Attention Allocation in Information-Rich Environments: The Case of News Aggregators

Mihai Calin, Elia Palme, Juliana Sutanto

Swiss Federal Institute of Technology (ETH)

Scheuchzerstrasse 7

8092 Zürich, Switzerland

{mcalin, epalme, jsutanto}@ethz.ch

Chrysanthos Dellarocas

Boston University

595 Commonwealth Avenue

Boston, MA 02215, USA

dell@bu.edu

September 12, 2012

ABSTRACT

News aggregators have emerged as an important component of digital content ecosystems, attracting traffic by hosting collections of links to third party content, but also creating conflict with content producers. Aggregators provide titles and short summaries (snippets) of articles they link to. Content producers claim that their presence deprives them of traffic that would otherwise flow to their sites. In light of this controversy, we conduct a series of field experiments whose objective is to provide insight with respect to how readers allocate their attention between a news aggregator and the original articles it links to. Our experiments are based on manipulating elements of the user interface of a Swiss mobile news aggregator. We examine how key design parameters, such as the length of the text snippet that an aggregator provides about articles, the presence of associated photos as well as of other related articles on the same story, affect (a) a reader's propensity to click on an article, (b) the amount of time that the reader spends on that article after clicking, and (c) the amount of time the reader spends on the aggregator. Our study sheds light into how the presence of aggregators affects the allocation of user attention to original article sites as well as how it impacts competition among articles written on the same topic.

JEL CLASSIFICATION: C93, D81, D83

ACKNOWLEGMENTS. Dellarocas gratefully acknowledges financial support from Microsoft Academic Relations

1. INTRODUCTION

Few industries have suffered more severe disruption by digital technologies than news and journalism. Traditional content creators, such as newspapers, are witnessing their geographical monopolies dissolving into the globally competitive Internet and some of their most important sources of revenue, such as classified ads, migrating to specialized online marketplaces like eBay and Craigslist. User-generated content, such as blogs and online reviews, has increased the supply of content that often competes head-on for readers' attention with professionally produced content.

An equally disruptive transformation is currently underway in these industries. The overwhelming amount of content available online has increased the importance of *curation and aggregation*, that is, of interfaces and services that help readers filter and make sense of the subset of content that is important to them. Historically such functions used to be the realm of professional editors: editors not only commissioned the production of content but also decided what content would be included in a newspaper or magazine issue and how it would be organized.

Web technologies allow this important function to be unbundled from content production. Specifically, the web's ability to place hyperlinks across content has enabled new types of players, commonly referred to as *content aggregators* or *web portals*, to successfully enter the professional content ecosystems, attracting traffic and revenue by hosting collections of links to the content of others (Dewan et al. 2004). Aggregators produce little or no original content; they usually provide titles and short summaries of the articles they link to (Figure 1). Well known aggregators include Google News, the Drudge Report and the Huffington Post. Table 1 provides a more extensive list of examples.

Figure 1 Example of a news aggregator article entry

URL	Description
news.google.com	Algorithmically aggregated headlines; a search engine of many of the world's news sources. Search results group similar articles together.
huffingtonpost.com	Hybrid of news aggregator and original content creator.
drudgereport.com	Aggregator containing selected hyperlinks to news websites all over the world; each link carrying a headline written by the site's editors.
digg.com	User-generated news aggregator. Users post links to stories. Votes promote stories to the front page.
techmeme.com	Blog aggregator tracking other technology blogs.
popurls.com	Meta-aggregator; aggregates links to the top stories chosen by other aggregators.
reddit.com	User-generated news aggregator. Users post links to stories. Votes promote stories to the front page.
newsvine.com	Community news aggregator; member voting determines the position of news stories.
newsnow.co.uk	U.K.-based news portal.

Table 1 Examples of popular news aggregators

Facing severe financial pressures, some content creators have turned against content aggregators, accusing them of "stealing" their revenues by free riding on their content.[1] Other market actors

[1] The recent dispute between the Associated Press and News Corporation with Google is perhaps the best example. See http://www.forbes.com/2009/04/06/google-ap-newspapers-business-media-copyright.html

point out that in today's "link economy", the links bring valuable traffic to the target nodes. Therefore content creators should be happy that aggregators exist and direct consumers to their sites (Karp 2007; Jarvis 2008). Key aggregator executives, such as Google's Eric Schmidt, assert that it is to their interest to see content creators thrive, since the value of links (and aggregators) is directly related to the quality of content that these point to.[2]

A central aspect of the debate focuses on the complex economic implications of the process of placing (for the most part) free hyperlinks across content nodes. The main argument in favor of aggregators is that, if links are chosen well, then they point to good quality content; as a result, they reduce the search costs of the consumers, which may lead to an aggregate increase in content consumption and to more traffic for higher quality sites. The main argument against aggregators is that some consumers satisfy their curiosity by reading an aggregator's short snippet of a linked-to article and never click through to the article itself. In fact, the question of whether aggregators are legally permitted to reproduce an article's title and snippet without obtaining permission from (and possibly paying) the content producer is still unresolved.[3]

Even though there is still an open question of whether the current generation of news aggregators is beneficial or harmful to the content ecosystems (Chiou and Tucker 2011; Athey and Mobius 2012), we believe that the ever-increasing volume of available content makes some form of aggregation an inevitable and valuable component of every content ecosystem. The key question, therefore, is not whether aggregators should exist, but rather how the partly symbiotic and partly competitive relationship between aggregators and content creators can be optimized for the benefit of both parties.

