# The Cost of Annoying Ads

Daniel G. Goldstein
Microsoft Research, NYC
102 Madison Ave., 12th Floor
New York, NY 10016
dgg@microsoft.com

R. Preston McAfee
Google Strategic Technologies
1600 Amphitheatre Parkway
Mountain View, CA 94043
preston@mcafee.cc

Siddharth Suri
Microsoft Research, NYC
102 Madison Ave., 12th Floor
New York, NY 10016
suri@microsoft.com

## ABSTRACT

Display advertisements vary in the extent to which they annoy users. While publishers know the payment they receive to run annoying ads, little is known about the cost such ads incur due to user abandonment. We conducted a two-experiment investigation to analyze ad features that relate to annoyingness and to put a monetary value on the cost of annoying ads. The first experiment asked users to rate and comment on a large number of ads taken from the Web. This allowed us to establish sets of annoying and innocuous ads for use in the second experiment, in which users were given the opportunity to categorize emails for a per-message wage and quit at any time. Participants were randomly assigned to one of three different pay rates and also randomly assigned to categorize the emails in the presence of no ads, annoying ads, or innocuous ads. Since each email categorization constituted an impression, this design, inspired by Toomim et al. [18], allowed us to determine how much more one must pay a person to generate the same number of impressions in the presence of annoying ads compared to no ads or innocuous ads. We conclude by proposing a theoretical model which relates ad quality to publisher market share, illustrating how our empirical findings could affect the economics of Internet advertising.

## Categories and Subject Descriptors

J.4 [**Social and Behavioral Sciences**]: Economics

## General Terms

Economics, Experimentation

## Keywords

display, advertising, quality, compensating differential

## 1. INTRODUCTION

Display advertising is the prevalent way for publishers to monetize content on the Web. Publishers receive payment from advertisers for placing ads near their content or in their applications. Payments can be determined by a contract or a real-time auction. In either arrangement, publishers are typically paid by the number of impressions they can deliver. Thus, they have an incentive to attract and retain users with valuable content, experiences, and applications, and have a disincentive to lose users due to annoyances.

Display ads vary in the extent to which they annoy users. Annoying ads are a source of tension for publishers since they both make money, through payments from advertisers, and cost money, through a decrease in page views due to users abandoning the site. This tension has led to conflict within publishing organizations between salespeople, who have an incentive in the form of commission to sell any ads, and management, who are concerned with long-term growth of users and traffic. The continued long-term display of annoying ads may exert negative effects on the publisher, the user, and the advertiser, which we discuss in turn.

First, annoying ads can exert negative effects on publishers. Apart from the user abandonment effects we investigate in this paper, annoying ads might signal that the website on which the ad is placed on lacks stability ("Why should I trust my email to a site that is so desperate for cash it accepts ads of such poor quality?"), reputability ("Why should I trust the objectivity of a site that is so in the pocket of advertisers it won't refuse any of them?"), or safety ("Why would I trust this publisher to protect me from phishing attacks, scams, malware, etc. if they are so indiscriminate about who they let advertise?").

Second, annoying ads can exert a negative impact on users. Ads with excessive animation can get in the way of the user consuming the publisher's content, undermining the very reason that brought them to the site. In what follows, we document users reporting that annoying ads distract them. Furthermore, we provide experimental evidence that annoying ads impair accuracy on a cognitive task.

Finally, annoying ads may harm the advertiser that created them. As will be shown, annoying ads are often characterized by exaggerated attempts to capture visual attention such as through fast-moving animation or bizarre imagery. While these manipulations do capture attention, they may also signal that the advertiser is desperate for business or low on resources, undermining the classical signal of quality that advertising is theorized to bring [15]. Furthermore, experiments have shown that too much animation can result in lower ad recognition rates compared to ads with moderate or no animation [21, 2]. In these ways, annoying ads may actually lower brand reputation and recall, two metrics advertisers typically strive to increase.

If annoying ads exhibit so many negative effects for publishers, users and advertisers, one may wonder why a publisher would run annoying ads at all. The answer may be that it is has been historically difficult to measure the mon-


*WWW*, 2013 Rio de Janeiro, Rio de Janeiro Brazil
ACM 978-1-4503-2035-1/13/05.

etary cost of annoying ads. The first and main contribution of this work is that we measure the compensating wage differential of annoying ads. That is, we measure how much more one must pay a user to do the same amount of work in the presence of annoying ads compared to innocuous ads or no ads. The compensating differential is important to measure because it captures some of the negative effects of advertising, which publishers need to heed as a lower bound when setting the price to run an ad.

In a two-experiment investigation, we compute the compensating differential for annoying ads. In the first experiment users randomly rated either an animated ad or its static counterpart. This design shows that animation has a negative impact on user ratings. For those ads that users rate as annoying we ask them to explain their thinking. An analysis of the ratings and comments yields a better of understanding of what users find annoying about these ads. This analysis will also exhibit how annoying ads negatively affect user perceptions of advertisers. These analyses are additional contributions of this work.

In the second experiment, we use those ads identified as more or less annoying, along with the recent methodological innovation of Toomim et al. [18], to estimate the pay rate increase necessary to generate an equal number of page views in the presence of annoying ads, compared to innocuous ads or no ads. This estimate is the cost of annoying ads in our experiment. We chose categorizing emails as the task to proxy for using a publisher's site because users either implicitly or explicitly need to categorize their emails as spam or not spam in the presence of ads when using free web-based email services such as Yahoo! Mail, GMail, and Mail.com. Finally, we provide a theoretical model of how our empirical findings could affect the display advertising industry, which is the third contribution of this work.

## 2. RELATED WORK

As mentioned in Section 1, we use the methodological innovation of Toomim et al. [18] for computing compensating differentials. Toomim et al. conducted a Mechanical Turk experiment in which participants randomly experienced an easy, medium, or hard version of a task at a randomly assigned pay rate. This allowed the authors to compute how much more one would have to pay a worker to do the hard task over the medium and easy tasks. The authors also exhibited this technique in an experiment in which participants were randomly assigned to use either an "ugly" or a "pretty" interface to do a task. We will use this technique to isolate the effect of the ad quality on user abandonment. Next, we describe prior experimental work which studies the impact of ad quality on behavior.

Dreze and Hussherr [4] conducted an experiment on the effectiveness of display advertisements using eye-tracking technology. Their conclusion, that users rarely focus directly on banner ads, is often referred to as banner blindness, a term coined by Benway [1]. Burke et al. [2] had participants perform visual search tasks in the presence of no ads, a static display ad, or an animated display ad. They found that ads did reduce search time, however, there was no significant difference between animated and static ads. Perhaps even more surprisingly, they did a *post hoc* test which found that animated ads were remembered less frequently than static ads.

Yoo and Kim [21] asked a similar research question. They conducted a larger-scale laboratory experiment in which participants were randomly exposed to web pages with ads with no animation, slow-moving animation or fast-moving animation. They found that more animation did increase attention to ads. Moreover, moderate animation increased ad recognition rates and brand attitudes. Highly animated ads, however, *decreased* recognition rates and brand attitudes. This result complements the results of Burke et al. [2]. Yoo and Kim [21] conclude that, "Web advertisers should be aware of the possibility that excessive animation can backfire against the original intention of effective communication."

