Presented by

Coalition for Better Ads

An experimental methodology to rank N ad experiences by consumers' perceptions

Ad Experience Research Group
April 2016

Table of Contents:

Abstract

Good ad experiences power an open and accessible web, where interested users can access diverse content regardless of their individual ability to pay. These ad experiences introduce new ideas and products, let people know of sales, and make the online marketplace work. Bad ad experiences — those that annoy or distract — threaten this ecosystem by disrupting this relationship between users, content creators, and brands. In previous work (Ad Experience Research Group 2016), we have shown that users have strong preferences for some types of ads over others, and that they rate them accordingly.

In this paper, we describe and present results from a new methodology that surveyed approximately 9370 participants to rate ad experiences in randomized and unbiased trials that mimic real-life web use. To our knowledge, this is the largest experiential study on ad annoyance ever undertaken.

In the study, each participant: (1) viewed four articles on a fictional website, three with ads and one without ads; (2) was presented with survey questions after each article to rate their overall experience, including ads (if any); (3) at the end of reading the four articles, ranked ad experiences — choosing the most and least preferred ads of the three they'd seen.

This paper presents findings for both mobile and desktop ads experiences.

In these studies we collected ratings and rankings for 37 mobile and 40 desktop web experiences. In order to reduce the number of participants needed to rate and rank these ads, we employed an iterative sampling algorithm to determine which ad experiences to show to each participant, and the Bradley-Terry algorithm (Turner and Firth, 2012) approach to generate an overall ranking for mobile ad experiences and an overall ranking for desktop ad experiences.

This methodology found strong differences between ad experiences. Ads that interrupt, distract, and prevent user control ranked as least preferred and rated as most annoying: these are ads that pop up in the middle of reading, ads that can't be dismissed before a countdown, ads with flashing animations, or ads that take up a high percentage of the content area of a page (i.e., comprising 35 - 50% of a mobile content). Ads that more seamlessly coexist with content are strongly preferred by consumers: these are small sticky ad units at the bottom of mobile pages, static ads at low densities, immediately-rendered prestitials and poststitials with no countdown.

Overall, this work showed that ads that can be used by content creators to fund their efforts, have differing levels of user preference. Creators and brands have a vested interest in improving user sentiment, and should take care when determining the best way to deliver advertising messages to consumers.

Introduction

Content publishers use display ads to generate revenue providing free content to their readers. These ads also provide readers with a way to learn about businesses and products they may like. However, ad experiences are not equally appreciated by users: some are more annoying than others, like animating or flashing ads (Goldstein, McAfee, Suri 2014). In a previous paper (Ad Experience Research Group 2016), we described a methodology for ranking and rating ad experiences that we created to understand how users perceived ads. While the methodology was successful in discerning the differences between some ad experiences, we hadn't designed it to rank more than three ad experiences or work outside of mobile web ad experiences.

To expand on our work and help publishers and advertisers develop better ad experiences, we modified the original methodology to enable users to rate and rank a large number of ad experiences. We also designed experiences for desktop to validate extending the methodology to other environments beyond mobile web.

First, this paper summarizes the previous methodology and experimental setup. Second, it describes how we updated the methodology to test any number of ad experiences. Finally, it presents and discusses initial results from the evaluation of 77 ad experiences (37 mobile, 40 desktop) and how the methodology was able to rank them.

Background

In Ad Experience Research Group 2016, we described an experimental methodology and survey tool that we developed to measure users' perception of ad experiences. We refer to "*ad experiences,*" rather than ads, because each experimental ad condition we tested was a combination of site layout and behavior, content, and one or more ads.

This methodology was based on two principles: *adopting the user's perspective* and *measuring ads in context*.

In the experiments, participants navigated to a content site on a mobile device and were asked to read an article. In the experimental conditions, the article contained one or more ads, while the control condition had no ads. After reading the article, participants were asked to rate their overall experience along three dimensions (satisfaction, predictability, and speed) and to rate the ad along seven dimensions (annoyance, usefulness, trust, visual appearance, distraction, inappropriateness, and creepiness).

The Single-Ad Experimental Framework

In the first version of the experiment, a between-subjects design, each participant saw only one condition: either one of three experimental ad conditions (sticky ad at the bottom of the

screen, inline animated ad, or a pop-up ad with a 10-second countdown) or the control condition. We found that survey ratings differed significantly in overall satisfaction and predictability scores between control and ad experiences. We also found significant differences among the three experimental conditions in annoyance and distraction: sticky ads were the least annoying and distracting; pop-up ads were the most annoying and distracting.

The Basic Multi-Ad Experimental Framework

To create a more efficient experimental paradigm that would also allow participants to rank ad experiences, we created a within-subjects variation of this paradigm: the Multi-Ad paradigm. In the Multi-Ad paradigm each participant read four web pages and answered survey questions after each: the control page had no ads, while the other three pages contained different ad experiences. At the end of the experiment, participants ranked the three ad experiences. We used the same three ad experiences used in the Single-Ad experiment, and we fully counterbalanced articles, ad creatives, and order of the four conditions across participants.

The Multi-Ad experimental framework confirmed the results of the Single-Ad experiment, and produced a ranking of the three ad experiences that was highly correlated with survey ratings. For the overall experience, each condition differed from the other in satisfaction and speed. Participants rated the control as most satisfying and fastest, followed by sticky ads then animated ads; pop-up ads were the least satisfying and were perceived as slowest. Ad experiences differed significantly in annoyance and distraction; other survey dimensions also detected significant differences across ad experiences.

The table below shows the average satisfaction and annoyance ratings for the three ad experiences in both Single-Ad and Multi-Ad experiments as comparison. Ad experiences with higher annoyance scores in the Single-Ad experiment (animated ads and pop-up ads) were also rated as more annoying in the Multi-Ad experiment, although rating differences between ad experiences were larger in the Multi-Ad experiment than in the Single-Ad experiment. We believe this is due to a combination of anchor effect (rating of the first ad experience influences ratings that follow) and fatigue (higher annoyance ad experiences were rated more annoying when they were shown toward the end of the experiment).

	Single-Ad Experiment		**Multi-Ad Experiment**	
Ad Experience	**Average Satisfaction**	**Average Annoyance**	**Average Satisfaction**	**Average Annoyance**
Sticky Ad	4.57	1.70	4.23	1.74
Animated Ad	4.56	2.44	3.80	2.86
Pop-up Ad	4.40	3.39	3.56	4.04

Table 1. Ad experiences ratings from previous single-ad and multi-ad experiments.

Average ranking scores in this experiment matched the survey rating scores: sticky ads were ranked the best and pop-up ads the worst.

Experiment Overview

In this paper we present two new experiments we ran using the extended multi-ad framework described below — one on desktop and one on mobile devices — in which we tested the 77 ad experiences described in the next section. Each participant saw only desktop ad experiences or only mobile ad experiences.

Our experiments used the same within-subject experimental design we validated in the original Multi-Ad experiment. See the Experiment Methodology section for more details.

Extending the Multi-Ad Experimental Framework

We further extended the framework to enable the testing of many ad experiences. The experimental flow was identical to what we used in the Multi-Ad experiment described above, but we now arranged our study to run in two phases, not one.

In both phases we aggregated data across all ad experiences by adopting an iterative sampling algorithm to select which ad experiences to show to each participant, and by using the Bradley-Terry algorithm to compile ranking data across participants into a single ranking scale.

Phase 1: A staged approach to creating the initial ranking

We broke our participants in this phase into three groups, and assigned each group to one of three stages of the experiment. In the first stage, we created a rough ranking of all the ad experiences we initially wanted to test. In the following two stages, we refined this ranking by showing participants ads that were ranked close together in the previous stages. (See the Separating Phase I Experiences into Three Stages section for more details.)

Phase 2: Adding new ad experiences to the ranking

In Phase 2, we showed participants new ad experiences paired with ad experiences we had already tested in Phase 1, and used their comparisons to add the new ad experiences to the initial ranking. This resulted in our final ranking of 37 mobile ad experiences.

Applying the Framework to Desktop

Our original multi-ad experiments tested 3 mobile ad experiences. We duplicated the same framework to test 40 desktop ad experiences. We created a simulated desktop news site

consisting of a simulated brand and real content derived from a high-circulation, award-winning global news publisher (see figure 1 for an example), using the two column layout that is common across the web. Our methodology was identical to what we used to evaluate mobile ad experiences, other than changing the layout of our article page and survey pages, and selecting desktop-appropriate ad experiences.

