t,j} + 1 \\ \delta_{\bar{s}_{t,j}}\left(\bar{s}'_j\right), & n'_j = n_{t,j} \end{cases} \tag{B.4}$$

C Word Counts and Information Quantity State Variables

In this section we provide technical details about the relationship between word counts, $w_{j,d}$, and consumers' state variables for information quantity, $K_{i,d,t}$. As shown in Equation (A.8), the state variable $K_{i,d,1}$ follows a binomial distribution with expectation $\mathbb{E}\left[K_{i,d,1}\right] = N\alpha_j/(1+\alpha_j)$. In the theoretical model, this expression characterizes the distribution of both consumer i's beliefs and the realizations of $K_{i,d,1}$. During estimation, we use a different distribution for the realizations of $K_{i,d,1}$. Specifically, we assume that conditional on the observed word count at the first site j visited on day d, consumer i's *realization* of $K_{i,d,1}$ is drawn from a binomial distribution with expected value

$$\mathbb{E}\left[K_{i,d,1} \middle| w_{j,d}\right] = N\, q\left(w_{j,d}\right) \tag{C.1}$$

where the function $q\left(w_{j,d}\right)$ translates the number of words published at site j on day d to an information scale. As mentioned in the main text, Equation (C.1) only applies to the realization of $K_{i,d,1}$, but not to subsequent values of $K_{i,d,t}$ obtained at steps $t > 1$, nor to the consumer's beliefs at any step.

In selecting a function $q\left(w_{j,d}\right)$, we face the following constraint: the parameters α_j must lie within the range of $(0,1)$, hence the function $q\left(w_{j,d}\right)$ must map $w_{j,d}$ to the interval $\left(0, \frac{1}{2}\right)$. The following half-logit function satisfies this restriction.

$$q\left(w_{j,d}\right) = \frac{2}{1+\exp\left(-w_{j,d}c_w\right)} - 1 \qquad c_w \equiv \frac{\log 3}{\max\left\{w_{j,d}\right\}} \tag{C.2}$$

Equation (C.2) is such that if a site publishes zero words on day d, consumer i would see a quantity of information with expected value $K_{i,d,1} = 0$; if the site publishes $\max\left\{w_{j,d}\right\}$ words, then consumer i would see a quantity of information with expected value $K_{i,d,1} = N/2$.

D Parameter Normalizations, Transformations, and Prior Distributions

We now provide discuss the parameter normalizations listed in Section 4.4, transformations of the data-augmented state variables, and the prior distributions of the remaining parameters.

The parameters N and κ_0 (related to news utility) cannot be separately identified from the individual taste parameters, λ_i. For example, doubling the number of bits N, while dividing κ_0 and each λ_i by two, yields the same choice probabilities. Moreover, simulations indicate

that while κ_0 is identified by the browsing data, there is insufficient variation to estimate this parameter with any meaningful precision. Hence we normalize both variables. We set $N = 30$, reflecting an upper limit of 30 celebrity news items each day, and $\kappa_0 = 4$, ensuring average utility per bit has finite variance. And as mentioned in Appendix A.2, the α_j's cannot be separately identified from α_0, hence we normalize $\alpha_0 = 1$ (note that all equations presented in the main text reflect this normalization).

Daily deviations in match position, $\nu_{j,d}$, and match signals from excerpts, $s_{j,k,d}$, are both latent constructs, and we cannot separately identify their scales. Instead, we set $\tau_\nu = 1$ and interpret τ_s as the ratio of their precisions. Average match locations, z_j, are also latent constructs, and we normalize them with respect to consumer's match preferences, v_i, by setting the mean of the z_j's to be zero. Finally, in order to avoid a degenerate posterior density for the v_i's of the type described in Roos and Shachar (2014), we set the prior intercept and scale of the v_i's to $\eta_v = 0$ and $\zeta_v = 1$, respectively.

Because the state variables for the amount of information, K, average utility per bit, $\overline{u}$, and average signal value for each site, $\overline{s}$ are unobserved, we use data augmentation (Tanner and Wong 1987; Rossi et al. 2005) to sample these state variables along with the model primitives and then integrate over them numerically. To improve the efficiency of our sampling procedure, we transform $\overline{u}$ and $\overline{s}$ in the following ways. First, we substitute σ in our theoretical model of news information with $\sigma^* \equiv \sigma\lambda^{-1}$, and the related state variable $\overline{u}$ with $\overline{u}^* = \overline{u}\lambda^{-1}$. Hence the state variable $\overline{u}^*$ does not depend on λ, and has a prior expectation of 1. Second, we define $s^*_{j,\ell,d} \equiv \left(s_{j,\ell,d} - z_j - \nu_{j,d}\right)\tau_s^{-\frac{1}{2}}$ so that $s^*_{j,\ell,d}$ follows a standard normal distribution independent of z_j and $\nu_{j,d}$. We enforce the identifying restrictions $\mathbb{E}\left(s^*_{j,\ell,d}\right) = 0$ and $\mathbb{V}\left(s^*_{j,\ell,d}\right) = 1$ via pairwise sampling of the s^*'s using the method of Musalem et al. (2009). A parallel strategy is used to sample the data augmented $\nu_{j,d}$'s.