---

[2] "CEO Eric Schmidt wishes he could rescue newspapers", Fortune January 7, 2009.

[3] Aggregators claim that the reproduction of titles and short snippets of text falls under the "fair use" provisions of copyright law. However, as stated by Isbell (2010), "for all of the attention that news aggregators have received, no case in the United States has yet definitively addressed the question of whether their activities are legal."

To provide answers to these questions, we conduct a series of field experiments whose objective is to provide insight with respect to how readers distribute their attention between a news aggregator and the original articles it links to. In addition, we look at how aggregators are impacting competition among articles written on the same story. Our experiments are based on manipulating elements of the user interface of a Swiss mobile news aggregator. We examine how key design parameters, such as the length of the text snippet that an aggregator provides about articles, the presence of associated photos as well as of other related articles on the same story, affect a reader's propensity to click on an article, the amount of time that the reader spends on that article after clicking, and the amount of time that the reader spends on the aggregator.

Gaining a firmer empirical understanding of these relationships will be valuable not only for aggregators seeking to optimize their own traffic patterns, but also in terms of informing the public discourse between aggregators and content creators on the need for equitable profit sharing agreements between the two parties.

In the rest of the paper we discuss related work (Section 2), develop our conceptual framework (Section 3), describe our experimental setting (Section 4) and present our data and results (Section 5). We conclude with a discussion of results, their key implications and outline opportunities for future research (Section 6).

2. RELATED WORK

The relationship between news aggregators and content producers is the subject of a small but growing body of work.

Dellarocas et al. (2010) examine how the ability to place costless hyperlinks to third party content affects the behavior of competing content producers who can now choose between spending the effort to write an original article of a story and simply linking to an article that someone else has written. The authors view aggregators as a limiting case of content nodes who are inefficient in original content production and, therefore, can only attract readers by placing

links to interesting third-party content. The paper identifies several properties of aggregators and motivates some of the questions we ask in this paper. We discuss it in more detail in Section 3.

Chiou and Tucker (2011) offer an empirical contribution to the discourse about the net impact of aggregators. They empirically examine the effect of the removal of all hosted articles by The Associated Press from Google News at the end of 2009 (due to a dispute in licensing negotiations) on what sites consumers visited. They find that the removal of The Associated Press's content was correlated with a decline in subsequent visits to traditional news sites immediately after visiting Google News as compared to other news aggregators that continued to host The Associated Press content. The results suggest the presence of a complementary relationship between aggregators and content producers, whereby aggregators send traffic to content producers that would otherwise not be reaching those sites.

In another empirical paper with a similar objective, Athey and Mobius (2012) look at how the addition of a localization feature on Google News affects the consumption of local content. They find that the addition of this feature increases local news consumption, including the number of direct visits to such sites (that presumably users have discovered via Google News). However, the effect diminishes over time.

Hong (2011) focuses on the potential for aggregators to induce information cascades that concentrate traffic to a few "popular" sites. The author provides evidence of an association between the number of visitors to a news aggregator site and the online traffic concentration of that site. The author suggests design interventions for alleviating the adverse impact of such phenomena.

Our work also relates to the broader discourse on how readers allocate their attention in content networks. For example, Wu and Huberman (2008) analyze the role that popularity and novelty play in attracting the attention of users to dynamic websites. Agarwal et al. (2009) propose novel spatial-temporal models to estimate click-through rates in the context of content recommendation. Roos et al. (2011) propose an ambitious model of browsing behavior in

hyperlinked media that takes into consideration a user's utility and beliefs about the quality of cross-linked content.

Compared to this broader literature, our aims are more focused, looking specifically on how consumers allocate attention between news aggregators and news articles and how design parameters of the aggregator affect this allocation.

3. CONCEPTUAL FRAMEWORK

Aggregators have emerged as an increasingly important layer in the hyperlinked content ecosystems. Dellarocas et al. (2010) model aggregators as content network nodes who attract readers by placing links to interesting third-party content. They distinguish between the impact of aggregators on individual pieces of content and on competition among the related pieces of content. Accordingly, we develop hypotheses on both questions.