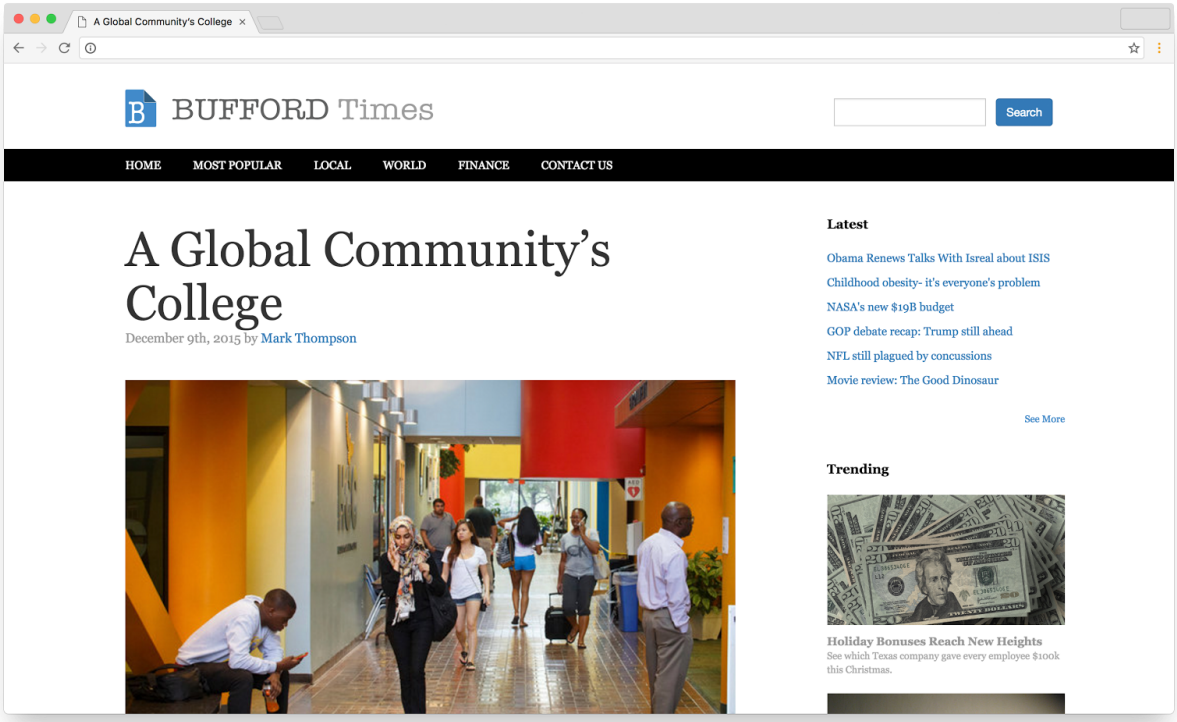


Figure 1. Example of the Desktop two-column layout used to test ad experiences.

We piloted the desktop study by testing six desktop ad experiences (bottom sticky ad, static inline ad, page-reflowing ad, animated ad, popup no countdown, 50% ad density). The results are in the table below.

Ad Experience	Overall (Relative) Rank Score	Average Annoyance	Average Satisfaction
Page-reflowing Ad	5.00	1.59	4.04
Static inline Ad	4.16	1.89	4.09
Animated Ad	3.65	2.27	3.89
50% density Ad	3.60	2.18	3.72
Bottom sticky Ad	2.38	3.06	3.75
Popup no countrown Ad	1.00	3.64	3.79

Table 2: Results from the first desktop multi-ad experience study

Experimental Methodology

We conducted two experiments — one experiment on mobile devices and one experiment on desktop devices. These experiments were conducted separately, meaning each participant saw only desktop ad experiences or only mobile ad experiences. The two experiments used the same methodology. This section details that methodology and highlights the few differences between the two new experiments and those described in the previous paper (Ad Experience Research Group 2016). The complete list of ad experiences tested is in the appendix, together with a description and a link to the live version.

Experiment Design

We utilized a within-subject experimental design in which each participant read four articles, three of which contained ads (experimental ad conditions) and one which did not (control condition). After reading each article, the participants answered three general questions about their experience on the webpage (e.g., "How satisfied were you with the overall experience on the web page?"). When the article they read contained an ad, participants next saw a survey page containing a screenshot of the article they just read and answered seven questions about that ad (e.g., "How annoying was the ad?"). At the end of the experiment, participants saw a page with screenshots of the three articles that contained ads. They were then asked to select which webpage they preferred least and which webpage they preferred most (ranking survey). The full text of these surveys is listed in the Appendix.

We limited each user to three ad experiences (four total articles) to minimize user fatigue and the number of ad experiences a user would have to remember when surveyed.

Experiment Staging

We split the 37 mobile ad experiences and 40 desktop ad experiences into two phases: Phase I (creating the initial ranking) and Phase II (adding new ad experiences to that ranking). Phase I included 17 mobile ad experiences and 18 desktop ad experiences. Phase II included 20 mobile ad experiences and 22 desktop ad experiences.

Within each phase, we developed a staged process to optimize the distribution of ad experiences shown to each participant. This was important, since we were testing a large number of ad experiences, but each of our 4,177 participants would see only three. If we were to simply randomize which experience was shown to each participant, a larger sample size would be needed.

Phase 1: Create the initial ranking

We divided our participants into three groups, or stages, in order to maximize the discerning power of our experiment with the smallest number of participants:

> **Stage 1** - We generated a set of ad experience groups that ensured a balanced number of between-experience pairwise comparisons. For example, we ensured that Pop-Up was assigned in a group along with Anchor just as much as with Flashing Animation. We assigned each of our participants to one of these ad experience groups. We then analyzed the data from Stage 1 using the B-T algorithm to generate a rough ranking of ad experiences (see *Estimating Overall Rank Score using B-T* in *Appendix*).
>
> *In stage 1 we used about 600 mobile participants and 800 desktop participants. Different numbers were used, as mobile and desktop stack ranks results differed, leading us to require more desktop participants to get reasonable confidence intervals.*
>
> **Stage 2** - We refined the initial ranking from Stage 1 by studying groups of ad experiences that were no more than five ranks apart in the ranking. For example, we re-paired experiences that were ranked third and fifth in the first stage, but not those that were ranked third and ninth. We then analyzed the data from Stages 1 and 2 using the B-T algorithm to come up with a more refined ranking.
>
> *In stage 2 we used about 600 mobile participants and 800 desktop participants.*
>
> **Stage 3** - We further refined the Stage 2 ranking by studying groups of ad experiences that were adjacent in our existing ranking. That is, their rank differences were not more than three.
>
> *In stage 3 we used about 700 mobile participants and 1000 desktop participants.*

Phase II: Adding new ad experiences to the ranking
In order to leverage ranking information gathered from Phase I, we showed participants In Phase II a new set of ad experiences. Each participant rated two new ad experiences from Phase II and one old ad experience from Phase I. The results allowed us to add new ad experiences to the initial ranking, creating a single stack rank. We chose Phase I experiences in a way that made it easier to find a new ad experience's position in the ranking.

For example, we paired the two Phase II ad experiences FSI_LARGEAD (Full-screen inline w/ large ad) and JANK_2FPS (Ad causing 2 frame-per-second scrolling) with the Phase I ad experience STATIC_INLINE (Static inline ad). Separately, we paired them with POPUPNOCD (Popup ad w/o countdown), and other ad experiences that were ranked between STATIC_INLINE and POPUPNOCD. This resulted in our final ranking of 37 mobile ad experiences and 40 desktop ad experiences.

Through this iterative experimental design, sample size was effectively used to achieve higher statistical power.

Using the Bradley-Terry Algorithm (B-T)
In order to provide a single ranking made from combining the rankings of many ad experience sets, we used the Bradley-Terry (B-T) Algorithm (Turner and Firth, 2012). The B-T Algorithm is a statistical algorithm that combines many pairwise comparisons into a single ranking, eliminating the need to have every participant review and rank all ad experiences. This approach allowed us to determine whether there is a significant preference difference between two ad experiences or whether the differences we observe was noise.

Study Participants
We used Answers Research to recruit 5295 US desktop users and 4259 US mobile users. We chose to use a vendor to recruit study participants because of the large and diverse population it provides, allowing us to study a representative sample of the United States internet-using population. The vendor recruits participants by sending out emails to internet users who have registered in their system. Each email informed potential participants the study would take around 15 minutes and included a link for them to take the study. Information on the breakdown of participant demographics can be found in the appendix "Demographic Distribution".

Material
These experiments varied ad experiences while using a similar set of ad creatives, text articles, and survey questions as the experiments described in our previous paper (Ad

Experience Research Group, 2016). Participants saw each article and each creative only once during the experiment, and articles and creatives were counterbalanced across the experimental conditions. The desktop and mobile experiments used the same set of ad creatives, text articles, and survey questions, although their formats were modified to fit each device .

Ad creatives

We built three ad creative concepts using fictional automotive, furniture, and insurance brands. We chose these advertiser verticals because they were likely to have broad appeal to our participants. The creatives have a medium-high level of visual polish.

We designed multiple variations of each creative concept to fit the sizes required by the 77 ad experiences tested in the experiment (see Figure 2). The pairing of creatives and ad experiences was fully counterbalanced across participants, so each participant saw a creative from each fictional advertiser (Sun Goose, Eagles Insurance, and Kosoi) during their study. This ensured that ratings and rankings of ad experience could not be attributed to a specific ad creative.