The prior distributions for the remaining parameters are:

$$\text{logit}\,\alpha_j \sim N(0,1) \qquad z_j \sim N(0,1) \qquad \tau_s^{-\frac{1}{2}} \sim Ga(.4,5) \Rightarrow \mathbb{E}\left(\tau_s^{-\frac{1}{2}}\right) = 2 \tag{D.1}$$

$$\eta, \phi|\zeta \sim N\left(0, 10^6\zeta^2\right),\ \zeta^{-2} \sim \chi^2_1 \text{for } \lambda \text{and } \gamma \qquad \gamma_w \sim N(0,1) \qquad \phi_v \sim N(0,1) \tag{D.2}$$

Plots depicting the prior (and posterior) distributions can be found in Section I of the Online Appendix.

E Alternative Model Specifications

We estimate and compare three alternatives to our full specification: 1) *myopic*—consumers are not forward-looking and do not anticipate seeing excerpts; 2) *no links*—with myopic consumers and no link data at all; and 3) *no words*—the full specification estimated without word

Table 8: Posterior Predictive Fit Based on Mean Absolute Percent Error of Various Summary Statistics

	Full	Myopic	No links	No words
Statistic				
Total traffic per site	21.1	23.6	23.9	19.2
Daily traffic per site	28.8	31.8	32.0	29.7
Daily transitions	41.0	45.0	45.9	53.6
Total visits per consumer	9.6	7.4	9.1	10.4
Model feature				
Forward-looking	X			X
Links	X	X		X
Word count data	X	X	X	

NOTES: Fit statistics describe the mean absolute percent error between posterior predictions and the observed data. All models are estimated using all 19,130 browsing observations.

count data. Comparisons between models are based on the posterior predictive fit with observed browsing based on the mean absolute percent error (MAPE) of four summary statistics: 1) total traffic per site, 2) daily traffic per site, 3) daily traffic between each pair of sites ("transitions"), and 4) total site visits per consumer.

Table 8 shows the full model has the best overall fit with observed browsing. The two models with forward-looking consumers and link data fit better on 3 out of 4 measures (doing worse on the number of site visits per consumer) compared to the models that limit the role of the link data. Of the two models with forward-looking individuals and links, the model with word count data fits better on 3 out of 4 measures (doing worse on total traffic per site).

Estimates for site parameters (match location, z_j, and information quantity, α_j) are qualitatively similar across models, with the following exceptions. First, estimates from both the myopic and no-links models differ from those in the full model, particularly with respect to the match location parameters. But there is little difference among these two models (myopic and no-links), suggesting that the explanatory value of excerpts is highest when consumers are modeled as forward-looking. Second, the greatest difference in parameter estimates arises between the full model and its no-word counterpart. When the word count data are not used for estimation, the estimate for average information quantity is somewhat higher for dlisted (0.34 vs. 0.24), and significantly higher for celebuzz (0.33 vs. 0.03). This suggests that in the case of celebuzz, word counts may be a poor indicator of its news quantity (but a good one for the other four sites).

Online Appendix

F Simulation Procedure

Here we document the simulation procedure on which the analysis reported in Section 2.4 is based, and provide further detail not reported in the main text. We simulate browsing for two types of consumers—1) myopic, and 2) forward-looking—under three types of excerpting—1) no links, 2) links are noisy signals ($\tau_s = .2$), and 3) links are informative signals ($\tau_s = 2$). We simulate 30,000 browsing sessions under each of the six conditions.

We set $\lambda = 0$ to ensure $\beta = 0$ at both sites, and $\gamma = 2$ to ensure the sites are not visited too often. The consumer's match preference is $v = 2$, and the two sites are located at $z = 0$ (thus providing equal match utility on average as reported in the main text). Finally, site L always links to site R (but not the reverse), hence $\omega_{L,R} = 1$ and $\omega_{R,L} = 0$.

Initiating a browsing session. Figure 6 shows the probability of initiating a browsing session (i.e., visit at least one site on any given day) under the six conditions. Forward-looking consumers are increasingly likely to initiate browsing sessions as links become more informative. When links are especially informative, the anticipated future benefits are even higher because consumers can choose to visit the linked site only when it provides very high match. Myopic consumers, on the other hand, are insensitive to the precision of link signals, since they cannot anticipate the future benefits from seeing excerpts.