Goldfarb and Tucker [5] conducted a field experiment in which they found that ads that matched the site's content or ads that were intrusive increased participant's self-reported intent to purchase. However, ads that were both intrusive and matched the website's content *reduced* intent to purchase. Ads were considered intrusive if, for example, they produced a popup window, took over the whole screen, played music, or obscured the web page text. The authors suggest that the reason for this interaction effect is that users are more sensitive to targeted and intrusive ads when the product advertised is privacy sensitive. In the context of sponsored search, Buscher et al. [3] found that ads that are relevant to the search terms received more visual attention than ads that were less relevant. This complements the results of Goldfarb and Tucker [5] which were found in the domain of display advertising.

Taken as a whole, these studies suggest there may be benefits to a small degree of animation or intrusiveness in advertising, but that too much animation or intrusiveness can have a detrimental impact on the ad effectiveness.

## 3. RATING THE QUALITY OF ADS

We next describe our experiments, both of which were conducted on Amazon's Mechanical Turk[1], an online labor market. Since it was originally built for jobs that are difficult for computers but are easy for humans (e.g., image recognition), jobs on Mechanical Turk are called Human Intelligence Tasks or HITs. There are two types of people on Mechanical Turk: requesters and workers. Requesters can post HITs and workers can choose which HITs to do for pay. After a worker submits a HIT, the requester can either accept or reject the work based on its quality. The fraction of HITs that a worker submits which are accepted is that worker's approval rating. This functions as a reputation mechanism. The Amazon API gives each worker account a unique, anonymous identifier. By tracking the IDs of the workers who accepted our HITs, we could enforce that participants were only allowed to participate in one of the two experiments, and they were only allowed to do that experiment one time.

There is a burgeoning literature on conducting behavioral experiments on Mechanical Turk [12, 11, 16, 6, 7, 20, 13, 9, 17]. In this setting, the experimenter takes on the role of the requester and the workers are the paid participants of the experiment. Mason and Suri [10] provide a how-to guide for conducting behavioral experiments on Mechanical Turk. We now describe the design and results of our first experiment, which served to identify sets of more and less annoying ads

[1] `http://www.mturk.com`

(henceforth "bad ads" and "good ads" for brevity) for use in the second experiment.

## 3.1 Method

The goal of this experiment is to rank a set of actual display ads in terms of annoyingness and to collect reasons why people find certain ads annoying. The preview page of the HIT explained that workers would first browse through all of the ads in the experiment and then rate them one by one. After accepting the HIT, participants were shown 9 pages of ads, with 4 ads per page, and instructions to take a quick look at each page. This was done to familiarize respondents with the items they would ultimately rate to reduce random order effects in their later use of a rating scale. This was also done so that participants could calibrate their use of the rating scale [14]. Next, users were shown each ad individually, again in random order, and asked to rate each ad on a 5-point scale with the following levels:

1. Much less annoying than the average ad in this experiment
2. A bit less annoying than the average ad in this experiment
3. Average for this experiment
4. A bit more annoying than the average ad in this experiment
5. Much more annoying than the average ad in this experiment

After rating every ad on this scale, participants were then shown only the ads they rated as annoying (i.e., "A bit more annoying than the average ad in this experiment" or "Much more annoying than the average ad in this experiment") and asked to write a few words as to why they found the ad annoying.

We used a pool of 144 ads, 72 of which were skyscrapers (either 120 or 160 pixels wide by 600 pixels high) and 72 of which were medium rectangles (300 pixels wide by 250 pixels high). The ads primarily came from Adverlicious[2], an online display advertising archive. A worker was randomly assigned to either see all skyscrapers or all medium rectangles. The 144 ads used in the experiment were created from 72 animated ads, from which we created an additional 72 static variants by taking a screenshot of each animated ad's final frame. The animation in display ads typically lasts only a few seconds and then settles into a final frame which serves as a static ad for the rest of the time the ad is in view. The final frame is usually a good representation of the ad since it often shows the advertiser name with the same general design, creative, and color scheme as the animated part. This technique, which resulted in 72 animated/static pairs, allowed us to study the effect of animation on annoyance, holding all other properties of the ads constant. For each ad pair, workers were randomly assigned to see either the static or animated variant. Since this random assignment was done per pair, each participant saw a mixture of both animated and static ads.

We paid each worker a $0.25 flat rate and a bonus of $0.02 per ad rated. Since we had users input free text, we used the Amazon API to restrict to U.S. workers to help ensure a good grasp of the English language. We also used the Amazon API to require workers have at least a 95% approval rating.

[2] `http://adverlicio.us`

| Category | Example Words | Count |
|---|---|---|
| Animation | move, motion, animate | 771 |
| Attentional Impact | annoy, distract, attention | 558 |
| Aesthetics | ugly, busy, loud, cheap | 435 |
| Reputation | scam, spam, fake | 122 |
| Logic | sense, weird, stupid | 107 |

**Table 1: Categorization of words found in the participants' comments on the ads they found annoying**

## 3.2 Results

The experiment ran for 18 hours and collected responses from 163 participants. We excluded participants who skipped more than one question, leaving 141 participants. Though this exclusion makes little difference in the results, we felt it best to only compare the ratings of people who rated a similar number of items, as ratings may change as a function of the number of items previously rated. The distribution was rather symmetric with an average rating of 2.9 on the five point scale. Since a rating of 3 corresponds to "Average for this experiment", the participants' ratings were well calibrated to our instructions.

Recall that we started with 72 ads and created static variants of each, resulting in the 144 ads in this experiment. Figure 1 (top) shows the average rating of each ad sorted from most annoying to least annoying. This plot shows the quite striking effect of animation on annoyance: the 21 most annoying ads were all animated, and the 24 least annoying ads were all static. This is further exemplified in Figure 1 (bottom) in which each animated ad is compared to its static variant. Here, the pairs are sorted by the annoyingness of the animated variant and the static version is placed at the same x-coordinate as its animated counterpart. The static versions tend to fall below the animated versions, often by several standard errors or more than one rating point. Put another way, we did not observe a case where animation significantly improved an ad on this annoyingness scale. Since the advertisers and products are held constant within each pair, it seems that animation alone is a cause of annoyance.

As mentioned, the 10 most and least annoying ads identified in this experiment will serve as the sets of "bad" and "good" ads in the next experiment which is described in Section 4. Figures 2 and 3 show examples from these two sets.

Recall that each participant who rated an ad as either "A bit more annoying than the average ad in this experiment" or "Much more annoying than the average ad in this experiment", was asked to write a few words as to why they found that ad annoying. In all there were 1846 such responses from the 141 respondents. First, we manually constructed a set of categories to characterize these reasons based on a 5% sample of the comments. We then analyzed the entire corpus treating each response as a bag of words. We looked at all words that occurred at least 10 times (excluding "stop words"), and assigned them to a relevant category. Then, for each category, we totaled up the number of times the words in that category appeared in the bag. The results are shown in Table 1.