Figure 2. Examples of the three ad creatives.

Articles

We used a news article reading task so that participants would see the ad while engaged in a typical online activity. We used four different articles of around 400 words each and at an approximately 13-14 year old reading level. The pairing between article, experimental condition, and creative was fully counterbalanced across participants.

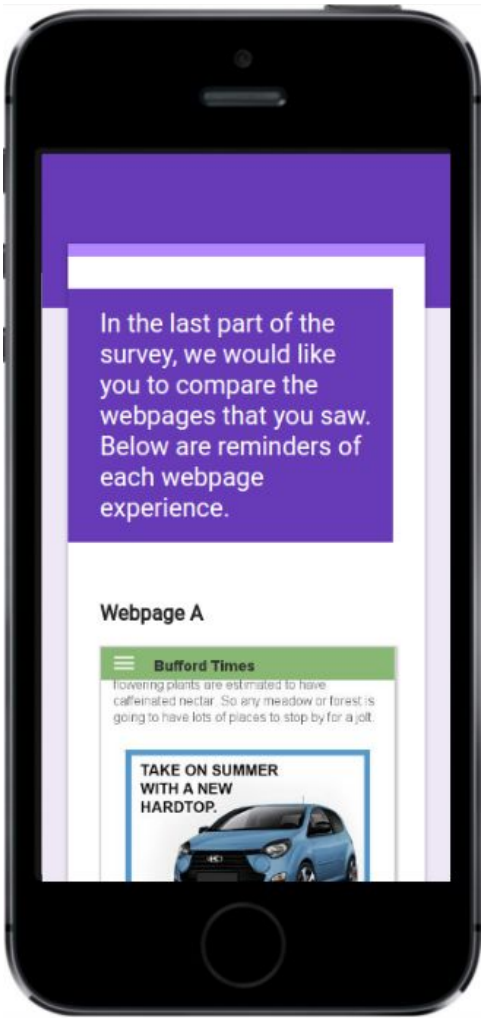
Comprehension, Overall Experience, and Ad-specific Questions

Participants answered a comprehension question after they read each article and their responses were used to confirm that they read the articles. Participants' reading comprehension accuracy was reasonably high (68.8%), suggesting most read the articles.

After each article, participants answered overall experience and ad-specific questions (if they just read an article with an ad experience). The overall experience questions covered satisfaction, speed, and predictability. The ad-specific questions covered annoyance, usefulness, trustworthiness, visual pleasantness, distraction, inappropriateness, and creepiness).

Final Ranking Survey

After reading the four articles and filling out the surveys, participants completed a ranking exercise. In this exercise, they saw screenshots of the three ad experiences they were exposed to earlier to refresh their memory, then selected which experience they most and least preferred.



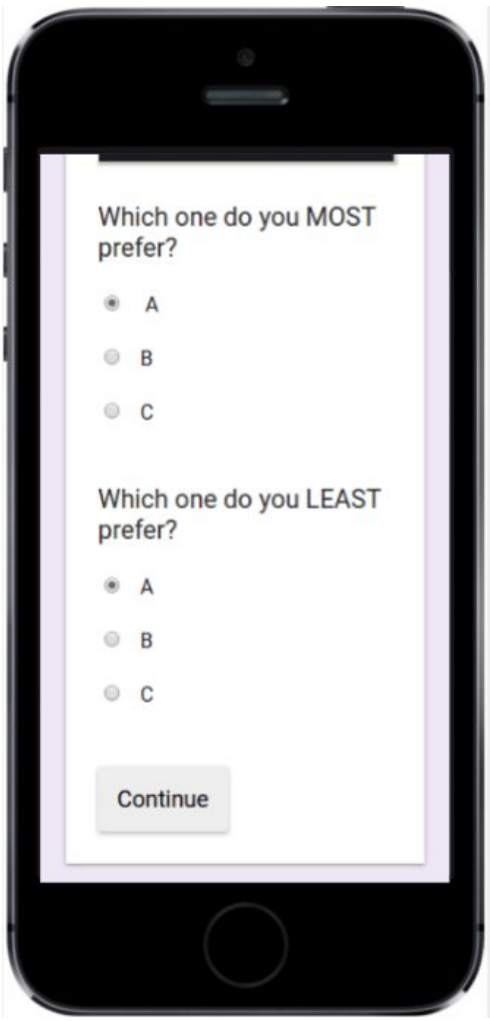


Figure 3. The top and bottom of the final ranking survey in the multi-ad experiment

Results and Discussion

Using Ad Experience Ratings to Better Understand Preference Rankings

In this section we'll first show the overall ranking results and then present and discuss the important metric scores.

While the rankings give us a strong overall signal on user preferences, the ad experience metric scores are valuable for understanding why a certain ad experience is ranked high or low (Is it distracting? Is it perceived as slow or unpredictable?). The metric scores are also a good measure of the overall user experience.

The Overall Ranking Chart

Each ad experience has a rank score estimate with a 95% confidence interval. In order to normalize the preferences, we assigned a score of 1 to the least preferred ad experience and a score of 5 to the most preferred ad experience.

For ads where the error bars are overlapping, the overall rank scores are similar. This means that participants' preferences for these ads do not significantly differ.

Overall mobile and desktop ranking charts are shown below in their respective sections.

Mobile Results

Pop-Up With Countdown ranked as the worst ad experience followed closely by 50% Ad Density. Anchor ranked as the best ad experience, significantly better than even its closest competition for the best spot: Animated Anchor. In addition, the graph reveals several other insights into how users feel about ad experiences.

Mobile Ad Experience Ranking

Figure 4. Overall Rank Score of the 37 mobile ad experiences, where 5 is the most favored experience and 1 is the least favored

Making the user wait makes an ad experience rank worse
We evaluated three different pairs of ad experiences that included a countdown and a no-countdown version: Pop-Ups, Prestitials and Poststitials, and in every case users ranked the with-countdown experience as worse than the no-countdown experience. This matched our expectations.

Timing matters: Pop-Ups are worse than Postitials
The Pop-Up ad experience uses an identical user interface to the Postitial ad experience but is rated significantly worse. The difference between the two experiences is when the ad appears: the Pop-Up appears 10 seconds after page load while the Postitial appears after a user finishes reading an article, then clicks a link to leave the page. We hypothesize this is due to the fact that the Pop-Up ad experience distracts a user while they are in the middle of reading the article.

Annoyance and Distraction affect rating differences the most across all ad experiences
Of the 10 metrics we collected, annoyance and distraction were most predictive of an ad experience's ranking. There were some outliers, however, meaning that experience preference is not entirely determined by these two metrics.

Annoyance and Distraction Metrics
In the two figures below, ad experiences are ordered according to their position in the overall rank (from the most preferred to the least preferred). The stack bars represent the participant ratings on a 1-5 scale for each ad metric.

Annoyance and distraction are the two metrics that show the largest rating differences across the 37 ad experiences. For the most part, the annoyance and distraction ratings are aligned with the overall preference ranking; ad experiences like Pop-Up with 10-second Countdown and Density 50 have the worst annoyance ratings, and Sticky ads have amongst the best.

Mobile Ad Experience Annoying Ratings

Figure 5. Annoyance rating distribution for each ad experience in the multi-ad mobile web experiment.