Share of sessions starting at the linking site. Because the two sites offer identical match utility in expectation, myopic consumers are equally likely to start their sessions at both sites, as seen in Figure 7. Forward-looking consumers behave the same when there are no links,

Figure 6: Probability of Initiating a Browsing Session

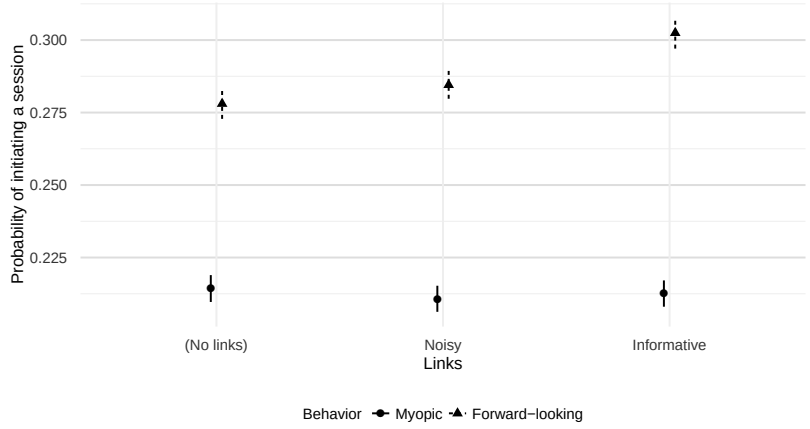


Figure 7: Share of Sessions Starting at Linking Site (L)

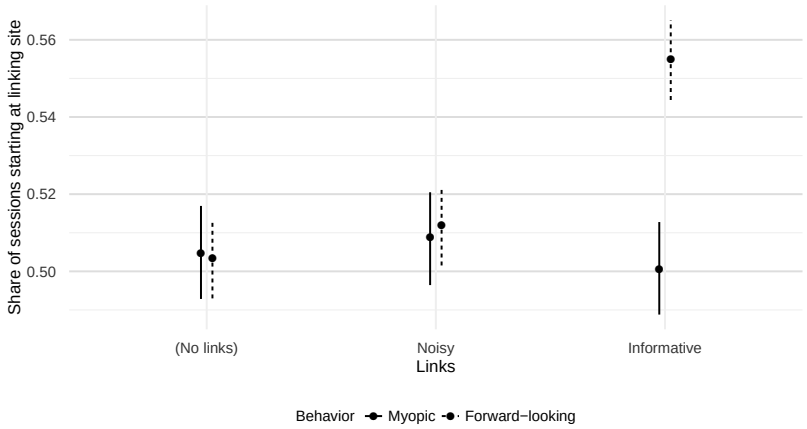
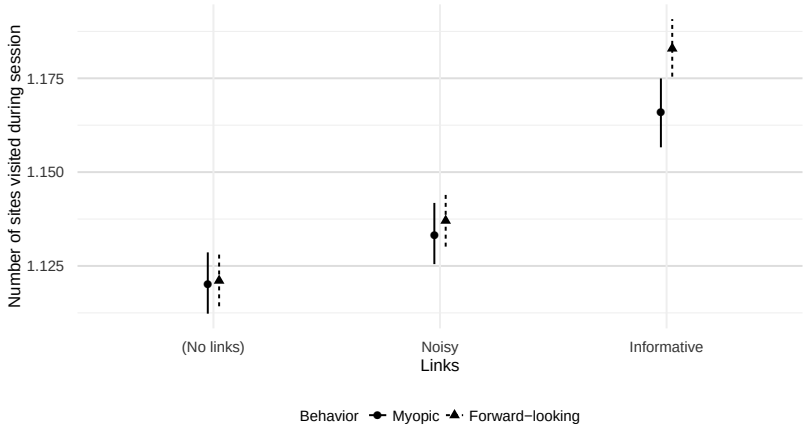


Figure 8: Number of Sites Visited Conditional on Browsing



but as links become more informative, they are increasingly likely to start their sessions at the linking site (L) given the anticipated future benefits from seeing excerpts from site R.

Number of sites visited per session. Figure 8 shows that as links convey more information, both myopic and forward-looking consumers visit more sites (conditional on having initiated a session—i.e., the denominator in this average is the number of sessions in each condition). The increase in session length is due to the consumer being more likely to visit site R after seeing an excerpt at site L. The even greater increase among forward-looking consumers is due to their greater likelihood of initiating their session at site L when links are informative.

Share of sessions visiting the linked site. Figure 9 shows that when links are informative, total traffic at the linked site is higher. The increase in traffic going to site R is highest if consumers are myopic, however, because forward-looking consumers *delay* their visits to the

Figure 9: Share of Sessions Visiting the Linked Site (R)