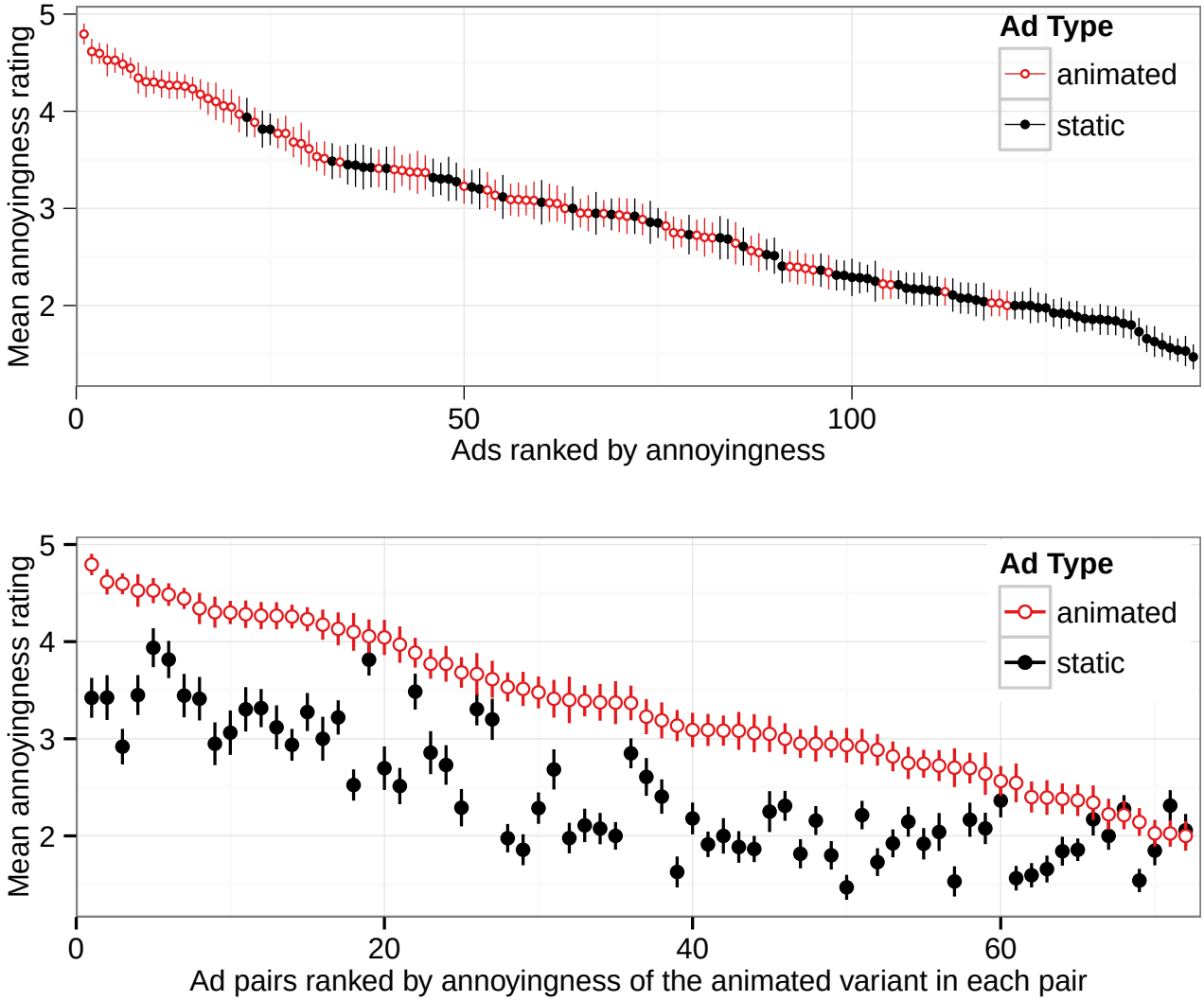
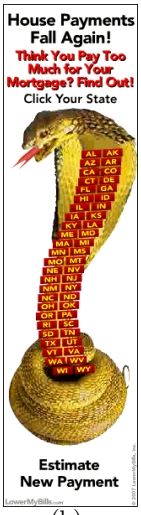



**Figure 1: The top panel ranks ads by annoyingness and shows that the 21 most annoying ads were animated and the 24 least annoying ads were static. The bottom panel ranks pairs of ads by the annoyingness of the animated variant. The static variants tend to fall below their animated versions, suggesting that animation increases annoyingness, even when the advertiser and product are held constant. Error bars are $\pm$ 1 standard error.**


(a)




(b)

**Figure 2: Two examples of "bad" ads. Figure 2(a) is a medium rectangle whereas Figure 2(b) is a skyscraper. In the animated version of Figure 2(a) the woman's arm rapidly moves up and down. Similarly, in Figure 2(b) the snake writhes wildly, sticks its tongue out, and pulses its eyes red.**

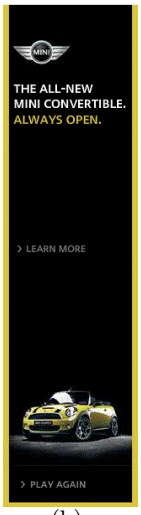
This rudimentary text analysis shows that the most frequent topic was the animation of the ad, e.g., "too much movement". The next most common topic was the ad's deleterious effect on user attention, e.g., "it diverts my attention from what is important; the content on the page". The third most common topic was the aesthetic of the ad, e.g., "another cheap looking ad that I would never click." This leads to the fourth most common complaint. These poor aesthetics would often lead people to believe the ad is a scam, e.g., "seems like a scam with a cheap design". Finally, many of the ads had a graphic that did not logically relate to the product such as a dancing wizard in an ad for online classes. This type of *non sequitur* also bothered users, e.g., "A dancing wizard has nothing to do with going to school."

We list here two observations from this analysis. First, the reasons listed above span the costs to the user, publisher and advertiser mentioned previously. Complaints about movement or animation and how they distract from the content show there is a cost to users and publishers. In addition, users were skeptical of aesthetically unappealing or illogical creatives, suggesting that annoying ads also have a cost for advertisers. Second, the chief complaint was about animation, which corroborates our finding that animation exerts a causal influence on annoyance.

# 4. MEASURING THE COST OF ADS

As seen in the previous section there might be a variety of attributes of an ad that a user might find annoying. If we view attributes of an ad as residing in a multidimensional space, the average ratings of the previous section are how users project of that multidimensional space onto a one-dimensional annoyingness scale. Thus, we use the method of Toomim et al. [18] along with ads from each end of this annoyingness scale as sets of "bad" and "good ads" to measure the cost of annoying ads.


(a)




(b)

**Figure 3: Two examples of "good" ads.**

## 4.1 Method

The participants were 1223 Mechanical Turk workers with at least a 90% approval rating who participated for a base pay of 25 cents and a bonus, which was not disclosed before the HIT was accepted to prevent selection effects. The experiment was advertised as an email classification task and ran for a period of two weeks. Upon accepting the HIT, participants were randomly assigned to one of nine conditions: three pay conditions and three ad conditions. The pay conditions offered a bonus of one, two, or three cents per five emails classified (i.e., .2, .4, or .6 cents per email), and the ad conditions varied whether "bad ads", "good ads", or no ads were displayed in the margin as the task was completed. No mention was made of pay conditions, ad conditions or random assignment, and a search on turkernation.com, a discussion forum for Mechanical Turk workers, found no mention of either experiment. A chi-squared test found no significant difference in the number of participants beginning work across the nine conditions.