Mobile Ad Experience Distraction Ratings

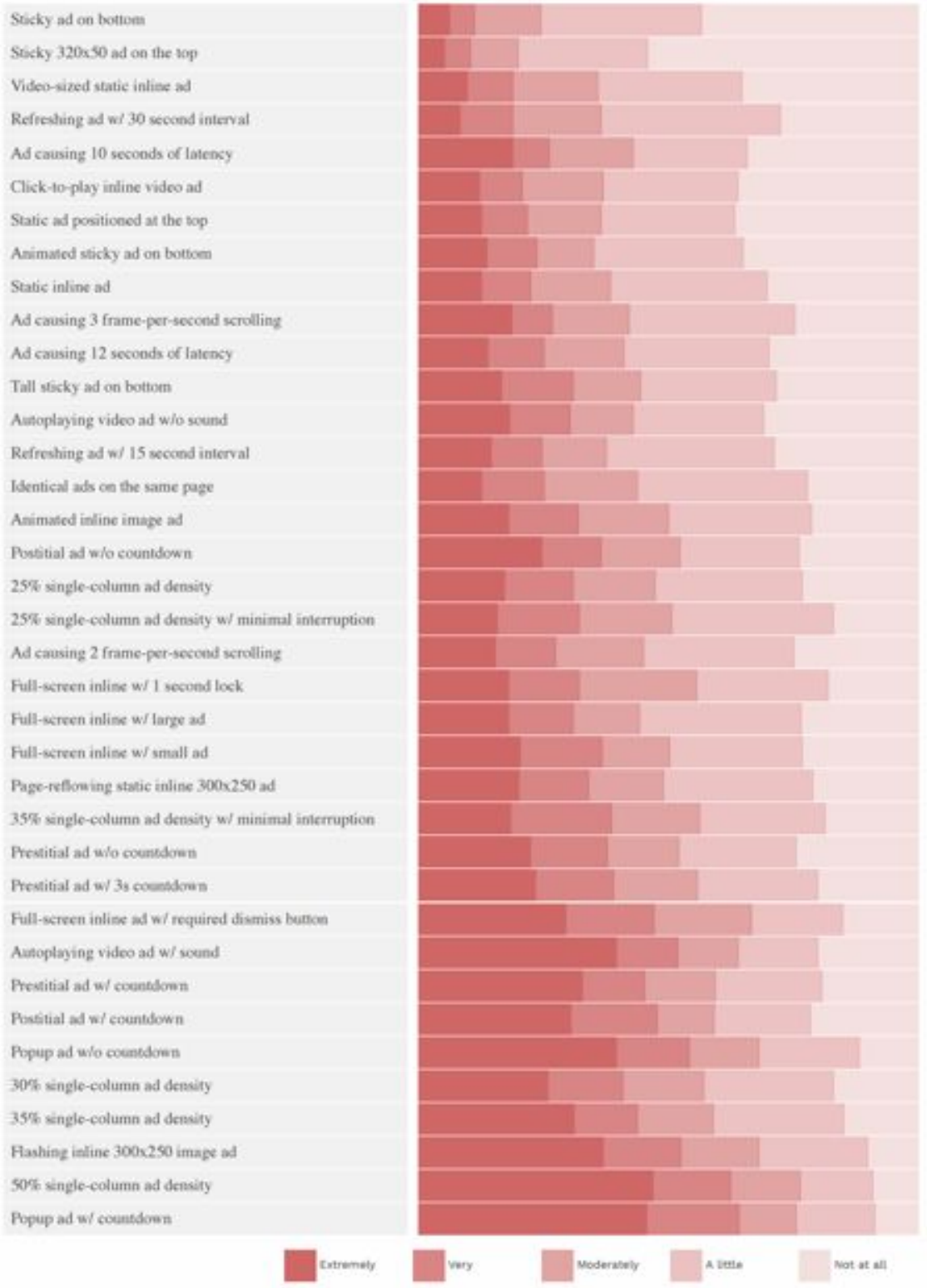


Figure 6. Distraction rating distribution for each ad experience in the multi-ad mobile web experiment.

The Other Metrics: Satisfying, Predictable, Fast, Useful, Trustworthy, Visually Pleasing, Inappropriate, and Creepy

For all of the metrics besides annoyance and distraction, **most** ads rate similarly to each other. There were a number of interesting takeaways from the ad experiences that rate differently from the rest. (The graphs visualizing these metrics can be found in the appendix.)

Usefulness and trustworthiness metrics did not change across ad experiences except for the most annoying ones

Pop-Ups, Density 50%, and Flashing Animated ads are some of the worst formats we tested according to their preference ranking and their annoyance and distraction metrics. Compared to other ad experiences, these formats also are rated worse in usefulness and trustworthiness scores than a lot of other ad experiences. This is especially interesting for the Pop-Up experiences because usefulness and trustworthiness are attributes one might expect to be related to the ad creative, not to the ad format. Participants in our studies perceived the Pop-Ups to be less useful and less trustworthy, even though they used the same creatives as ad experiences with less obtrusive placement. *Users' trust and perception of utility from a given creative is affected by the placement of that creative*.

High ad density experiences were rated as slower and less satisfying than similarly-ranked experiences

50% and 35% Density stand out in this regard — they affect the webpage metrics differently than how they affect the ad-specific and overall preference metrics. 50% Density rates significantly worse than Pop-Up with Countdown in terms of speed and satisfaction, even though Pop-Up with Countdown rank worse than 50% Density on the final preference question; the same flip occurs for 35% Density vs Flashing Animation. This seems to indicate that when a mobile webpage has high ad density, visitors are unsatisfied by their experience on that webpage but might not attribute that feeling to the ads.

Not all animation is disfavored by users. Reasonable animations are visually pleasing, but flashing animations are not

Despite using the same basic creative and same advertisers, Flashing Animated ads are rated as much less visually pleasing than their standard animating counterparts (flashing animations are those that quickly shift between 2-3 different states, creating a visual effect that makes the ad look like it's "flashing" on and off). This shows poor animations can impact users' perceptions of the ad overall. It also calls out that not all animation is bad — standard animating ads are considered equally visually pleasing as static inline ads.

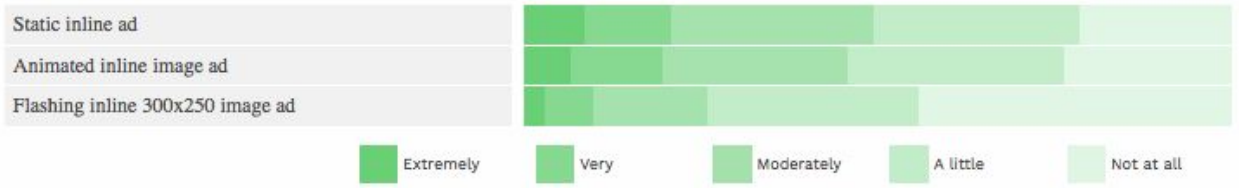


Figure 7. Visually Pleasing metrics for STATIC_INLINE, ANIMATED_300x250, and FLASHING_ANIMATION

Desktop Results

Pop-Up with Countdown ranked as the worst ad experience followed by Pop-Up without Countdown. Low ad density and ads that were not placed inline with the content were generally the most preferred. The full preference rank graph is below:

Desktop Ad Experience Rankings

Figure 8. Overall Rank Score of the 40 desktop ad experiences, where 5 is the most favored and 1 is the least favored

Countdowns make an ad experience worse
With Popups and Prestitials, the variation of the ad experience with a countdown rank worse than the variation without. This result matches the results from the mobile experiences.

Placement makes a big difference in formats on desktop
Sticky ads that are placed on top rank much better than sticky ads that are placed on the bottom of a desktop page. This differs from the mobile experiences, where a sticky ad on the bottom rank the best.

Annoyance and Distraction Metrics

In the two figures below, ad experiences are ordered according to their position in the overall rank (from the most preferred to the least preferred). The stack bars represent participants' ratings on a 1-5 scale for each ad metric. (We've included some of the ad metrics below. Graphs of other experiences can be found in the appendix.)

As with mobile, annoyance and distraction are highly correlated to each other and to the overall ranking.