3.1 Aggregator's impact on individual articles

Dellarocas et al. (2010) show that the primary impact of an aggregator on the content ecosystems is the sum of two opposite effects:

On the one hand, a *search cost reduction* effect arises from the fact that aggregators generally place links to well-chosen content and provide some information (e.g. text snippet or photos) about this content that helps users decide whether it matches their interests. This effect is positive; it increases the overall consumption of content and benefits high quality content producers.

On the other hand, a *free riding* effect is primarily due to readers who browse through aggregator the headlines and snippets and never click through to the original articles. This effect is at the core of the controversy between aggregators and original content producers. The latter claim that, if aggregators did not exist, readers would be spending their time browsing the "front page" of

original content producer's sites. The free riding effect reduces the content producer's profits and incentives to produce quality content.

Dellarocas et al. (2010) identify the *aggregator click-through rate*, that is, the probability that an aggregator's user will click through to an article linked to through the aggregator, as a key determinant of whether the net impact of aggregators on a content ecosystem is positive or negative. However, they do not attempt to describe what this parameter depends on and consider it as a constant.

An important contribution of the current work is to elaborate on the factors that determine an article's aggregator click-through rate, as well as the amount of time that readers spend on the original article, conditional on clicking through.

News is an instance of what economists call *experience goods* (Varian 2001). Readers are, thus, unsure about the utility they will derive from reading a particular article until they have read it (in which case it is too late). In such situations, readers may be reluctant to spend the cognitive effort required unless they have some sufficiently strong signal that an article matches with their interest. This is the precise reason why information good industries make pervasive use of previews, reviews, and samples (Varian 2001). We claim that in the context of news aggregators, text snippets play the role of a preview. This, in turn, suggests that click-through rates have a non-monotonic relationship with article snippet lengths. Specifically, article summaries that are too short may not provide sufficient information that would allow users to decide whether an article is interesting enough; this may hurt their click-through rate. On the other hand, article summaries that are too long contain most of an article's salient information; this would make clicking-through redundant for many readers. Our reasoning leads to the following hypothesis:

Hypothesis 1: News aggregator article click-through rates exhibit an inverse U-shaped relationship with article abstract/snippet length.

Content producers care not only about the click-through rates, but also about the time readers actually spend on their sites. Drawing on the previous line of reasoning we hypothesize that

article snippet length will, similarly, have a non-monotonic relationship with the amount of time clicking readers spend on an article: When snippets are very short, readers do not receive sufficient information about an article. Many, thus, click through, only to discover that the article was not as relevant or interesting as they thought, in which case they quickly move away. On the other hand, when snippets are very long, readers may find that full articles do not contain much additional information relative to what was already seen on the snippet. We expect that this will, similarly, reduce the time that readers spend on the article itself. The following hypothesis ensues:

Hypothesis 2: Article reading times (conditional on clicking) exhibit an inverse U-shaped relationship with the abstract/snippet length provided by the news aggregator.

3.2 Aggregator's impact on competition among articles

Aggregators often collect several articles written about the same story and display them next to each other. This increases the competition among articles and complicates the user's attention allocation problem because the question now is not only whether to click-through but also which article they should select. This additional effect of aggregators has, so far, received little attention and is, therefore important to explore further.

An interesting question is whether the presence of multiple related articles increases or decreases a user's propensity to click on *any* of these articles. Both possibilities are plausible. On the one hand, one can argue that the more articles are available on a story, the more likely that at a user will find at least one of them appealing. Furthermore, the presence of multiple articles tends to signal important stories that are worth reading about. These arguments lead to the following hypothesis:

Hypothesis 3a: The presence of multiple articles about the same story increases the group click-through rate.

On the other hand, the combined presence of multiple article snippets provides more information about the story itself and may satisfy the curiosity of readers who will then decide it is not necessary to read the original article(s). This leads to the alternative hypothesis:

Hypothesis 3b: The presence of multiple articles about the same story decreases the group click-through rate.

Another important question is what factors make users choose among the competing articles. It is well documented here that the link's position matters a lot. The higher the article is on the list, the higher the probability that it will be chosen (see for example Ghose and Yang 2009). What has not been researched is the impact, if any, of an article's snippet length and picture on the choice probability. To construct hypotheses around this we draw upon theories of attention from cognitive sociology (Quinlan and Dyson 2008). Such theories are based on the premise that attention is necessary because our central mechanisms cannot cope with the amount of sensory stimulation present at any given time (Quinlan and Dyson 2008). There are two general theories of attention: *early selection* theory and *late selection* theory. According to the original early selection theory, all stimuli reaching the sensory system are processed to the point at which certain physical attributes are analyzed and one single stimulus is selected (Broadbent 1958). Physical attributes of a snippet include images accompanying the snippet, the length of the snippet and the position of the snippet. Selection occurs early in the stream of processing, and a stimulus' (snippet's) probability of being selected increases directly as a function of its physical intensity or spotlights (Broadbent 1958; Quinlan and Dyson 2008). For example, having an accompanying image while other snippets do not have accompanying image may increase a snippet's likelihood of being selected (clicked through). Another example, being the longest snippet may attract attention and increase its click through rate. This line of reasoning leads to the following hypothesis:

Hypothesis 4a: When several articles about the same story compete for user attention, ceteris paribus, readers are more likely to click on articles that are higher up, have photos and whose snippet lengths are longer than the average group snippet length.