In all conditions, the task consisted of classifying the content of emails as "spam", "personal", "work" or "e-commerce" related. Emails were drawn from the public-domain Enron email dataset[3] with one email presented per page, along with accompanying ads, if any. In the "bad ads" condition, two ads randomly drawn from the 10 most annoying ads in our first experiment were displayed in the margins around the email being classified. Figure 4 is a screenshot of the bad ads condition. The "good ads" condition was the same, except the ads were drawn from the 10 least annoying ads. In both conditions, ads were drawn randomly from their respective pools with each page load, and the urls for the ads were such that ad blocking software would not filter them out. The "no ads" condition simply had whitespace in the margin. The width of the page and text area was held constant across conditions and chosen so that it would be visible to the vast majority of Web browsers.

[3] `http://www.cs.cmu.edu/~enron/` Identifying information such as email addresses, phone numbers and the name "Enron" were removed.

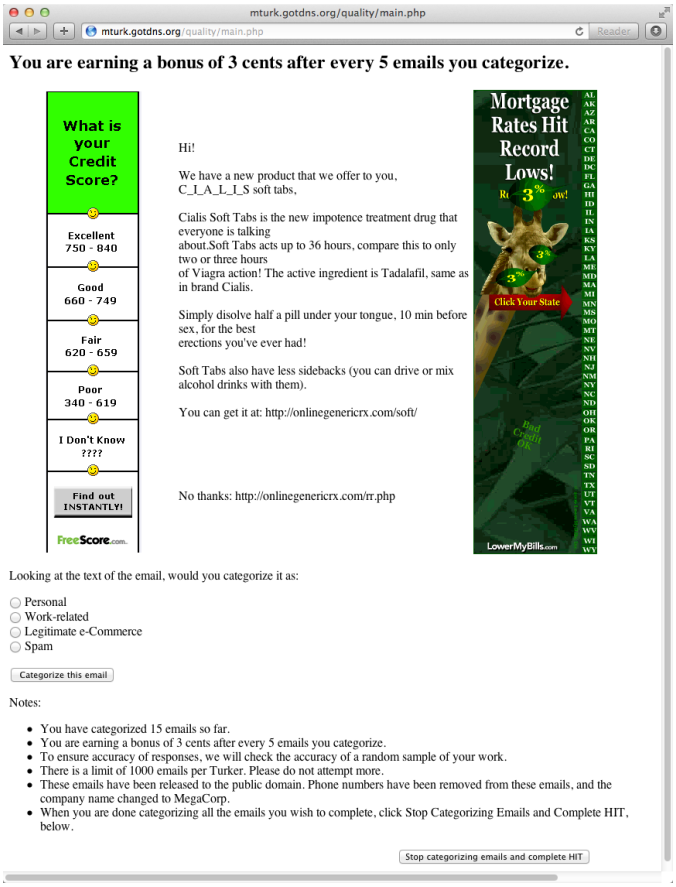



**Figure 4: Screenshot of the email categorization task showing the bad ads condition.**

At the bottom of each email classification page, participants were shown how many emails they had rated, their pay rate, and a review of the instructions. The footer included two buttons: one allowing them to submit and rate another email, and a second allowing them to stop categorizing and collect their payment. Participants were allowed to classify up to 1000 emails.

## 4.2 Results

Let an impression be one participant viewing one email (and its accompanying ads, if any), regardless of whether the participant classifies the email or quits before classifying it. Since an email is presented as soon as the user acknowledges the instructions, each of the 1223 participants generated at least one impression. The overall distribution of impressions per person is skewed with a mean of 61, a median of 25 and first and third quartiles of 6 and 57. Being bounded by 1 from below and effectively unbounded from above (only two participants reached the upper limit), these impressions constitute count data. In particular, they are overdispersed count data relative to the Poisson (observed variance / theoretical Poisson data variance is 228.7, p<.0001) and thus well suited to a negative binomial generalized linear model (GLM) [19]. Model 1 in Table 2 provides the coefficients of a negative binomial GLM of impressions on pay rate and dummy variables for the presence of "good ads" or no ads, relative to the baseline of "bad ads". Relative to a baseline

| | Model 1 | Model 2 |
|---|---|---|
| (Intercept) | $3.43^{***}$ | $3.43^{***}$ |
| | (0.12) | (0.12) |
| Good ads | $0.17^{*}$ | |
| | (0.10) | |
| No ads | $0.22^{**}$ | |
| | (0.10) | |
| Pay rate | $26.47^{***}$ | $26.61^{***}$ |
| | (4.80) | (4.80) |
| Good ads or no ads | | $0.19^{**}$ |
| | | (0.08) |
| AIC | 12158.57 | 12156.85 |
| BIC | 12184.12 | 12177.29 |
| Log Likelihood | −6074.29 | −6074.43 |
| Deviance | 1481.00 | 1481.04 |
| Number of observations | 1223 | 1223 |

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

**Table 2: Negative binomial GLM of impressions on ad condition and pay rate. Bad ads lead to fewer impressions than good ads or no ads. Coefficients are expressed in log impressions; predicted values are displayed in Figure 5. Pay rate is in dollars per five impressions (.01,.02,.03). Standard errors are in parentheses.**

of "bad ads", both the "good ads" condition and the no ads condition led to substantially more impressions (19% and 25% more impressions, respectively). Model 2 is the same as Model 1 but replaces the two ad dummies with one new dummy representing the "good ads" and no ads conditions combined and results in a similar conclusion. As the coefficients in Table 2 are expressed in log terms, the effects of the conditions on raw impressions is most easily seen in Figure 5, which also makes clear that the difference in impressions between the "good ads" and "no ads" conditions is not significant.

The model expressed in Table 2 and Figure 5 can be used to estimate the compensating differential of annoying ads in this experiment. Since the curves are slightly non-linear, a range of compensating differentials could be calculated across the pay rate and ad conditions. To get a simple, single approximation we use the middle, "good ads" condition to estimate the effect of pay raises. We take the average of the .2 to .4 and .4 to .6 cent differences, giving an estimated increase of 16.58 impressions resulting from a .2 cent per impression pay raise. When summarizing the effect of ad quality, we use the number of impressions at the .4 cent pay rate. Moving from "bad ads" to no ads, impressions increase by 12.68. The pay raise required to achieve a 12.68 impression increase is .153 cents per impression (= .2∗12.68/16.58) or $1.53 CPM (cost per thousand impressions). That is, in this experiment, a participant in the "bad ads" condition would need to be paid an additional $1.53 per thousand impressions to generate as many impressions as a person in the condition without ads. Similarly, moving from the "bad ads" condition to the "good ads" condition resulted in an additional 9.52 impressions per person. It would require a pay raise of .115 cents per impression (= $.2 * 9.52/16.58$) to generate 9.52 additional impressions, meaning that people in