Desktop Ad Experience Annoyance Ratings

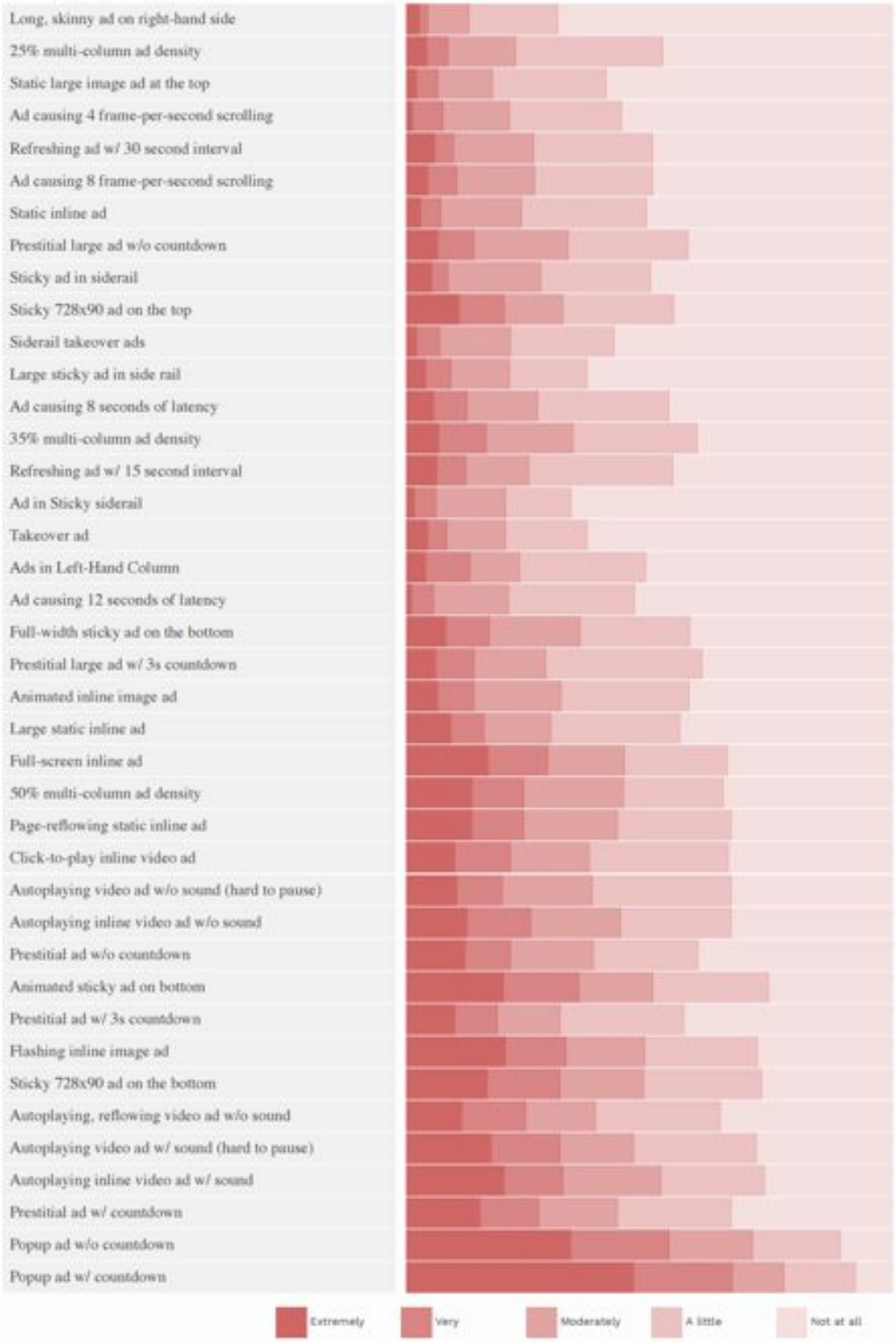


Figure 9. Annoyance rating distribution for each ad experience in the multi-ad desktop web experiment.

Desktop Ad Experience Distraction Ratings

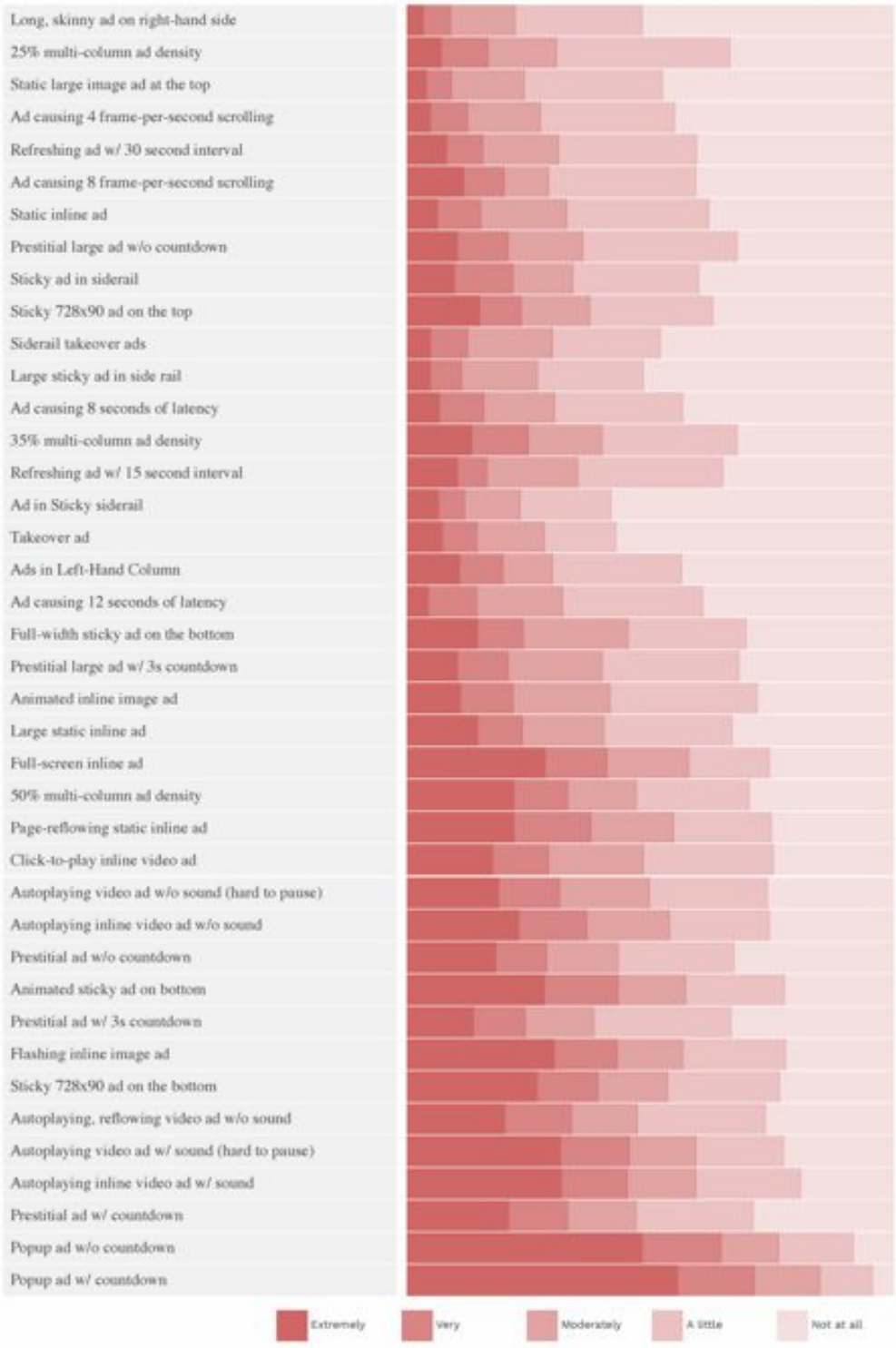


Figure 10. Distraction rating distribution for each ad experience in the multi-ad desktop web experiment.

Pop-ups on desktop are especially bad

Whether or not they were immediately dismissible, both pop-up conditions were rated as significantly more annoying and distracting than other experiences. This differed from the mobile results. Pop-ups also have much worse ratings on other metrics (satisfaction, inappropriate). This is despite using the same creatives as the other ad experiences, showing once again that the format has an impact on the perception of the ad creative.

Sticky Ads that are not full width are more annoying than those that are

We tested the same 728x90 creatives in desktop stickies in two different ways. In one, the normal content of the page was visible to the left and right of the creative; in the other, the left and right sides of the ad are filled in with black. We hypothesize that full width ads are less annoying because when the ad only takes up a portion of the bottom part of the screen it is perceived as covering content. If the ad takes up the full width of the browser (as it does on mobile), the ad appears to be a more natural part of the website's design.

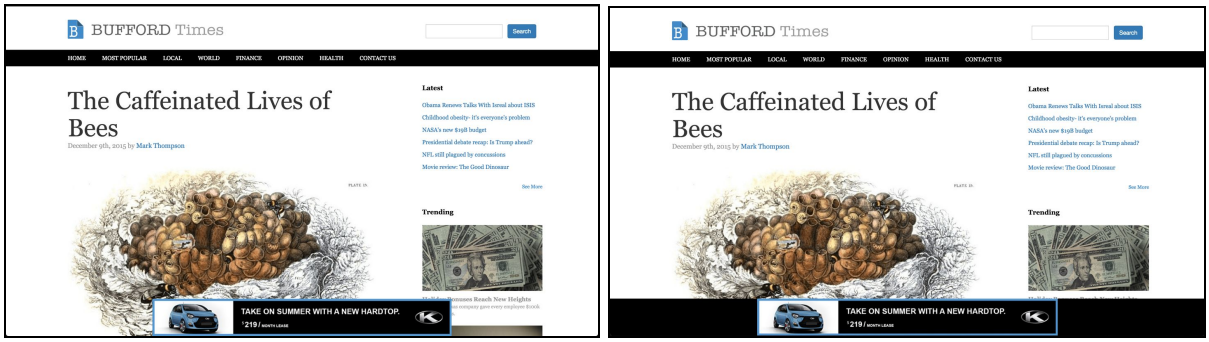


Figure 11. Difference between a desktop sticky and a full width desktop sticky

The Other Metrics: Satisfying, Predictable, Fast, Useful, Trustworthy, Visually Pleasing, Inappropriate, and Creepy

Other metrics were not as sensitive as annoyance and distraction - however, useful and trustworthy ratings were significantly lower for ad experiences rated as *really* annoying or distracting, as seen on mobile ad experiences.

Comparing Mobile and Desktop

In most cases, ad experiences that were shared between mobile and desktop ranked similarly in both environments.

Sticky ad experience ratings differ in mobile and desktop environments.

Both bottom sticky experiences on desktop rank far worse than the mobile sticky experience. The 728x90 sticky bottom ad that doesn't take up the whole width of the browser ranked as the 5th worst experience on desktop, but it is the best on mobile.

Ad Density ratings differ in mobile and desktop environments.
Participants were more sensitive to ad density on mobile than they were on desktop. We hypothesize that this was due to how ads are laid out in the two environments. In mobile, the ads are all laid out within the content. On desktop, the ads take positions on the left and right sidebars as well as within the content.