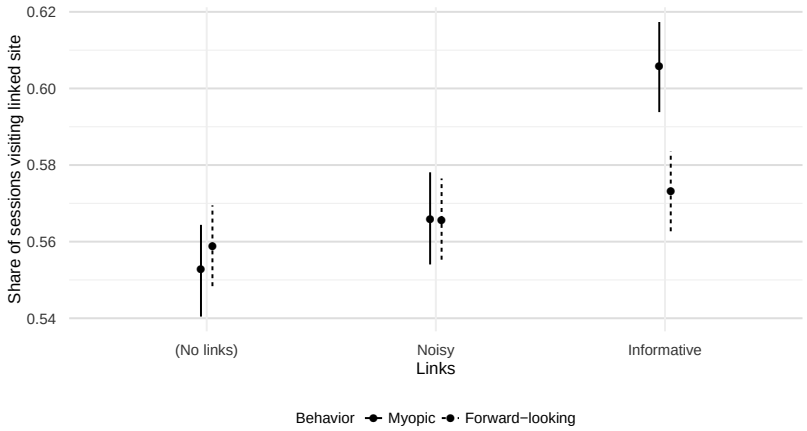
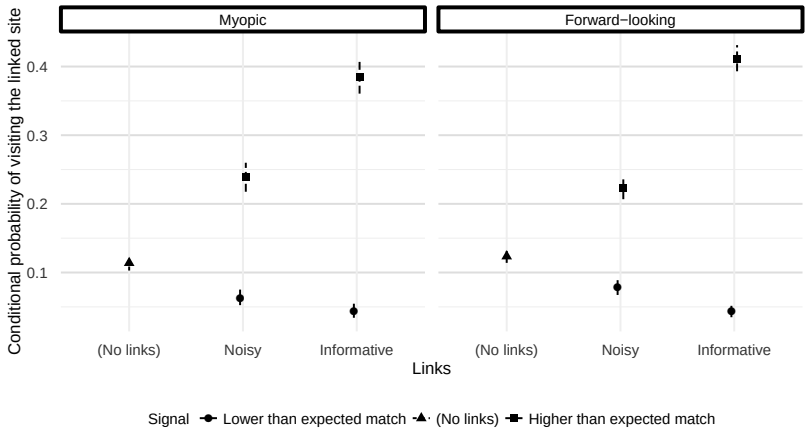


Figure 10: Effect of Signal Valence on Probability of Visiting Linked Site



NOTES. Probabilities are calculated conditional on having chosen to visit the linking site (L) first in the session.

linked site, and sometimes choose to end their session before visiting R.

Effect of signal valence. Figure 10 shows the asymmetric effect of match signals on visit probabilities by considering only sessions that begin at site L. When the excerpt at site L signals higher than average match, then the probability of subsequently visiting site R increases. (The amount of the increase is the same for forward-looking and myopic consumers.) Similarly, when the excerpt at L signals lower than average match, then the probability of subsequently visiting site R decreases. The magnitude of the decrease, however, is smaller than the magnitude of the increase because the probability of subsequently visiting R is already low to start with. That is, there is a floor effect limiting the damage that low match signals can inflict on the excerpted site.

G Sampling Algorithm

The general sampling procedure is outlined in Algorithm 1. This algorithm is a straightforward application of IJC (Imai et al. 2009), with a few differences. At the lines marked [1] in Algorithm 1, a single procedure calculates both $p(\theta|\mathscr{W})$ and $\mathscr{D}_\theta$ using automatic differentiation. In the MMALA procedure (Girolami and Calderhead 2011), the value of $\mathscr{D}_\theta$ would typically contain derivatives of the log posterior density function. In our setting, however, $\mathscr{D}_\theta$ contains the derivatives of the log posterior function while ignoring the contributions to these derivatives from the IJC emax approximation subroutine. The loss in precision in calculating $\mathscr{D}_\theta$ is compensated for by lower computational burden.

At the lines marked [2] and [3] in Algorithm 1, the function $f(\cdot,\cdot)$ indicates the MMALA proposal distribution described in Girolami and Calderhead (2011). At line [2], the proposal distribution is created conditional on the current parameter vector θ and the derivatives of the log posterior density function evaluated at the point θ, $\mathscr{D}_\theta$. At line [3], the proposal distribution is created conditional on the proposed parameter vector θ^c and the derivatives of the log posterior density function evaluated at the point θ^c, $\mathscr{D}_{\theta^c}$. The proposal distributions are not symmetric, and therefore do not cancel out of the Metropolis-Hastings accept/reject ratio.

Finally, the line marked [4] indicates calculation of a new value function iteration for step t of the IJC sample, as described in (Imai et al. 2009). IJC recommend increasing the efficiency of the sampler by using θ^c to calculate the next approximation of the emax function. Because θ^c has greater distance from θ compared to a random walk sampler owing to the MMALA proposal distribution, we θ to be more efficient when calculate our estimate of the emax function.

H Proofs of Results in Appendix A

Claim 1. After visiting one or more sites, the consumer's updated beliefs about the distribution of the remaining unseen bits is

$$\tilde{\pi}_{b,t}|k_{b,t} = 0, h_t \sim Beta(\alpha_0, 1 + A_t) \tag{H.1}$$

where h_t is a vector indicating which sites have been visited prior to step t, and $A_t \equiv \sum_{j=1}^{J} h_t \alpha_j$ is the sum of the α_j's for those sites.