Contrary to the above theory, the late selection theory put forward the idea of selecting based on semantic descriptions instead of merely physical descriptions. According to the late selection theory, selection takes place at a much later stage of processing and involves much more complex categories than physical stimulus properties, i.e., selection does not take place until every stimulus has undergone perceptual or semantic analysis (Deutsch and Deutsch 1963; Duncan 1980; Kahneman 1973; Quinlan and Dyson 2008). In the context of our study, the late selection theory implies that the snippets of a particular topic will all be semantically analyzed before one of them is selected (clicked through). Thus instead of being selected based on the intensity of their physical characteristics, the snippets attract attention and in turn are being selected based on their content. In such cases, the cost-benefits reasoning we used to derive out hypotheses in Section 3.1 would apply. We posit that: (a) snippet lengths that are shorter than the average group snippet length are more likely to leave readers with more uncertainty about their quality and relevance, whereas (b) snippet lengths that are longer than the average group snippet length are more likely to convey most of the article's salient information and, thus, make more readers feel that clicking through to the original article would be redundant. This line of reasoning leads to the following hypothesis:

Hypothesis 4b: When several articles about the same story compete for user attention, ceteris paribus, readers are less likely to click on articles whose snippet lengths deviate from the average group snippet length.

4. EXPERIMENTAL SETTING

Our field experiments are conducted using the *Newscron* (originally called *Nüwsli*) Switzerland news aggregator, developed by some of the authors at the Swiss Federal Institute of Technology (ETH) in Zurich. The front-end of the application is a client that can be freely downloaded and

installed on iPhones from the Apple's App Store. The back-end of the application is a server that collects and stores news articles that can be accessed by the client. The server collects daily news articles from Switzerland's major newspapers (in all three national languages: German, French, and Italian) and stores them in a database on daily basis. Often, several newspapers publish the exact same article (syndication). The server identifies and ignores copies of the same article. The server performs a semantic analysis of article texts to group them together into topics (stories). This leads to the following data structure: every article belongs to a topic; a topic contains one or more articles. Our algorithm ensures that all collected articles have the minimum required snippet length (i.e., at least 343 characters).

User interaction with Newscron is designed as a three step process (see Figures 2, 3, 4). First, the user is presented with a list of topics/stories. When the user clicks on a topic, he/she sees all articles belonging to the particular topic, which are sorted by their publication dates. Only an outline (title, snippet, and (if available) picture) of each article is displayed. At this step, we can assume that the user is already interested in a particular topic. This second step thus has the advantage of removing topic-specific interest variable from the equation and having the click-through only influenced by variables such as the snippet length, the existence of image, and the number of articles in a particular topic. Finally, by clicking on the article's dedicated and labeled button at the bottom of the article's outline, the user is directed to the newspaper's website for full article read.

The application offers the user three methods of browsing and discovering topics. The default method is through the list of featured news (see Figure 2). Featured news contains the most read topics of the day. Alternatively users can browse the latest topics by category, by publisher, or by publication date. In addition the application keeps track of the previously read topics, which can easily be retrieved by the user.

Newscron tracks users' interactions with the software by logging each transaction on the back-end server. The logs allow us to analyze the users' behaviors.

Figure 2 First level - Topics

Figure 3 Second level – Article outline of a topic

Figure 4 Third level -Full article

98 characters (-60%)	147 characters (-40%)	196 characters (-20%)
		
245 characters (default)	294 characters (+20%)	343 characters (+40%)
		

Table 2 Preview of the Different Snippet Lengths (% Change)

To test our hypotheses, we manipulated the articles' snippet length at the back-end of the Newscron application (at the article-level). The default snippet length used in our app is equal to 245 characters, which is the average number of characters of snippets in Google News. We reduced/increased this default snippet length in increments of 20%, which is 2*the standard deviation of snippets in Google News. We thus defined 6 different snippet lengths ranging from -60% to +40% of the default length (Table 2). We chose -60% because it is the shortest length that is supported by the user interface and +40% because it is the longest snippet possible subject to copyright agreements we have with the news providers.

During our experiment the snippet length that was displayed when user *i* accessed article *j* was randomized. This means that different users might encounter the same article with different snippet lengths. Furthermore, different articles within the same topic group could be displayed with different snippet lengths.

This field experiment lasted for 2 weeks during which we had 2,163 users, 4,981 unique articles, and 26,559 topic-level access events that resulted in 17,654 article click-through events.

Table **3** depicts the salient statistics of our experimental dataset.