Conclusions

The methodology described in this paper was validated by rating and ranking 77 ad experiences at scale across platforms. By applying an algorithm to optimize which ad experiences were shown to each participant, and the Bradley–Terry model to analyze the resulting data, we were able to efficiently rank ad experiences by preference. The results imply that specific aspects of ad experiences can affect their preferability among a representative sample of the population.

The results also provide new insights into how a publisher's ad format, the format of the creative contained within, and the overall density of ads on a page can affect user experience and sentiment.

Our methodology allows for the relative determination of user preferences between different experiences. As a whole, we found that experiences that block or distract users by virtue of their density, use of a countdown timer, or use of overlayed placement not part of a natural user journey performed poorly across the measured criteria. We also found that users can perceive an ad creative as less useful or more creepy if that creative appears in a more obtrusive ad placement.

Furthermore, when we investigated the relationship between user-provided scores on 10 absolute dimensions (e.g., annoyance, distraction, satisfaction), we found that these components explained 92% of the preference variation in the ranking scores. This shows that for formats, these metrics effectively explain user preferences with ad formats.

The methodology described in this paper is scalable across experiences and platforms and should be replicable by publishers, advertisers, or other organizations that have a stake or interest in the ads ecosystem. These benefits come from:

1. **An experiential study structure**
   Because the methodology is experiential in nature, it can measure any experience that can be applied to the task asked of participants (in this example, reading an article). This study structure can also be applied to different tasks (ex: watching videos) while keeping its core design.

2. **A phased design for analysis**
   The framework we chose can continue to accept new data and expand its results. It is able to continue accepting new ad experiences into the stack rank.

Next steps

We plan to expand the set of experiences we test, specifically:

1. **Continue to expand the set of ad experiences to be tested on mobile web**
   To date, we have identified over 200 representative, isolated experiences that can be tested.
2. **Test across geographical regions**
   Geography and culture may impact how users perceive ad experiences. We will be looking into testing these ad experiences in places other than the US, and have identified partners with a global user base.
3. **Test different user contexts, specifically video**
   User expectations for ads in video content or other contexts may differ when compared to ads in article content (e.g., autoplay ads may be expected, or standard-sized static formats may be used as overlays). We will look into modifying the framework to evaluate video content and video ad experiences.
4. **Determine if any ads are significantly worse than others**
   In future analyses with additional ad experiences, we will look for ways to determine which experiences are particularly more annoying and likely to cause users to abandon a site, create negative advertiser or publisher brand perceptions, or even install ad blockers.

References

Abdi H. (2010). Partial least squares regression and projection on latent structure regression (PLS-regression). *Wiley Interdisciplinary Reviews: Computational Statistics*, *2*(1), 97–106. doi:10.1002/wics.51

Goldstein, D.G., Suri, S., McAfee, R.P., Ekstrang-Abueg, M. & Diaz, F. (2014). The Economic and Cognitive Costs of Annoying Display Advertisements. *Journal of Marketing Research, LI,* 742-752.

Ad Experience Research Group (2016). *An experimental methodology to rank N ad experiences by consumers' perceptions. From* betterads.org/research/rankingpaper/.

Turner, H., & Firth, D. (2012). Bradley-Terry Models in R: The BradleyTerry2 Package. *Journal of Statistical Software*, *48*(9). doi:10.18637/jss.v048.i09

Appendix

The survey instrument

Demographic Questions
Before you begin, we have a few questions about your background.
Age

- 18-21 years old
- 22-34 years old
- 35-44 years old
- 45-54 years old
- 55-64 years old
- 65+ years old

Gender

- Male
- Female
- Other

To what extent do you agree with the following statement?
Overall, I consider online advertising a good thing.

- Strongly agree
- Agree
- Neither agree nor disagree
- Disagree
- Strongly disagree

What is your current employment status?

- Employed
- Out of work
- Student
- Retired
- Other

What is your annual household income before taxes?

- Less than $30,000
- $30,000 to $49,999
- $50,000 to $99,999
- $100,000 to $149,999
- $150,000 or more

- Prefer not to answer

Overall Experience Questions

How satisfied were you with the OVERALL EXPERIENCE on the web page?

- Very satisfied
- Somewhat satisfied
- Neither satisfied nor dissatisfied
- Somewhat dissatisfied
- Very dissatisfied

To what extent does each of the following words describe your experience on the WEB PAGE?

Predictable

- A great deal
- A lot
- A moderate amount
- A little
- Not at all

Fast

- A great deal
- A lot
- A moderate amount
- A little
- Not at all

Ad-specific questions

Participants only saw this section after reading an article that contained an ad in it. We showed a screenshot of the article they just read, with the ad highlighted, to refresh their memory of the ad.

We would like to ask you about an ad (boxed in red) that you may have noticed in Webpage A.*

[Participant sees screenshot of the exposed ad experience with the ad highlighted in red]

How USEFUL was the ad?

- Extremely useful
- Very useful
- Moderately useful
- Slightly useful
- Not at all useful

How ANNOYING was the ad?

- Not at all annoying
- Slightly annoying
- Moderately annoying
- Very annoying
- Extremely annoying

To what extent does each of the following words or phrases DESCRIBE THE AD?

Trustworthy

- A great deal
- A lot
- A moderate amount
- A little
- Not at all

Visually pleasing

- A great deal
- A lot
- A moderate amount
- A little
- Not at all

Distracting

- Not at all
- A little
- A moderate amount
- A lot
- A great deal

Inappropriate

- Not at all
- A little
- A moderate amount
- A lot
- A great deal

Creepy

- Not at all
- A little
- A moderate amount
- A lot

- A great deal

Ad Ranking Exercise

In the last part of the survey, we would like you to compare the ads that you saw.

[Participant sees screenshot of the exposed ad experience with the ad highlighted in red]

Which one do you LEAST prefer?

- A
- B
- C

Which one do you MOST prefer?

- A
- B
- C

Ad Experiences We Tested

Mobile Web Ad Experiences

Mobile Web Experiences Tested in a Text and Photo Site Context		
Name	**Description**	**Link**
Static ad positioned at the top	A 300x250 image ad is placed above the article content.	http://poetic-glass-136423.appspot.com/experience?exp=STATIC_TOP&advertiser_id=1&article_id=31
Static inline ad	A 300x250 image ad is placed between two paragraphs in the article.	http://poetic-glass-136423.appspot.com/experience?exp=STATIC_INLINE&advertiser_id=1&article_id=31
35% single-column ad density	Image ads are evenly interspersed with content such that while reading the article, a user always sees ads as 35% of the content.	http://poetic-glass-136423.appspot.com/experience?exp=DENSITY35&advertiser_id=1&article_id=31
Flashing inline 300x250 image ad	A 300x250 animated GIF image ad is shown between two paragraphs of content. The animation flashes and has a lot of quick movement. It repeats every 1 second.	http://poetic-glass-136423.appspot.com/experience?exp=FLASHING_ANIMATION&advertiser_id=1&article_id=31
Sticky ad on bottom	A 320x50 image ad is shown on the bottom of the user's screen - it stays there regardless of the user's scrolling.	http://poetic-glass-136423.appspot.com/experience?exp=ANCHOR&advertiser_id=1&article_id=31
Popup ad w/ countdown	A popup ad appears 5 seconds after the article loads. A full-page popup ad appears on top of the content with a 300x250 image ad. The popup ad has a 10 second timer, after which an easy-to-find close button appears (the ad can't be dismissed for 10 seconds).	http://poetic-glass-136423.appspot.com/experience?exp=POPUP&advertiser_id=1&article_id=31
Postitial ad w/ countdown	After a user completes the article, they are shown a full-page ad container with a 300x250 image ad. The ad has a 10 second timer, after which an easy-to-find skip button appears (the ad can't be dismissed for 10 seconds).	http://poetic-glass-136423.appspot.com/experience?exp=POSTITIAL_CD&advertiser_id=1&article_id=31