Proof. Denoting by α_t the value of α_j for the site visited at step t, we can write the likelihood of *not* observing each bit b after visiting $t-1$ sites as $(1-\tilde{\pi}_b)^{\alpha_1} \cdots (1-\tilde{\pi}_b)^{\alpha_{t-1}} = (1-\tilde{\pi}_b)^{A_t}$. Hence

Algorithm 1: MCMC sampling procedure. At each iteration, parameters are sampled in blocks. The value function is then iterated and the result either replaces the oldest saved iteration or is appended to the set of saved iterations.

initialize saved MCMC samples: Θ
saved value function iterations: $\mathscr{W}$
foreach *MCMC iteration* t **do**
  foreach *parameter block* $\theta \equiv \theta_b^{(t-1)}$ **do**
    Propose new θ using mMALA proposal distribution:
      calculate marginal posterior probability: $p(\theta|\mathscr{W})$
        derivatives of log posterior probability: $\mathscr{D}_\theta$ // [1]
      set $\left(\mu, \Sigma\right) \leftarrow f(\theta, \mathscr{D}_\theta)$ // [2]
      draw $\theta^c \leftarrow N(\mu, \Sigma)$
    Maintain detailed balance:
      calculate $p(\theta^c|\mathscr{W})$ and $\mathscr{D}_{\theta^c}$ // [1]
      set $\left(\mu^\circ, \Sigma^\circ\right) \leftarrow f(\theta^c, \mathscr{D}_{\theta^c})$ // [3]
    Accept or reject proposal:
      set $\alpha \leftarrow \frac{p(\theta^c)N(\theta|\mu^\circ,\Sigma^\circ)}{p(\theta)N(\theta^c|\mu,\Sigma)}$
      draw $u \leftarrow U(0,1)$
        if $u < \alpha$ **then set** $\theta_b^{(t)} \leftarrow \theta^c$
      else set $\theta_b^{(t)} \leftarrow \theta$
  Iterate value function using IJC:
    draw $I \leftarrow p(I|\theta^{(t)})$
    calculate $\widehat{W} \leftarrow f(I, \theta^{(t)})$ using IJC // [4]
  Save parameters and value function:
    append $\mathscr{W} \leftarrow \{\widehat{W}, I, \theta^{(t)}\}$
    append $\Theta \leftarrow \theta^{(t)}$

the posterior distribution of $\tilde{\pi}_b$ for any bits b that have not yet been seen is:

$$\frac{(1-\tilde{\pi}_b)^{A_t}\,\tilde{\pi}_b^{\alpha_0-1}\,(1-\tilde{\pi}_b)^{1-1}\,[B(\alpha_0,1)]^{-1}}{\int_0^1 (1-\tilde{\pi}_b)^{A_t}\,\tilde{\pi}_b^{\alpha_0-1}\,(1-\tilde{\pi}_b)^{1-1}\,[B(\alpha_0,1)]^{-1}\,d\tilde{\pi}_b} = \frac{(1-\tilde{\pi}_b)^{(A_t+1)-1}\,\tilde{\pi}_b^{\alpha_0-1}}{B(\alpha_0,A_t+1)} \tag{H.2}$$

This is the p.d.f. of the $Beta\,(\alpha_0, A_t+1)$ distribution. □

Claim 2. The distribution of the total number of *new* bits (of the $N-K_t$ that remain) at the next site j is binomial with expected value

$$\mathbb{E}\left[K'-K_t|I_t,j\right] = (N-K_t)\left(1-\frac{B\left(\alpha_0,1+A_t+\alpha_j\right)}{B\left(\alpha_0,1+A_t\right)}\right) \tag{H.3}$$

where $B(\cdot,\cdot)$ represents the beta function, and I_t represents the set of state variables at step t, including K_t and h_t (and thus A_t).

Proof. The conditional probability bit b is found at site j at step t is $\tilde{\rho}_{j,b,t}|\tilde{\pi}_{b,t} = 1-\left(1-\tilde{\pi}_{b,t}\right)^{\alpha_j}$. The marginal predictive probability is found by integrating over the updated distribution for $\tilde{\pi}_{b,t}$:

$$\mathbb{E}\left[\tilde{\rho}_{j,b,t}|I_t\right] = \int_0^1 \left(1-\left(1-\tilde{\pi}_{b,t}\right)^{\alpha_j}\right) Beta\left(\tilde{\pi}_{b,t}|\alpha_0,1+A_t\right) d\tilde{\pi}_{b,t} = 1-\frac{B\left(\alpha_0,1+\alpha_j+A_t\right)}{B\left(\alpha_0,1+A_t\right)} \tag{H.4}$$

Expected probabilities for all unseen bits are i.i.d, hence the number of bits encountered, among the $N-K_t$ remaining, follows a binomial distribution. □

Claim 3. After observing K_t bits with an average utility of $\overline{u}_t$, the consumer's updated belief about the average utility from information that day is

$$\tilde{\sigma}_t|I_t \sim Inv\text{-}Ga\left(\kappa_0+K_t+1,\,\kappa_0\lambda+K_t\overline{u}_t\right) \tag{H.5}$$

where I_t includes the state variables K_t and $\overline{u}_t$.

Proof. The consumer's beliefs are that individual bit utilities u_b are i.i.d. exponential with scale $\tilde{\sigma}_t$, hence the sum of K_t such utilities follows a gamma distribution:[13]

$$K_t\overline{u}_t|\tilde{\sigma}_t \sim Ga\left(K_t,\tilde{\sigma}_t\right) \tag{H.6}$$

The prior distribution for $\tilde{\sigma}$ is conjugate to this likelihood, hence the updated posterior for $\tilde{\sigma}$ is also inverse-gamma, with shape κ_0+K_t+1 and scale $\kappa_0\lambda+K_t\overline{u}_t$. □

[13] Note that any dependencies that might exist among the u_b's, conditional on their common expectation σ, do not affect consumers' choices because consumers cannot meaningfully update their beliefs. Each day, a consumer observes one draw of the vector u (and typically just partially observes some of the u_b's). In the absence of repeated observations from this distribution, it would be impossible to update beliefs about dependencies among u_b's.