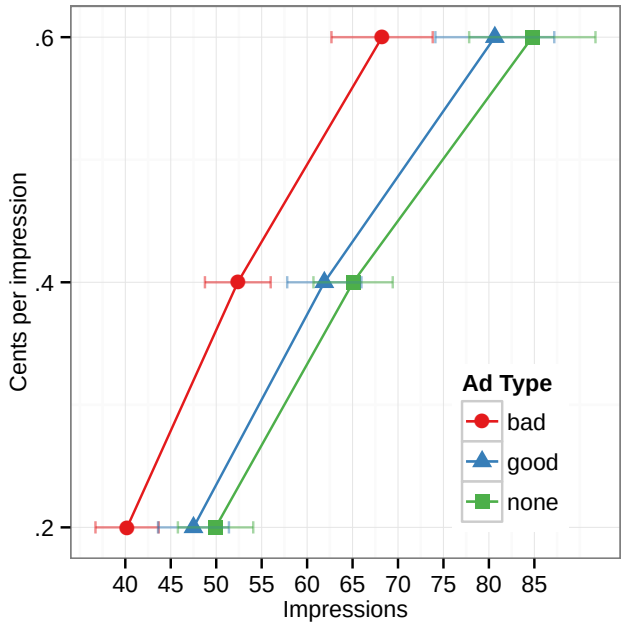



**Figure 5: Estimated impressions per condition based on the negative binomial model. Error bars are ± 1 standard error.**

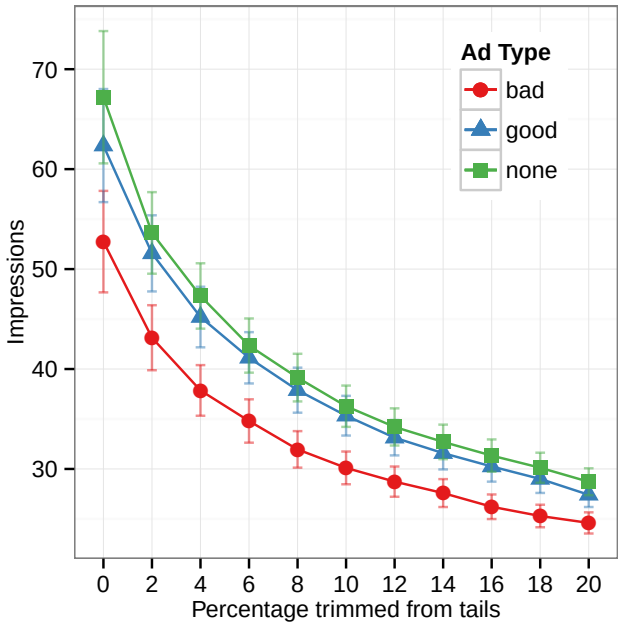



**Figure 6: Robustness check: Trimmed means of the effect of advertisement quality on impressions, collapsed across pay conditions. Means drop as observations are excluded from the tails, reflecting skewness, while the ad annoyingness effect remains. Error bars are ± 1 standard error.**

the "bad ads" conditions would need to be paid an additional $1.15 CPM to generate as many impressions as in the "good ads" condition. By the same logic, a pay raise of $0.38 CPM would be required to have the "good ads" condition generate as many impressions as the no ads condition.

With highly skewed data, it is possible that extreme observations can bias the results, even when working in log terms as in the negative binomial model. As a robustness check of the general effect of the ad conditions on the number of impressions generated, Figure 6 shows how the mean number of impressions per ad condition changes as extreme observations are eliminated from the tails of each ad condition. Means and standard errors decrease as more observations are trimmed, reflecting the pull of extremely large observations. Nonetheless, the relative position of the ad conditions remains stable: annoying ads are associated with fewer impressions and the difference between the "good ads" and "no ads" conditions is slight.

In addition to having an effect on impressions, annoying advertisements may have a measurable deleterious effect on user experience. For example, a user that is distracted by annoying ads might have a harder time concentrating or care less about site content. Because the email corpus used contains the correct classifications (spam or not spam) for every email, it is possible to test the effect of ad and pay conditions on email classification accuracy. Overall accuracy in the experiment was .91. Table 3 shows regressions of accuracy (proportion of correct email classifications) on ad conditions, using the same independent variables as in Table 2 with the addition of the number of impressions generated per participant. Relative to the "bad ads" condition, accuracy was significantly higher in the "good ads" and "no ads" conditions. That is, annoying ads harmed accuracy. The number of observations is lower than the total number of participants since 166 individuals exited before classifying a single email and thus have undefined accuracy scores. Note that these 166 participants did generate one impression, that is, they saw one email to classify and its accompanying ads (if any), but chose to be paid immediately instead of submitting a classification. The coefficient on the impressions variable is negative such that accuracy dropped 1 percentage point per 128 emails classified.

## 5. THEORETICAL MODEL

In this section we give a theoretical prediction as to how our empirical findings matter for internet advertising. We will exhibit a relationship between a publisher's market share and the user cost of the ads they choose to display. We begin by defining some notation. Let $A$ be the cost to a typical user from ads run, measured in dollars. Let $v$ denote the value of the publisher's content to a typical user, also measured in dollars. Let $u = v - A$ be the net utility to the user consuming the content (their value minus their cost). Finally, $R(A)$ is the revenue associated with user ad cost $A$. It is assumed to be increasing and concave.

In this setting a publisher will select $A$ and $v$ to maximize revenue for the chosen user utility. Competition with alternative publishers will influence only the net user utility offered. This insight applies also to different kinds of ads. If the publisher has a portfolio of ads $a_1, a_2, \ldots$, where $a_i$ is the user cost, with revenue $r_i$, the publisher will select a set $S$ to maximize revenue subject to a target user cost. Indeed,

| | Model 1 | Model 2 |
|---|---|---|
| (Intercept) | $0.90^{***}$ | $0.90^{***}$ |
| | (0.01) | (0.01) |
| Impressions | $-7.8e{-5}^{***}$ | $-7.8e{-5}^{***}$ |
| | (0.00) | (0.00) |
| Good ads | $0.02^{**}$ | |
| | (0.01) | |
| No ads | $0.03^{***}$ | |
| | (0.01) | |
| Pay rate | 0.14 | 0.14 |
| | (0.43) | (0.43) |
| Good ads or no ads | | $0.02^{***}$ |
| | | (0.01) |
| Number of observations | 1057 | 1057 |

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

**Table 3: Linear regression of classification accuracy on impressions, ad condition, and pay rate. Excludes participants who classified no emails, as their accuracy is undefined. Pay rate is in dollars per five impressions. Standard errors are in parentheses.**

the function $R$ would be:

$$R(A) = \max_S \sum_{i \in S} r_i \text{ subject to } \sum_{i \in S} a_i \leq A.$$

This function is poorly behaved since it has flat spots and jumps. Moreover, the selection of ads is equivalent to the knapsack problem and thus NP-hard. Nevertheless, since it is very unlikely that showing a single ad would maximize the revenue of the advertiser, the greedy algorithm (which displays the ads from highest to lowest by the ratio of $r_i/a_i$ until the user cost approaches $A$) involves at worst the percentage loss in filling $A$, and thus is small when $A$ is large relative to the individual ads, as in our context. This approximate solution has the property that $R$ is non-decreasing and approximately concave, as assumed.

We simplify the problem by considering a differentiable $R$, which may be thought of as the consequence of an infinite number of very small ads. In our setting, the assumption of many small ads is more plausible than in many economic settings simply because ad space per user generally sells for less than a U.S. penny.