Tall sticky ad on bottom	A 320x100 image ad is shown on the bottom of the user's screen - it stays there regardless of the user's scrolling.	http://poetic-glass-136423.appspot.com/experience?exp=TALL_ANCHOR&advertiser_id=1&article_id=31
Animated sticky ad on bottom	A 320x50 animated image ad is shown on the bottom of the user's screen - it stays there regardless of the user's scrolling.	http://poetic-glass-136423.appspot.com/experience?exp=ANIMATED_ANCHOR&advertiser_id=1&article_id=31
Prestitial ad w/o countdown	Before the article loads, a user is shown a full-page ad container with a 300x250 image ad. The interstitial has a skip button and can be immediately dismissed.	http://poetic-glass-136423.appspot.com/experience?exp=PRESTITIAL_NOCD&advertiser_id=1&article_id=31
Animated inline image ad	A 300x250 animated GIF image ad is shown between two paragraphs of content. The animation is noticeable, but not blinking or flashing. It repeats every 5 seconds.	http://poetic-glass-136423.appspot.com/experience?exp=ANIMATED&advertiser_id=1&article_id=31
Page-reflowing static inline 300x250 ad	While a user is reading an article, a 300x250 image ad appears and causes the page's contents to reflow (the text shifts down in the viewport).	http://poetic-glass-136423.appspot.com/experience?exp=REFLOW&advertiser_id=1&article_id=31
25% single-column ad density	Image ads are evenly interspersed with content such that while reading the article, a user always sees ads as 25% of the content.	http://poetic-glass-136423.appspot.com/experience?exp=DENSITY25&advertiser_id=1&article_id=31
Postitial ad w/o countdown	After a user completes the article, they are shown a full-page ad container with a 300x250 image ad. The interstitial has a skip button and can be immediately dismissed.	http://poetic-glass-136423.appspot.com/experience?exp=POSTITIAL_NOCD&advertiser_id=1&article_id=31
50% single-column ad density	Image ads are evenly interspersed with content such that while reading the article, a user always sees ads as 50% of the content.	http://poetic-glass-136423.appspot.com/experience?exp=DENSITY50&advertiser_id=1&article_id=31
Prestitial ad w/ countdown	Before the article loads, a user is shown a full-page ad container with a 300x250 image ad. The ad has a 10 second timer, after which an easy-to-find skip button appears (the ad can't be dismissed for 10 seconds).	http://poetic-glass-136423.appspot.com/experience?exp=PRESTITIAL_CD&advertiser_id=1&

		article_id=31
Popup ad w/o countdown	A popup ad appears 5 seconds after the article loads. A full-page popup appears on top of the content with a 300x250 image ad. The popup ad has an easy-to-find close button and can be immediately dismissed.	http://poetic-glass-136423.appspot.com/experience?exp=POPUPNOCD&advertiser_id=1&article_id=31
Sticky 320x50 ad on the top	A 320x50 image ad is shown on the top of the user's screen - it stays there regardless of the user's scrolling.	http://poetic-glass-136423.appspot.com/experience?exp=ANCHOR_TOP&advertiser_id=1&article_id=31
Video-sized static inline ad	A 300x167 static image ad is placed between two paragraphs in the article.	http://poetic-glass-136423.appspot.com/experience?exp=VIDEOSIZEDSTATIC&advertiser_id=1&article_id=31
Ad causing 10 seconds of latency	A standard 300x250 image ad is placed inline with the content. Once the image ad would be in view, the ad and the content below it are forced to wait for 10 seconds before being visible.	http://poetic-glass-136423.appspot.com/experience?exp=LATENCY_INLINE_10&advertiser_id=1&article_id=31
Refreshing ad w/ 30 second interval	A standard 300x250 image ad is placed inline with the content. The ad is exchanged in-place with another ad every 30 seconds.	http://poetic-glass-136423.appspot.com/experience?exp=REFRESH_30&advertiser_id=1&article_id=31
Click-to-play inline video ad	A 300x167 video ad is placed between two paragraphs in the article. It is paused until a user decides to play the ad.	http://poetic-glass-136423.appspot.com/experience?exp=VIDEO_NOAUTOPLAY&advertiser_id=1&article_id=31
Ad causing 3 frame-per-second scrolling	A 300x250 image ad is placed inline with the content. The content is forced to have a jittery scroll, such that the scrolling only seems to occur at a pace of 3 frames per second (good performance is 60 frames per second).	http://poetic-glass-136423.appspot.com/experience?exp=JANK_3FPS&advertiser_id=1&article_id=31
Ad causing 12 seconds of latency	A standard 300x250 image ad is placed inline with the content. Once the image ad would be in view, the ad and the content below it are forced to wait for 12 seconds before being visible.	http://poetic-glass-136423.appspot.com/experience?exp=LATENCY_INLINE_12&advertiser_id=1&article_id=31
Autoplaying video ad w/o sound	A 300x167 video ad is placed between two paragraphs in the article. It autoplays on mute until a user decides to pause or unmute the ad.	http://poetic-glass-136423.appspot.com/experience?exp=VIDEO_AU

		TOPLAYMUTED&advertiser_id=1&article_id=31
Refreshing ad w/ 15 second interval	A standard 300x250 image ad is placed inline with the content. The ad is exchanged in-place with another ad every 15 seconds.	http://poetic-glass-136423.appspot.com/experience?exp=REFRESH_15&advertiser_id=1&article_id=31
Identical ads on the same page	Three 300x250 ads are interspersed with the content. The three ads all contain the same content.	http://poetic-glass-136423.appspot.com/experience?exp=DUPLICATES&advertiser_id=1&article_id=31
25% single-column ad density w/ minimal interruption	Image ads are placed together in a "chunk" within content such that while reading the article a user is only interrupted once. 25% of the total height of the content is ads.	http://poetic-glass-136423.appspot.com/experience?exp=DENSITY_CHUNKED_25&advertiser_id=1&article_id=31
Full-screen inline w/ 1 second lock	A 300x250 ad with dark padding surrounding it that takes up the full size of the screen is placed inline with the content. When a user scrolls it into view, the ad blocks the user from scrolling past the ad for 1 second. After 1 second, the user can continue scrolling the article.	http://poetic-glass-136423.appspot.com/experience?exp=FSI_1SECLOCK&advertiser_id=1&article_id=31
Full-screen inline w/ large ad	A 300x600 ad is surrounded by black padding, forcing the ad to take the full screen. It is placed inline with the content such that users can scroll through the ad.	http://poetic-glass-136423.appspot.com/experience?exp=FSI_LARGEAD&advertiser_id=1&article_id=31
Ad causing 2 frame-per-second scrolling	A 300x250 image ad is placed inline with the content. The content is forced to have a jittery scroll, such that the scrolling only seems to occur at a pace of 2 frames per second (good performance is 60 frames per second).	http://poetic-glass-136423.appspot.com/experience?exp=JANK_2FPS&advertiser_id=1&article_id=31
Full-screen inline w/ small ad	A 300x250 ad with dark padding surrounding it that takes up the full size of the screen is placed inline with the content such that users can scroll through the ad.	http://poetic-glass-136423.appspot.com/experience?exp=FSI_SCROLLABLE&advertiser_id=1&article_id=31
35% single-column ad density w/ minimal interruption	Image ads are placed together in a "chunk" within content such that while reading the article a user is only interrupted once. 35% of the total height of the content is ads.	http://poetic-glass-136423.appspot.com/experience?exp=DENSITY_CHUNKED_35&advertiser_id=1&article_id=31

Prestitial ad w/ 3s countdown	Before the article loads, a user is shown a full-page ad container with a 300x250 image ad. The ad has a 3 second timer, after which an easy-to-see skip button appears (the ad can't be dismissed for 3 seconds).	http://poetic-glass-136423.appspot.com/experience?exp=PRESTITIAL_CD3&advertiser_id=1&article_id=31
Full-screen inline ad w/ required dismiss button	A 300x250 ad with dark padding surrounding it that takes up the full size of the screen is placed inline with the content. When a user scrolls it into view, the ad blocks the user from scrolling past the ad. In order to continue reading, the user must press the close button on the ad.	http://poetic-glass-136423.appspot.com/experience?exp=FSI_NEEDSDISMISS&advertiser_id=1&article_id=31
Autoplaying video ad w/ sound	A 300x167 video ad is placed between two paragraphs in the article. It autoplays with sound until a user decides to pause or unmute the ad.	http://poetic-glass-136423.appspot.com/experience?exp=VIDEO_AUTOPLAYSOUND&advertiser_id=1&article_id=31
30% single-column ad density	Image ads are evenly interspersed with content such that while reading the article, a user always sees ads as 30% of the content.	http://poetic-glass-136423.appspot.com/experience?exp=DENSITY30&advertiser_id=1&article_id=31

Table 3: Mobile Web Experiences tested in a Text and Photo Site Context

Desktop Web Experiences

Desktop Web Experiences Tested in a Text and Photo Site Context		
Name	**Description**	**Link**
Static inline ad	A 300x250 image ad is placed between two paragraphs in the article.	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_STATIC&advertiser_id=1&article_id=31
Sticky 728x90 ad on the top	A 728x90 image ad is shown on the top of the user's screen - it stays there regardless of the user's scrolling.	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_TOPANCHOR&advertiser_id=1&article_id=31
Flashing inline image ad	A 300x250 animated GIF image ad is shown between two paragraphs of content. The animation flashes and has a lot of quick movement. It repeats every 1 second.	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_FLASHING_ANIMATION&advertiser_id=1&article_id=31