Claim 4. The expected information utility from the content at site j is the expected average utility per remaining bit, times the expected number of bits at site j:

$$\mathbb{E}\left[\beta_{j,t}|I_t\right] = \left[\left(\frac{\alpha_j}{1+A_t+\alpha_j}\right)(N-K_t)\right]\left[\lambda + \left(\frac{K_t}{\kappa_0+K_t}\right)(\overline{u}_t - \lambda)\right] \tag{H.7}$$

Proof. By the law of iterated expectations:

$$\mathbb{E}\left[\beta_{j,t}|I_t\right] = \mathbb{E}\left[K'\overline{u}' - K_t\overline{u}_t|I_t, j\right] = \mathbb{E}\left[K'\mathbb{E}\left[\overline{u}'|K', I_t\right]\middle|I_t, j\right] - K_t\overline{u}_t \tag{H.8}$$

First, the inner expectation is

$$\mathbb{E}\left[\overline{u}'|K', I_t\right] = \frac{1}{K'}\left\{\mathbb{E}\left[\tilde{\sigma}_t|I_t\right](K'-K_t) + \overline{u}_t K_t\right\} = \frac{\kappa_0\lambda + K_t\overline{u}_t}{\kappa_0+K_t}\left(1-\frac{K_t}{K'}\right) + \overline{u}_t\frac{K_t}{K'} \tag{H.9}$$

Then substituting this into the outer expectation yields

$$\mathbb{E}\left[\beta_{j,t}|I_t\right] = \mathbb{E}\left[K'\left\{\frac{\kappa_0\lambda + K_t\overline{u}_t}{\kappa_0+K_t}\left(1-\frac{K_t}{K'}\right) + \overline{u}_t\frac{K_t}{K'}\right\}\middle|I_t, j\right] - K_t\overline{u}_t = \mathbb{E}\left[K'-K_t|I_t, j\right]\frac{\kappa_0\lambda + K_t\overline{u}_t}{\kappa_0+K_t} \tag{H.10}$$

The result is obtained by substituting $\mathbb{E}\left[K'-K_t|I_t, j\right]$ from Equation (H.3) and rearranging terms. □

Claim 5. The p.d.f. of the conditional distribution $\overline{u}'|K', I_t$ is

$$p\left(\overline{u}'|K', I_t\right) = \begin{cases} \frac{K'\left(\frac{K'\overline{u}'-K_t\overline{u}_t}{\kappa_0\lambda+K'\overline{u}'}\right)^{K'-K_t}\left(\frac{\kappa_0\lambda+K_t\overline{u}_t}{\kappa_0\lambda+K'\overline{u}'}\right)^{\kappa_0+K_t+1}}{\left(K'\overline{u}'-K_t\overline{u}_t\right)B(\kappa_0+K_t+1, K'-K_t)}, & K' > K_t \\ \delta_{\overline{u}_t}\left(\overline{u}'\right) & K' = K_t \end{cases} \tag{H.11}$$

Proof. Start with the following two distributions:

$$\beta_{j,t}|K', \tilde{\sigma}_t, I_t \sim Ga\left(K'-K_t, \tilde{\sigma}_t\right) \quad \text{and} \quad \tilde{\sigma}_t|I_t \sim Inv\text{-}Ga\left(\kappa_0+K_t+1, \kappa_0\lambda+K_t\overline{u}_t\right)$$

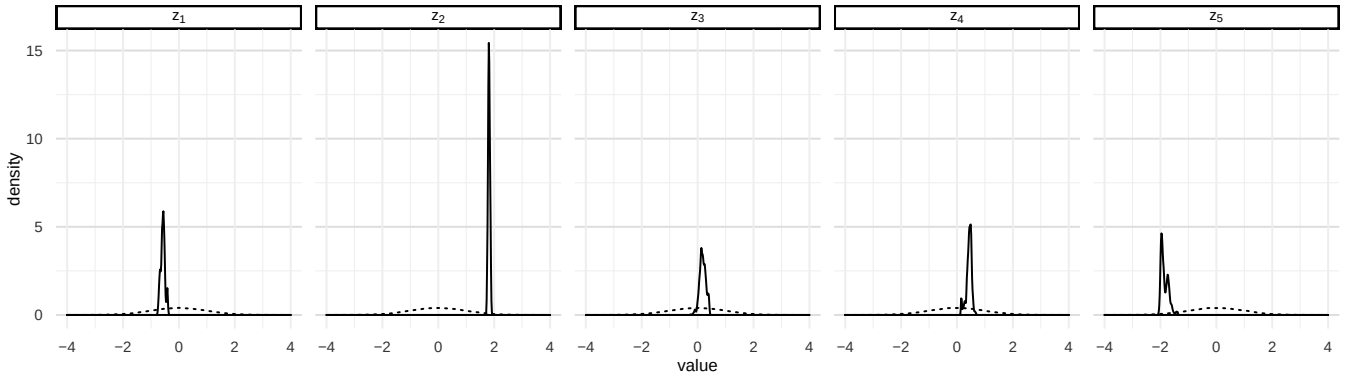
These define the joint distribution $p\left(\beta_{j,t}, \tilde{\sigma}_t|K', I_t\right)$. First integrate over $\tilde{\sigma}_t$ to get the marginal distribution