Measurement	**Value**
Unique Articles	5,312
Unique Users	2,163
Total topic access events	26,559
Total article click-through events	17,654
Average number of articles accessed by users	8.46
Average reading time (in seconds)	78
Average decision time (in seconds)	11

Table 3 Parameters of the dataset

5. RESULTS

5.1 Baseline effects

An important objective of the field experiment is to analyze the impact of abstract/snippet lengths on article click-through rates, article reading time (conditional on clicking), and article decision time (conditional on clicking). To control for the other characteristics of news articles besides the snippet lengths, we added several control variables into the analysis such as the full article length (in characters), the existence of an image that accompanied the snippet (0 or 1), the topic category (0: International, 1: Local, 2: Business, 3: Technology, 4: Entertainment, 5: Sports, 6: Life, 7: Motors, 8: Culture), and the article publisher (1 to 15). Table 4 summarizes our independent, dependent, and control variables.

We used logistic regression to analyze (binary) click-through rates and Poisson regression to analyze decision time and reading time, both of which can be thought of as count variables.[4] To account for idiosyncratic behavior at the user and article level, we used appropriate random effects. The results of our regressions are summarized in Tables 5, 6 and 7.

[4] Poisson regression is justified in this case because our data set does not exhibit overdispersion. Specifically, the mean and variance of reading times (78 and 80 seconds respectively) and decision times (11 and 12 seconds respectively) are comparable, thus fulfilling the conditions for applying Poisson regression.

Variable	Description
Dependent Variables	
CLICK-THROUGH	Binary variable; whether an article was opened in the browser to be read in full length
DECISION-TIME	The time needed by a user to decide whether to CLICK-THROUGH
READING-TIME	The time needed by a user to read the full article on the publisher's webpage
Independent Variables (for logistic and Poisson regressions)	
SNIPPET-98	News snippet shown along with the title, having a length of 98 characters
SNIPPET-147	As above; 147 characters
SNIPPET-196	As above; 196 characters
SNIPPET-245	As above; 245 characters
SNIPPET-294	As above; 294 characters
SNIPPET-343	As above; 343 characters
Independent Variables (for multinomial logit regression)	
SNIPPET-1	News snippet shown along with the title, having a length lower than the average length of the snippets of all related-articles
SNIPPET0	As above; length equal to the average length
SNIPPET+1	As above; length higher than the average length
POSITION-TOP	Article is on the top position of the list of related-articles
POSITION-SECOND	Article is at the second position of the list of related-articles
POSITION-LOW	Article is at the third or lower position of the list of related-articles
Control Variables	
HAS-IMAGE	Binary variable; whether an image is displayed together with the title and snippet
ARTICLE-LENGTH	The length in characters of the full article on the publisher's website
RELATED-ARTICLES	Number of articles of the same topic that were displayed at the same time in form of a list
CATEGORY	The category to which the article belongs (see explanation above)
PUBLISHER	The publisher of the news

Table 4 Variables Summary

Variables	**Coef.**	**Std. error**	**z**
Independent Variables			
SNIPPET-98	baseline		
SNIPPET-147	-.285	.055	-5.18***
SNIPPET-196	-.478	.054	-8.81***
SNIPPET-245	-.654	.054	-12.21***
SNIPPET-294	-.796	.054	-14.81***
SNIPPET-343	-.838	.053	-15.70***
Control Variables			
HAS-IMAGE	-.334	.062	-5.36***
CATEGORY	Estimated but not shown		
PUBLISHER	Estimated but not shown		
Wald Chi-square(27): 462.73; Prob>Chi-square: 0.000 *-p<0.05; ** - p < 0.01; *** - p < 0.001;			

Table 5 Click-Through Rates Result

Variables	**Coef.**	**Std. error**	**z**
Independent Variables			
SNIPPET-98	baseline		
SNIPPET-147	.249	.010	24.29***
SNIPPET-196	.415	.020	41.36***
SNIPPET-245	.537	.010	54.48***
SNIPPET-294	.667	.010	67.59***
SNIPPET-343	.756	.010	77.98***
Control Variables			
HAS-IMAGE	.246	.027	9.06***
RELATED-ARTICLES	.184	.005	37.98***
CATEGORY	Estimated but not shown		
PUBLISHER	Estimated but not shown		
Wald Chi-square(28): 9850.26; Prob>Chi-square: 0.000 *-p<0.05; ** - p < 0.01; *** - p < 0.001;			

Table 6 Decision Time Rates Result

Click-through rates. Focusing only on topics that contain a single article, we find (Table 5) that click-through rates monotonically decrease with snippet length, i.e. longer snippets are associated with lower click-through rates. The presence of an accompanying image also reduces the click-through rate.

Decision time. Conditional on clicking-through, the time spent on the aggregator (reading snippets etc.) until a user decides to click on the linked article monotonically increases with the snippet length (Table 6). It also has a positive association with the presence of an image and the number of related articles simultaneously present for the same topic.