Without loss of generality we can model the publisher as selecting $A$ and $u$, with $v = u + A$ determining $v$. Let $c(v)$ be the cost of content to the publisher; it is assumed increasing and convex. Then the publisher's net profit for a given value of $u$ is,

$$\pi(u) = \max_A R(A) - c(u + A)$$

The first order condition for the optimal value $A^*$ is,

$$0 = R'(A^*) - c'(u + A^*).$$

The concavity of $R$ and assumed convexity of $c$ ensure that the function $R(A) - c(u + A)$ is concave in $A$, and thus any interior solution of the first order conditions is a global maximum[4]. Moreover, differentiating the first order conditions gives,

$$\frac{dA^*}{du} = \frac{c''(u + A^*)}{R''(A^*) - c''(u + A^*)} \in (-1, 0)$$

The function $\pi$ has the following property by the envelope theorem,

$$\pi'(u) = -c'(u + A^*) < 0.$$

Differentiating this gives,

$$\begin{aligned} \pi''(u) &= c''(u + A^*)\left(1 + \frac{dA^*}{du}\right) \\ &= -c''(u + A^*)\left(1 + \frac{c''(u + A^*)}{R''(A^*) - c''(u + A^*)}\right) \\ &= -c''(u + A^*)\left(\frac{R''(A^*)}{R''(A^*) - c''(u + A^*)}\right) \\ &< 0 \end{aligned}$$

Thus $\pi$ is decreasing and concave. Provided there is no upper bound on advertising or on content value, any $u$ can be accommodated.

Let $x(t) \in [0, 1]$ be the publisher's market share at time $t$. We will suppress the time dependence for clarity but note that $x$ and $u$ are both time dependent. Let $u^*$ be a reference utility level—think of it as the utility offered by imperfect substitute products. If there is only one rival offering a competitive product it would certainly react to changes in $u$. In this case we are modeling the benefit the publisher would receive before the competitor has time to react. If the substitute is competitively supplied, however, then we can take $u^*$ as a given, since the competition has already been forced it to its optimum value. We suppose that the process determining consumers switching to alternative services depends on the net utility offered, according to the logistic equation

$$x'(t) = \lambda(u - u^*)x(1 - x). \tag{1}$$

This equation is commonly used in population growth because population growth depends on the existing population $x$ (in our context, this might be the word of mouth of existing users influencing adoption), the fraction of non-users $1 - x$ who can be converted, the speed $\lambda$ at which they convert, and the difference in net utility $u - u^*$. This functional form also arises through the replicator dynamics [8].

Rewriting Equation 1, we can think of the publisher as choosing the share growth $x'$, which dictates user utility

$$u = u^* + \frac{x'}{\lambda x(1 - x)}.$$

The publisher's flow revenue is $x\pi(u)$. Let $r$ be the interest rate facing the publisher. The publisher maximizes present discounted which is

$$\int_0^\infty e^{-rt} x\pi(u)dt = \int_0^\infty e^{-rt} x\pi\left(u^* + \frac{x'}{\lambda x(1 - x)}\right)dt.$$

THEOREM 1. *If $-\frac{r}{\lambda}\frac{\pi'(u^*)}{\pi(u^*)} \geq 1$, the publisher's market share converges to $x^* = 0$. Otherwise, the terminal market share is given by*

$$x^* = 1 + \frac{r}{\lambda}\frac{\pi'(u^*)}{\pi(u^*)}.$$

[4]We will focus only on interior solutions, but the analysis readily generalizes to the border cases when $A^* = 0$ or $u + A^* = 0$. Indeed, the only properties used of $\pi$ are that it is decreasing and concave, both of which are preserved in the border cases.

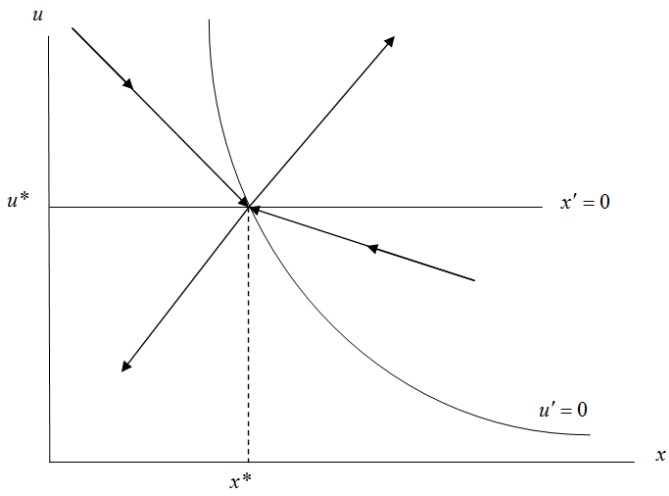



**Figure 7: Phase diagram relating market share to user utility as described by Theorem 1**

*In addition, if $x < x^*$, then user utility $u > u^*$ and is decreasing over time. If $x > x^*$, $u < u^*$ and is increasing over time.*

The proof of this theorem is given the in Appendix. The solution is illustrated in the phase diagram given in Figure 7. The equilibrium for any starting market share $x$ involves the path pointing toward $(x^*, u^*)$. The value of $u$ adjusts to put the publisher on this path. Starting with a low market share, the publisher sets a high user utility which is a combination of low advertising and high content quality, and then gradually degrades user utility and increases advertising intensity. In contrast, a publisher who starts with a high market share will set a very low content quality and high advertising intensity, and then gradually improve the user experience. An increase in the interest rate decreases $x^*$, the asymptotic market share. An increase in the competitive level $u^*$ increases $x^*$ when $\pi$ is log-convex and vice-versa.

There are several conclusions one can draw from this model. First, since the terminal market share predicted in Theorem 1 depends on $\pi$, which depends on $A$ and $u$, the model justifies the ratio of the revenue to user cost as the key metric for advertising selection. Second, in a competitive advertising market, all ads will sell for a constant times the user cost. Annoying ads will run only when their revenue is very high or the publisher is extremely willing to sacrifice user experience for revenue. Third, a legacy publisher, whose market share is large because they initially faced little competition, will start with a lower user experience involving both more ads and worse content than an entrant. This will result in the legacy publisher seeing a fast decline in user base. The legacy publishers content will gradually improve until a stable point is reached. Finally, if consumers react sufficiently slowly to changes in content (that is, $\lambda$ is small), a legacy publisher will gradually go extinct rather than offer a better user experience.

# 6. CONCLUSION

The first study reported here showed that people find animated advertisements more annoying than static ones, holding all else constant. This study also identified five categories of complaints about annoying ads providing a first pass at identifying undesirable features. We used the good and bad ads from this study to measure the compensating wage differential in the second study. The main result of this paper is that annoying ads lead to site abandonment and thus fewer impressions than good ads or no ads. In what might be seen as good news for publishers, good ads and no ads led to roughly equal numbers of impressions. Annoying ads impaired people's ability to carry out an email classification task, suggesting that annoying ads have a real cost to users beyond mere annoyance. Finally, we provided a theoretical model that computes a dynamic equilibrium, which permits studying not only properties of the steady state, but the adjustment to that state as well. This model can be used to understand the behavior of legacy publishers, who inherited a large market share, in the face of competition from new entrants.