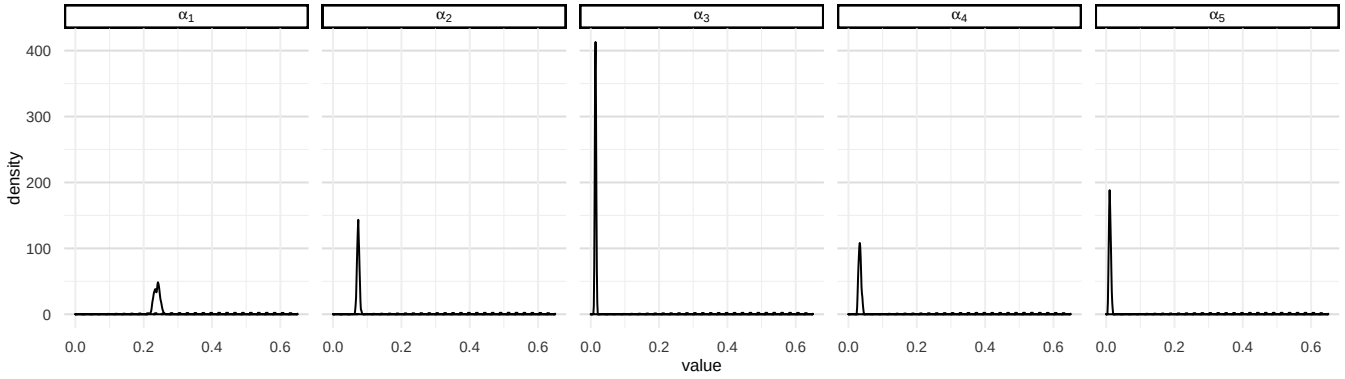
$$p\left(\beta_{j,t}|K', I_t\right) = \frac{\left(\frac{\beta_{j,t}}{\beta_{j,t}+\kappa_0\lambda+K_t\overline{u}_t}\right)^{K'-K_t}\left(\frac{\kappa_0\lambda+K_t\overline{u}_t}{\beta_{t,j}+\kappa_0\lambda+K_t\overline{u}_t}\right)^{\kappa_0+K_t+1}}{\beta_{t,j}B\left(\kappa_0+K_t+1, K'-K_t\right)} \tag{H.12}$$

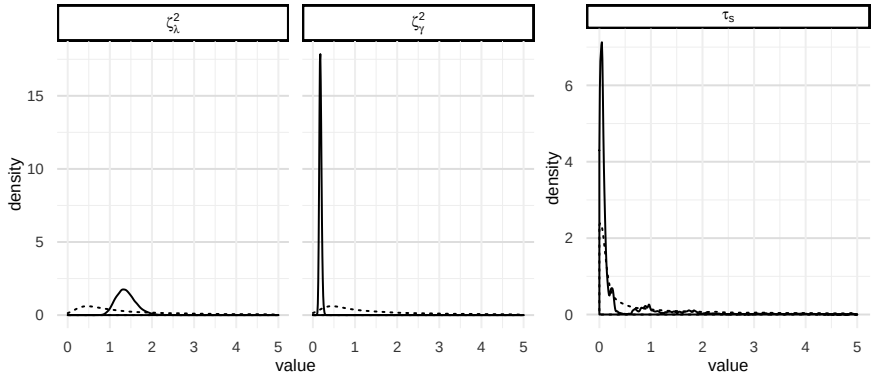
Next, perform the change of variables $\beta_{j,t} = \overline{u}'K' - \overline{u}_t K_t$ to obtain the distribution of $\overline{u}'|K'$ for $K' > K_t$. □