Reading time. Conditional on clicking-through, reading time exhibits an inverse U-shaped relationship with snippet length: reading time increases when snippet length increases from 98 to 294 characters and decreases when snippet length reaches 343 characters (Table 7). As expected, reading time increases with the length of the article

Variables	**Coef.**	**Std. error**	**z**
Independent Variables			
SNIPPET-98	baseline		
SNIPPET-147	.050	.004	13.37***
SNIPPET-196	.037	.004	8.95***
SNIPPET-245	.059	.004	15.83***
SNIPPET-294	.077	.004	20.22***
SNIPPET-343	.050	.004	13.01***
Control Variables			
HAS-IMAGE	.010	.038	.27
ARTICLE-LENGTH	2.2e4	9.41e-6	7.69***
RELATED-ARTICLES	.033	.002	13.41***
CATEGORY	Estimated but not shown		
PUBLISHER	Estimated but not shown		
Wald Chi-square(28): 2071.16; Prob>Chi-square: 0.000 *-p<0.05; ** - p < 0.01; *** - p < 0.001;			

Table 7 Reading Time Rates Result

The number of related articles of the same topic has a positive association with reading time. We tentatively explain this by postulating that the number of related articles is a proxy for a topic's importance and that readers spend more time on articles of more important topics.

5.2 Competition effects

As discussed in Section 4 an important function of aggregators is to collect related articles on the same topic. Since the aggregator displays a snippet for each article, two pertinent questions arise: (a) how the presence of multiple articles affects the probability that any article in the group is clicked and (b) what factors determine which article is chosen.

To answer the first question we collapsed each article group (topic) access into a single record and performed logistic regression. Our dependent variable recorded whether at least one article within that group was clicked by the user. The results are shown in Table 8. The number of related articles has an inverse U-shaped relationship with the group click through rates. Click-through rates attained their maximum value when a topic had around 4 related articles.

Variables	**Coef.**	**Std. error**	**z**
Independent Variables			
RELATED-ARTICLES	.567	.065	8.72***
RELATED-ARTICLES^2	-.071	.010	-6.68***
Control Variables			
HAS-IMAGE	.143	0.037	-3.83***
Wald Chi-square(3): 90.71; Prob>Chi-square: 0.000 *-p<0.05; ** - p < 0.01; *** - p < 0.001;			

Table 8 Topic Click-through Results

To answer the second question we performed discrete choice analysis. Specifically, we looked at article groups where one or more linked articles were clicked on and used a multinomial logit regression model to identify which covariates of each choice alternative had a statistically significant association with the probability that the alternative in question would be clicked on by

the user. For this analysis, we replaced raw snippet lengths with a variable that indicated whether an article's snippet was longer than (+1), equal to (0) or shorter than (-1) the group's average snippet length. We also added a control variable that indicated an article's position in the group (0: top position, 1: second position, 2: third or lower position).

The regression results are shown in Table 9. As expected, an article's position has an important effect on it being chosen, with the topmost article being chosen most often. Interestingly, an article's snippet length relative to the group's average snippet length was also found to be a significant predictor. The baseline case corresponds to the case where the snippet length is shorter than the average (-1). Compared to the baseline case, we notice that having longer than average snippets (+1) has a statistically significant positive effect on the choice probability. Moreover, the presence of an image increases an article's within-group choice probability. This finding is interesting and should be contrasted with the fact that the presence of an image is associated with a decrease in an article's absolute click-through rate when there is no related article.

Overall, the above findings provide support for Hypothesis 4a and suggest that the article selection process is consistent with the predictions of early selection theories from cognitive psychology.

Variables	**Coef.**	**Std. error**	**z**
SNIPPET-1	baseline		
SNIPPET0	-.156	.480	-.324
SNIPPET+1	.290	.113	2.57*
POSITION-TOP	baseline		
POSITION-SECOND	-2.07	.113	-18.30***
POSITION-LOW	-2.02	.182	-11.09***
ARTICLE-LENGTH	3.25e-4	6.35e-5	5.12***
HAS-IMAGE	2.06	.262	7.88***
Log-Likelihood: -333.81 *-p<0.05; ** - p < 0.01; *** - p < 0.001;			

Table 9 Article Choice Analysis Result

6. DISCUSSION, IMPLICATIONS, AND ONGOING RESEARCH

News aggregators have emerged as an important component of digital content ecosystems. A better understanding of how news aggregator design parameters affect the allocation of reader attention between aggregators and original articles is useful, both in terms of informing aggregator design, as well as in terms of informing the current controversy that exists between the two parties. In this study, we conducted field experiments with that objective in mind.

Our results indicate that there is, indeed, a degree of substitution between the amount of information of a news article that is displayed on news aggregators and the cumulative time that readers are likely to spend on the original article site. We found a negative relationship between an article's snippet length on the aggregator and the probability that a user will click the link to the original article site: the longer the snippet, the lower the click-through rate. Moreover, we found a positive relationship between an article's snippet length and the amount of time readers spend on the aggregator until they decide to click on the linked article. Interestingly, the presence of an image has the same effect to that of increasing the snippet length on the article's click-through rate: it is associated with a decrease in click-through rate and an increase in a reader's average decision time.