We calculated the compensating wage differential in our experiment of bad ads to no ads to be \$1.53 CPM, bad ads to good ads to be \$1.15, and good ads to no ads to be \$.38 CPM. Some care must be taken in interpreting these numbers. While we picked a task—classifying emails—that should be familiar and common for most internet users, this task may not be representative of other internet tasks like reading news stories or searching for products to purchase. Abandonment rates may differ with different tasks and the effects of advertising may vary as well. While virtually every web service features competition, the switching costs vary from very low in consuming news to relatively high in changing email services. Because our users on Mechanical Turk have an outside option of working on an alternative HIT, we expect our results to be most applicable to situations involving lower switching costs. Nevertheless, we expect that our finding that annoying ads cost the user at least \$1 CPM over more pleasant ads will be obtained in some other environments.

For these reasons, we suggest further studies be done on Mechanical Turk, as field experiments, and in laboratories to measure this differential on similar and different tasks. If studies across various domains with a variety of tasks and outside options arrive at similar differentials, more credence can be placed on these numbers. We view this work as a first step in this direction. If future work arrives at similar estimates across a variety of publishers, such estimates could serve as a useful lower bound for what a publisher should charge to run these ads. Moreover, it will be valuable to use the compensating differentials approach to price the various bad aspects of ads, including animation and poor aesthetics.

This work also suggests a variety of policy recommendations. Most directly, the \$1 CPM user cost of bad ads has practical consequences for publishers, especially as bad ads often command lower CPMs. It is a reason that publishers should insist on a substantial premium for annoying advertisements. Moreover, a publisher could randomize which users see which ads and track both time spent on the page and the frequency with which users return to the site. This type of experimentation would capture longer term effects of annoying ads than those studied here. Also, publishers could give users an option to close or replace an ad. A replacement event would allow the publisher to infer that a user would prefer a random ad over the ad currently shown. Advertisers with a high closure rate should be charged more. Furthermore, it would be reasonable to assume that more

annoying ads would be closed or replaced faster than less annoying ads. Ad replacement would help the user by removing the annoying ad and the publisher by making it possible to charge for two impressions.

## 7. ACKNOWLEDGMENTS

We thank Randall A. Lewis, Justin M. Rao, and David H. Reiley for helpful conversations.

## APPENDIX

In this section we give the proof of Theorem 1.

PROOF. Define $y = \log\left(\frac{x}{1-x}\right)$. Note, $y' = \frac{x'}{x} + \frac{x'}{1-x} = \frac{x'}{x(1-x)}$, and $x = \frac{e^y}{1+e^y}$. Furthermore, $1 + e^y = 1 + \frac{x}{1-x} = \frac{1}{1-x}$. Thus we can reformulate the publisher's optimization problem as that of maximizing $\int_0^\infty e^{-rt}\frac{e^y}{1+e^y}\pi\left(u^* + \frac{1}{\lambda}y'\right)dt$. Let $F(y, y', t) = e^{-rt}\frac{e^y}{1+e^y}\pi\left(u^* + \frac{1}{\lambda}y'\right)$. The Euler equation for this problem is

$$\begin{aligned} 0 &= \frac{\partial F}{\partial y} - \frac{d}{dt}\frac{\partial F}{\partial y'} \\ &= e^{-rt}\frac{e^y}{(1+e^y)^2}\pi\left(u^* + \frac{1}{\lambda}y'\right) \\ &\quad -\frac{1}{\lambda}\frac{d}{dt}e^{-rt}\frac{e^y}{1+e^y}\pi'\left(u^* + \frac{1}{\lambda}y'\right) \\ &= \frac{x}{\lambda}e^{-rt}[\lambda(1-x)\pi(u) + r\pi'(u) \\ &\qquad -\pi'(u)\lambda(u-u^*)(1-x) - \pi''(u)u'] \end{aligned}$$

Thus,

$$\pi''(u)u' = \lambda(1-x)\pi(u) + r\pi'(u) - \lambda(u-u^*)(1-x)\pi'(u).$$

A steady state of the system holds when $x' = u' = 0$, or $u = u^*$ and $0 = \lambda(1-x^*)\pi(u^*) + r\pi'(u^*)$. This is equivalent to

$$1 - x^* = -\frac{r}{\lambda}\frac{\pi'(u^*)}{\pi(u^*)}.$$

If $-\frac{r}{\lambda}\frac{\pi'(u^*)}{\pi(u^*)} \geq 1$, all optimal paths involve $x \to 0$ as there is no internal steady state. When $-\frac{r}{\lambda}\frac{\pi'(u^*)}{\pi(u^*)} < 1$, there is an interior steady state. The $u' = 0$ curve occurs when $0 = \lambda(1-x)\pi(u) + r\pi'(u) - \lambda(u-u^*)(1-x)\pi'(u)$. Thus, near $(x^*, u^*)$,

$$\begin{aligned} \left.\frac{du}{dx}\right|_{u'=0} &= \frac{\lambda\pi(u) - \lambda(u-u^*)\pi'(u)}{\lambda(1-x)\pi'(u) + r\pi''(u) - \lambda(1-x)\pi'(u) - \lambda(u-u^*)(1-x)\pi'''(u)} \\ &= \frac{\lambda\pi(u) - \lambda(u-u^*)\pi'(u)}{+r\pi''(u) - \lambda(u-u^*)(1-x)\pi'''(u)} \approx \frac{\lambda\pi(u^*)}{r\pi''(u^*)} < 0. \end{aligned}$$

We can obtain insight about the paths near this solution by a first order Taylor approximation. The strategy looks like this. Write

$$\begin{pmatrix} x' \\ u' \end{pmatrix} = \begin{pmatrix} \lambda(u-u^*)x(1-x) \\ \lambda(1-x)\frac{\pi(u)}{\pi''(u)} + r\frac{\pi'(u)}{\pi''(u)} - \lambda(u-u^*)(1-x)\frac{\pi'(u)}{\pi''(u)} \end{pmatrix}$$

$$= \begin{pmatrix} g(x,u) \\ h(x,u) \end{pmatrix}.$$

$$\begin{pmatrix} x' \\ u' \end{pmatrix} \approx \left.\begin{pmatrix} \partial g/\partial x & \partial g/\partial u \\ \partial h/\partial x & \partial h/\partial u \end{pmatrix}\right|_{(x,u)=(x^*,u^*)} \begin{pmatrix} x - x^* \\ u - u^* \end{pmatrix}.$$

Locally the behavior of the general system is approximated by the behavior of the linear system. The only challenging term in the matrix is

$$\begin{aligned} \left.\frac{\partial h}{\partial u}\right|_{u=u^*,x=x^*} &= \frac{\partial}{\partial u}\lambda(1-x)\frac{\pi(u)}{\pi''(u)} + r\frac{\pi'(u)}{\pi''(u)} \\ &\quad -\lambda(u-u^*)(1-x)\left.\frac{\pi'(u)}{\pi''(u)}\right|_{u=u^*,x=x^*} \\ &= -\lambda(1-x^*)\frac{\pi(u^*)\pi'''(u^*)}{\pi''(u^*)^2} \\ &\quad +r\left(1 - \frac{\pi'(u^*)\pi'''(u^*)}{\pi''(u^*)^2}\right) \\ &= r \end{aligned}$$