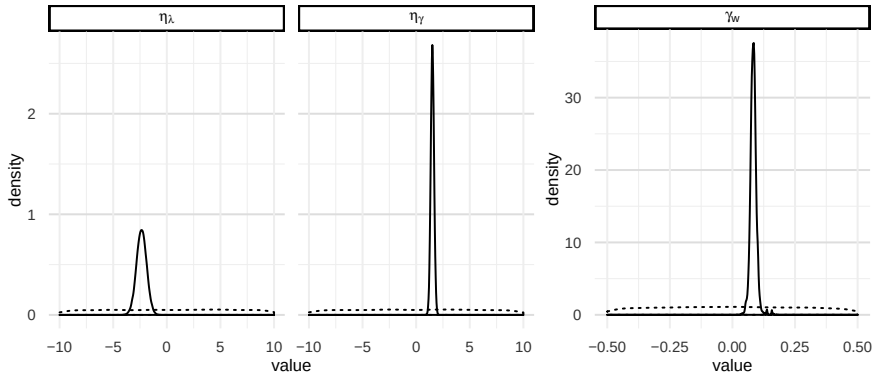
I Marginal Priors and Posteriors

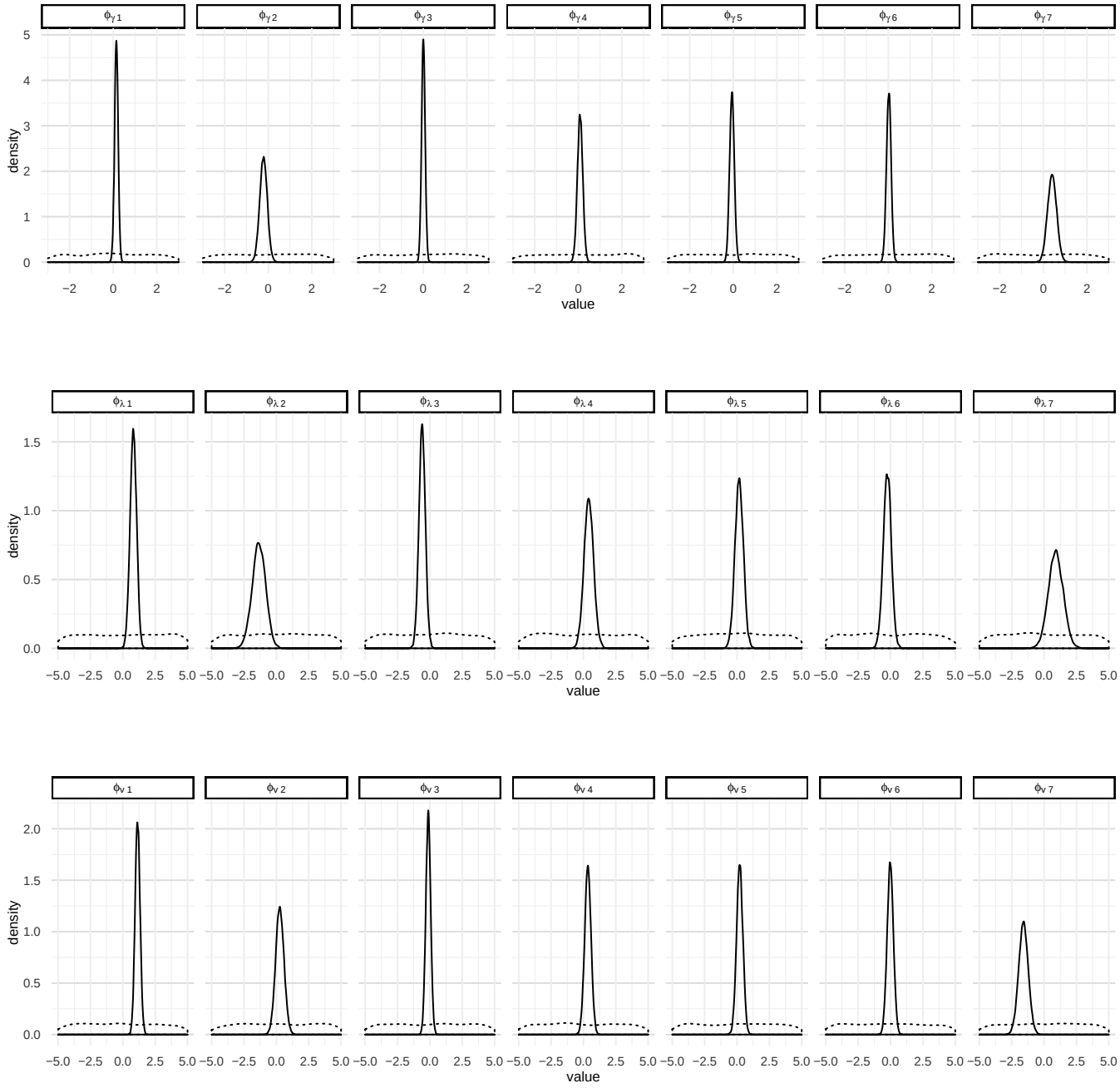
The following plots show the marginal prior (dashed lines) and posterior (solid lines) distributions for each model parameter in the full model. Plots show kernel-smoothed densities based on samples drawn from each distribution, and are truncated for parameters with very diffuse priors (relative to the posterior).


z_1
z_2
z_3
z_4
z_5
density
value


α_1
α_2
α_3
α_4
α_5
density
value


ζ_λ^2
ζ_γ^2
τ_s
density
value


η_λ
η_γ
γ_w
density
value



References

Girolami, M., and B. Calderhead. 2011. "Riemann manifold Langevin and Hamiltonian Monte Carlo methods." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73 (2): 123–214.

Imai, S., N. Jain, and A. Ching. 2009. "Bayesian estimation of dynamic discrete choice models." *Econometrica* 77 (6): 1865–1899.