Moderating the above results, we found that when there is a click-through, the amount of time spent on the original article has an inverted U relationship with the snippet lengths. This finding suggests that very short snippets do not provide adequate information, resulting in more readers clicking on their respective linked articles but then deciding that the articles were actually not very interesting to them.

Aggregators typically group together articles that refer to the same story, thus increasing competition among related articles. In this study, we also examined how the aggregation of articles into topic groups affects the allocation of readers' attention. We found an inverse U-shaped relationship between the number of articles in a topic group and the probability that readers will click on at least one article from that group. Our tentative explanation for this finding is that, the more articles are available on a topic, the more likely that at a user will find at

least one of them appealing. Furthermore, the presence of multiple articles tends to signal important stories that are worth reading about. However, when there are plenty of related articles, the combined presence of multiple snippets may satisfy the readers' curiosity who then may not feel the need to click on any of the linked articles. This is a previously unnoticed side effect of news aggregators that can be potentially detrimental to content producers and thus deserves more attention.

Recognizing its potential practical significance, in our ongoing work we plan to delve deeper into this result. Specifically, we will perform text analysis of snippets of related articles to determine the extent to which these snippets offer complementary information on a story. We will score the snippets of each article group along this dimension and will use this score as an independent variable in our regressions.[5]

With respect to the competition among the related articles, we examined what factors determine which article(s) in a group will likely to be chosen by the readers. As expected, articles positioned at the top of the list were most often being chosen. Controlling for the position, articles with an image were more likely to be chosen. Interestingly, the choice probability was higher for articles whose snippet length was longer than the average snippet length of the related articles.

The above set of findings suggest that, when faced with the choice among competing articles, on average readers behave according to the predictions of early selection theories from cognitive psychology: they choose based on external physical characteristics, such as position, presence of an image or a large abstract, rather than based on an exhaustive analysis of snippet content.

In our ongoing research we will delve deeper into understanding this behavior. According to perceptual load theory (Lavie 1995, 2001; Lavie and Tsal 1994), depending on the user's

[5] Paragraphs in italics describe ongoing research that we expect will be partially or fully completed before July. If the paper is accepted the results of this additional work will be reported in the workshop.

perceptual load, he/she may perform early selection by means of the physical attributes of the snippets, such as the position and presence of accompanying image (when perceptual load is high), or late selection by semantically analyzing the content of each snippet (when perceptual load is low). We will examine whether users of our app exhibit this range of behaviors. The time when the application is launched until it is closed is marked as one session. For each session, we will take note the decision time, which is computed as the time duration from selecting a topic to selecting a snippet. Decision time will be our proxy measure of perceptual load. The sessions with decision time below that of the average decision time represent high perceptual load; whereas the sessions with decision time above that of the average decision time represent low perceptual load. In a situation where the perceptual load is high, we predict that physical features such as the position and accompanying image will be more significant predictors of the snippets selection. When the perceptual load is low, we predict that the content of individual snippets, as well as complementarities in the information contained in neighboring snippets, might emerge as significant predictors.

Taken collectively, our results produce a nuanced picture of the impact of article snippet length on its click-through rate. Longer snippets generally reduce click-through rates. However it is important to highlight that very short snippets seem to generate a lot of "spurious" clicks, where readers "leave" the original article shortly after "their arrival" because they didn't find it as interesting as they expected.

As with any study, the findings in this study should be viewed with regard to the study limitations. Our aggregator has multi-level interface (see Figures 2, 3, and 4). Our experimental treatments were performed at the second level (topic level) of the aggregator's interface (see Figure 3). Our data thus showed how varying snippet lengths affected the probability of transitioning from the second to the third level. We observed that, once a reader clicks to the second level, he/she has already indicated his/her interest for the topic. Conditional on being interested in the topic, the decision to click a linked article is primarily based on the reader's assessment of how much more useful information the original article contains relative to the snippet. It will be interesting to conduct another experiment at an aggregator's top level (see

Figure 2) where the different topics of the day are aggregated. For technical reasons, such manipulations are not possible in the iPhone version of Newscron. *Currently we are in the process of conducting another experiment using the iPad version of the app that will allow us to manipulate at the top (topic) level.*

In conclusion, we would like to emphasize that, although this work offers novel insight into the relationship between news aggregators and content producers, its objective is not to provide answers to the question of whether aggregators are, on balance, beneficial or harmful to the content ecosystems. What the current work establishes is that aggregators indeed extract an "attention tax" from content producers, in the form of users who never click through to the original articles. We demonstrate that the fraction of such users depends on the design parameters of the aggregator and that there is a substitution relationship between the amount of time that readers spend on the aggregator and the time readers spend on the original articles. We further show that competitive pressure might limit the extent to which individual content producers deviate from an aggregator's norms: our multinomial logit results suggest that a publisher's unilateral decision to shorten the snippet lengths of its articles might put them at disadvantage in situations where there are several related articles on the same topics. What is outside the scope of the current research is the impact that aggregators have in increasing the overall consumption of content (e.g. because they reduce search costs by organizing content). Despite recent attempts to provide partial answers (Chiou and Tucker 2011; Athey and Mobius 2012), the latter still remains an elusive and interesting empirical question for future research.