Thus,

$$\begin{pmatrix} x' \\ u' \end{pmatrix} \approx \begin{pmatrix} 0 & \lambda x^*(1-x^*) \\ -\lambda\frac{\pi(u)}{\pi''(u)} & r \end{pmatrix}\begin{pmatrix} x - x^* \\ u - u^* \end{pmatrix}$$

The eigenvalues of the linear system are determined by solutions $\mu$ to

$$0 = \det\begin{pmatrix} -\mu & \lambda x^*(1-x^*) \\ -\lambda\frac{\pi(u)}{\pi''(u)} & r-\mu \end{pmatrix}$$

$$0 = \mu^2 - r\mu + \lambda^2 x^*(1-x^*)\frac{\pi(u^*)}{\pi''(u^*)}$$

solving for $\mu$ gives,

$$\mu = \frac{1}{2}\left(r \pm \sqrt{r^2 - 4\lambda^2 x^*(1-x^*)\frac{\pi(u^*)}{\pi''(u^*)}}\right)$$

Because $\pi''(u^*) < 0$, there is one positive and one negative eigenvalue and both are real. Thus, the behavior of the system is a saddle, as illustrated in Figure 7. There are infinitely many paths consistent with equilibrium given by the differential equations. Which one is the right one? In the case when $-\frac{r}{\lambda}\frac{\pi'(u^*)}{\pi(u^*)} \geq 1$, all paths that don't violate transversality lead to $x = 0$. Suppose $x$ is a candidate limit. Consider setting $u = u^* + \Delta$ for $t$ units of time. The firm earns

$$\begin{aligned} \Psi &\approx \left(\int_0^t e^{-rs}ds\right)x\pi(u^* + \Delta) \\ &\quad + \left(\int_0^t e^{-rs}ds\right)(x + \lambda x(1-x)\Delta t)\,\pi(u^*)y \\ &= \frac{1}{r}\left(1 - e^{-rt}\right)x\pi(u^* + \Delta) \\ &\quad + \frac{1}{r}e^{-rt}\left(x + \lambda x(1-x)\Delta t\right)\pi(u^*) \end{aligned}$$

$$\begin{aligned} \tfrac{1}{t}\left.\tfrac{\partial\Psi}{\partial\Delta}\right|_{\Delta=0} &= \tfrac{1}{t}\left(\tfrac{1}{r}\left(1 - e^{-rt}\right)x\pi'(u^*) + \tfrac{\lambda}{r}e^{-rt}x(1-x)t\pi(u^*)\right) \\ &= \tfrac{\lambda}{r}x\pi(u^*)\left(\lambda\left(\tfrac{1-e^{-rt}}{t}\right)\tfrac{\pi'(u^*)}{\pi(u^*)} + e^{-rt}(1-x)\right) \\ &= \tfrac{\lambda}{r}x\pi(u^*)\left(\tfrac{\lambda}{r}\tfrac{\pi'(u^*)}{\pi(u^*)} + (1-x)\right) \end{aligned}$$

Thus, it pays to increase a convergent $x$ if and only if $x < 1 - \frac{r}{\lambda}\frac{\pi'(u^*)}{\pi(u^*)}$, implying that this is only candidate for convergent paths when $-\frac{r}{\lambda}\frac{\pi'(u^*)}{\pi(u^*)} < 1$. □

# A. REFERENCES


[1] Jan P. Benway. Banner blindness: The irony of attention grabbing on the world wide web. In *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting*, volume 1, pages 463–467, October 1998.

[2] Moira Burke, Anthony Hornof, Erik Nilsen, and Nicholas Gorman. High-cost banner blindness: Ads increase perceived workload, hinder visual search, and are forgotten. *ACM Transactions on Computer-Human Interaction*, 12(4):423–445, December 2005.

[3] Georg Buscher, Susan T. Dumais, and Edward Cutrell. The good, the bad, and the random: an eye-tracking study of ad quality in web search. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '10, pages 42–49, New York, NY, USA, 2010. ACM.

[4] Xavier Drèze and François-Xavier Hussherr. Internet advertising: Is anybody watching? *Journal of Interactive Marketing*, 17(4), 2003.

[5] Avi Goldfarb and Catherine Tucker. Online display advertising: Targeting and obtrusiveness. *Marketing Science*, 30(3):389–404, May–June 2011.

[6] Daniel G. Goldstein, R. Preston McAfee, and Siddharth Suri. The effects of exposure time on memory of display advertisements. In *Proceedings of the 12th ACM conference on Electronic commerce*, EC '11, pages 49–58, New York, NY, USA, 2011. ACM.

[7] Daniel G. Goldstein, R. Preston McAfee, and Siddharth Suri. Improving the effectiveness of time-based display advertising. In *Proceedings of the 13th ACM conference on Electronic commerce*, EC '12, 2012.

[8] Ed Hopkins. Two competing models of how people learn in games. *Econometrica*, 70(6):2141–2166, November 2002.

[9] John J. Horton, David G. Rand, and Richard J. Zeckhauser. The online laboratory: Conducting experiments in a real labor market. *Experimental Economics*, 14(3):399–425, 2011.

[10] Winter Mason and Siddharth Suri. Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods*, 44(1):1–23, March 2012.

[11] Winter Mason and Duncan J. Watts. Collaborative learning in networks. *Proceedings of the National Academy of Sciences*, December 2011.

[12] Winter A. Mason and Duncan J. Watts. Financial incentives and the performance of crowds. In Paul Bennett, Raman Chandrasekar, and Luis von Ahn, editors, *Proceedings of the ACM SIGKDD Workshop on Human Computation*, pages 77–85. ACM, 2009.

[13] Gabriele Paolacci, Jesse Chandler, and Panagiotis G. Ipeirotis. Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, 5:411–419, 2010.

[14] Allen Parducci and Linda F. Perrett. Category rating scales: Effects of relative spacing and frequency of stimulus values. *Journal of Experimental Psychology*, 89(2):427–452, 1971.

[15] John G. Riley. Silver signals: Twenty-five years of screening and signaling. *Journal of Economic Literature*, XXXIX:432–478, 2001.

[16] Aaron D. Shaw, John J. Horton, and Daniel L. Chen. Designing incentives for inexpert human raters. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*, CSCW '11, pages 275–284, New York, NY, USA, 2011. ACM.

[17] Siddharth Suri and Duncan J. Watts. Cooperation and contagion in web-based, networked public goods experiments. *PLoS One*, 6(3), 2011.

[18] Michael Toomim, Travis Kriplean, Claus Pörtner, and James A. Landay. Utility of human-computer interactions: Toward a science of preference measurement. In *Proceedings of CHI 2011: ACM Conference on Human Factors in Computing Systems*, 2011.

[19] W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S*. Springer, New York, 2002.

[20] Jing Wang, Siddharth Suri, and Duncan Watts. Cooperation and assortativity with dynamic partner updating. *Proceedings of the National Academy of Sciences*, 109(36):14363–14368, Septemeber 2012.

[21] Chan Yun Yoo and Kihan Kim. Processing of animation in online banner advertising: The roles of cognitive and emotional responses. *Journal of Interactive Marketing*, 19(4):18–34, 2005.