From a methodological perspective, this work highlights the feasibility of conducting field experiments with "real users" using apps developed in research labs and then released to the public. Our results suggest that experiments with even a few thousands of users can expose many of the effects that are also present in much larger scale applications. There is, thus, an interesting methodological discussion to be had on the merits of working with larger, but less flexible, secondary data sets obtained from third-parties vs. with primary data sets obtained from smaller scale apps developed for the explicit purpose of conducting experimental studies.

REFERENCES

Agarwal, D., Chen B-C. Elango P. 2009. Spatio-temporal models for estimating click-through rate. Proceedings of the 18th International Conference on World Wide Web (WWW'09), 21-30.

Athey, S. and Mobius, M. (2012) The Impact of News Aggregators on Internet News Consumption: The Case of Localization. Working Paper, February 24, 2012.

Broadbent, D. (1958). Perception and Communication, Pergamon Press.

Chiou, Lesley and Tucker, Catherine, Copyright, Digitization, and Aggregation (December 13, 2011). NET Institute Working Paper No. 11-18. Available at SSRN: http://ssrn.com/abstract=1864203 or doi:10.2139/ssrn.1864203

Dellarocas, Chrysanthos N., Katona, Zsolt and Rand, William M., Media, Aggregators and the Link Economy: Strategic Hyperlink Formation in Content Networks (September 30, 2010). NET Institute Working Paper No. 10-13; Boston U. School of Management Research Paper No. 2010-30; Robert H. Smith School Research Paper No. RHS 06-131. Available at SSRN: http://ssrn.com/abstract=1694370 or http://dx.doi.org/10.2139/ssrn.1694370

Deutsch, J. and Deutsch, D. (1963). "Attention: Some theoretical considerations", Psychological Review, vol. 70, pp. 80-90.

Dewan, R. M., Freimer, M. L., Seidmann, A. and Zhang, J. (2004) Web portals: Evidence and analysis of media concentration. J. Mgmt Info. Sys., 21, 2, 2004, 181-199.

Duncan, J. (1980). "The locus of interference in the perception of simultaneous stimuli", Psychological Review, vol. 87, pp. 272-300.

Ghose, A. and Yang, S. An Empirical Analysis of Search Engine Advertising: Sponsored Search in Electronic Markets. Management Science October 2009 vol. 55 no. 10 1605-1622.

Hong, Sounman, A Simple Model of News Aggregators, Information Cascades, and Online Traffic (September 13, 2011). Available at SSRN: http://ssrn.com/abstract=1927035

Isbell, K. (2010). The Rise of the News Aggregator: Legal Implications and Best Practices. Berkman Center, Research Publication No. 2010-10.

Jarvis, J. (2008) The link economy vs. the content economy. Buzzmachine blog, June 18, 2008. (http://www.buzzmachine.com)

Kahneman, D. (1973) Attention and effort, Prentice-Hall.

Karp, S. (2007) The Web's Link-Driven Attention Economy. Publishing 2.0, December 15, 2007. (http://publishing2.com/)

Lavie, N. (1995) "Perceptual load as a necessary condition for selective attention", Journal of Experimental Psychology: Human Perception and Performance, vol. 21, pp. 451–468.

Lavie, N. (2001) "The role of capacity limits in selective attention: Behavioral evidence and implications for neural activity", In J. Braun and C. Koch (Eds.), Visual attention and cortical circuits, MIT Press, pp. 49–68.

Lavie, N. and Tsal, Y. (1994) "Perceptual load as a major determinant of the locus of selection in visual attention", Perception and Psychophysics, vol. 56, pp. 183–197.

Quinlan, P. and Dyson, B. (2008) "Chapter 8: Attention", Cognitive Psychology, Prentice Hall, pp. 273-311.

Roos, J., Mela, C. F., Shachar, R. 2011. Hyper-Media Search and Consumption. Working Paper, Duke University.

Varian, Hal R. 2001. "Markets for Information Goods," in Monetary Policy in a World of Knowledge-Based Growth, Quality Change and Uncertain Measurement. Kunio Okina Tetsuya Inoue, ed. Macmillan Press.

Wu, F. and Huberman, B. 2008. Popularity, Novelty and Attention, Proceedings of the 9th ACM Conference on Electronic Commerce (EC'08), 240-245.