Autoplaying inline video ad w/ sound	A 640x390 video ad is placed between two paragraphs in the article. It automatically starts playing and sound is enabled on the player.	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_VIDEOAUTOPLAYSOUND&advertiser_id=1&article_id=31
Page-reflowing static inline ad	While a user is reading an article, a 640x360 image ad appears and causes the page's contents to reflow (the text shifts down in the viewport).	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_INLINE_REFLOW&advertiser_id=1&article_id=31
Sticky 728x90 ad on the bottom	A 728x90 image ad is shown on the bottom of the user's screen - it stays there regardless of the user's scrolling.	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_ANCHOR&advertiser_id=1&article_id=31
Static large image ad at the top	A 970x250 image ad is placed above the article content.	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_MASTHEAD&advertiser_id=1&article_id=31
35% multi-column ad density	Image ads are placed in the side rails and in content such that while reading the article, a user always sees ads as 35% of the content.[1]	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_DENSITY35&advertiser_id=1&article_id=31
Large static inline ad	A 640x360 static image ad is placed between two paragraphs in the article.	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_VIDEOSIZEDSTATIC&advertiser_id=1&article_id=31
Prestitial ad w/ countdown	Before the article loads, a user is shown a full-page ad container with a 300x250 image ad. The ad has a 10 second timer, after which an easy-to-see skip button appears (the ad can't be dismissed for 10 seconds).	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_PRESTITIAL_CD&advertiser_id=1&article_id=31
Popup ad w/o countdown	A popup ad appears 5 seconds after the article loads. A full-page popup appears on top of the content with a 300x250 image ad. The popup ad has an easy-to-see close button and can	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_PRESTITIAL_NOCD&advertiser_id=1&article_id=31

[1] See 'Calculating Ad Density' for more details.

	be immediately dismissed.	
50% multi-column ad density	Image ads are placed in the side rails (in a 3-column layout) and in content such that while reading the article, a user always sees ads as 50% of the content.	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_DENSITY50&advertiser_id=1&article_id=31
Prestitial ad w/o countdown	Before the article loads, a user is shown a full-page ad container with a 300x250 image ad. The interstitial has a skip button and can be immediately dismissed.	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_PRESTITIAL_NOCD&advertiser_id=1&article_id=31
Animated inline image ad	A 300x250 animated GIF image ad is shown between two paragraphs of content. The animation is noticeable, but not blinking or flashing. It repeats every 5 seconds.	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_ANIMATED&advertiser_id=1&article_id=31
Click-to-play inline video ad	A 640x390 video ad is placed between two paragraphs in the article. It is paused until a user decides to play the ad.	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_VIDEONOAUTOPLAY&advertiser_id=1&article_id=31
Popup ad w/ countdown	A popup ad appears 5 seconds after the article loads. A full-page popup ad appears on top of the content with a 580X400 image ad. The popup ad has a 10 second timer, after which an easy-to-see close button appears (the ad can't be dismissed for 10 seconds).	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_POPUPIMAGE&advertiser_id=1&article_id=31
Autoplaying inline video ad w/o sound	A 640x390 video ad is placed between two paragraphs in the article. It automatically starts playing, but is muted.	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_VIDEOAUTOPLAYMUTED&advertiser_id=1&article_id=31
25% multi-column ad density	Image ads are placed in the side rails and in content such that while reading the article, a user always sees ads as 25% of the content.	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_DENSITY25&advertiser_id=1&article_id=31
Long, skinny ad on right-hand side	A 120x600 ad that fills the left or right side of a publisher's page	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_RIGHT_SKINNYAD&advert

		iser_id=1&article_id=31
Ad causing 4 frame-per-second scrolling	A 300x250 image ad is placed inline with the content. The content is forced to have a jittery scroll, such that the scrolling only seems to occur at a pace of 4 frames per second (good performance is 60 frames per second).	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_JANK_4FPS&advertiser_id=1&article_id=31
Refreshing ad w/ 30 second interval	A standard 300x250 image ad is placed inline with the content. The ad is exchanged in-place with another ad every 30 seconds.	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_REFRESH_30&advertiser_id=1&article_id=31
Ad causing 8 frame-per-second scrolling	A 300x250 image ad is placed inline with the content. The content is forced to have a jittery scroll, such that the scrolling only seems to occur at a pace of 8 frames per second (good performance is 60 frames per second).	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_JANK_8FPS&advertiser_id=1&article_id=31
Prestitial large ad w/o countdown	Before the article loads, a user is shown a full-page ad container with a 800x750 image ad. The interstitial has a skip button and can be immediately dismissed.	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_PRESTITIAL_NEW&advertiser_id=1&article_id=31
Sticky ad in siderail	A 300x250 sticky ad on side rail. It stays on screen regardless of how much a user scrolls	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_SIDERAILSTICKY&advertiser_id=1&article_id=31
Siderail takeover ads	Two 300x600 ads are shown on the left and right-hand sides of the content.	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_TAKEOVER_SIDE&advertiser_id=1&article_id=31
Large sticky ad in side rail	A 300x600 sticky ad on side rail. The side rail stays on screen separately from the user's scroll.	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_SIDERAIL600&advertiser_id=1&article_id=31
Ad causing 8 seconds of latency	A standard 300x250 image ad is placed inline with the content. Once the image ad would be in view, the ad and the content below it are forced to wait for 8 seconds before being visible.	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_LATENCY_8&advertiser_id=1&article_id=31
Refreshing ad w/ 15 second interval	A standard 300x250 image ad is	http://poetic-glass-1364

	placed inline with the content. The ad is exchanged in-place with another ad every 15 seconds.	23.appspot.com/experience?exp=DESKTOP_REFRESH_15&advertiser_id=1&article_id=31
Ad in Sticky siderail	A 300x250 ad in a sticky side rail. The side rail stays on screen regardless of how much a user scrolls	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_SIDERAIL250&advertiser_id=1&article_id=31
Takeover ad	A massive ad takes the place of the page's background	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_TAKEOVER_FULL&advertiser_id=1&article_id=31
Ads in Left-Hand Column	Ads take up the entire left-hand column of the page's layout	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_LEFT_ADCOLUMN&advertiser_id=1&article_id=31
Ad causing 12 seconds of latency	A standard 300x250 image ad is placed inline with the content. Once the image ad would be in view, the ad and the content below it are forced to wait for 12 seconds before being visible.	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_LATENCY_12&advertiser_id=1&article_id=31
Full-width sticky ad on the bottom	A 728x90 image ad is shown on the bottom of the user's screen - it stays there regardless of the user's scrolling. The rest of the width of the page on either side of the ad is covered by black bars.	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_ANCHOR_FW&advertiser_id=1&article_id=31
Prestitial large ad w/ 3s countdown	Before the article loads, a user is shown a full-page ad container with a 800x750 image ad. The ad has a 3 second timer, after which an easy-to-see skip button appears (the ad can't be dismissed for 3 seconds).	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_PRESTITIAL_NEWCD3&advertiser_id=1&article_id=31
Full-screen inline ad	A Full-screen ad is placed inline with the content such that users can scroll through the ad.	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_FSI&advertiser_id=1&article_id=31
Autoplaying video ad w/o sound (hard to pause)	A 640x390 video ad is placed between two paragraphs in the article. It	http://poetic-glass-136423.appspot.com/experie

	autoplays on mute until a user decides to pause or unmute the ad. The video ad cannot be paused by clicking on the center of the video.	nce?exp=DESKTOP_NEWVIDEO_AUTOPLAYMUTED&advertiser_id=1&article_id=31
Animated sticky ad on bottom	A 728x90 animated image ad is shown on the bottom of the user's screen - it stays there regardless of the user's scrolling.	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_ANCHOR_ANIMATED&advertiser_id=1&article_id=31
Autoplaying video ad w/ sound (hard to pause)	A 640x390 video ad is placed between two paragraphs in the article. It autoplays with sound until a user decides to pause or mute the ad. The video ad cannot be paused by clicking on the center of the video.	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_NEWVIDEO_AUTOPLAYSOUND&advertiser_id=1&article_id=31
Prestitial ad w/ 3s countdown	Before the article loads, a user is shown a full-page ad container with a 300x250 image ad. The ad has a 3 second timer, after which an easy-to-see skip button appears (the ad can't be dismissed for 3 seconds).	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_PRESTITIAL_CD3&advertiser_id=1&article_id=31
Autoplaying, reflowing video ad w/o sound	A 640x390 video ad slides into view when a user scrolls to where the video ad would be. It is placed between two paragraphs in the article. It autoplays on mute until a user decides to pause or unmute the ad. The video ad cannot be paused by clicking on the center of the video.	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_VIDEOREFLOW&advertiser_id=1&article_id=31

Table 4: Desktop Web Experiences tested in a Text and Photo Site Context

Estimating Overall Rank Score Using Bradley-Terry Algorithm

The data we collected are all individual-level ranks of three ad combinations, out of about 20 ad experiences (17 in mobile, 18 in desktop). The total number of ad experiences is not directly relevant to the analysis. The information creates a pool of pair-comparison data that fits with the Bradley-Terry model and consolidates all ad comparison results; it also provides parameter estimates of Overall Rank Score, indicating how ad experiences are preferred relative to each other.

Bradley-Terry Model

In this application, ads are in a "contest" judged by a participant. We assume that α_i/α_j is the odds that Ad i beats Ad j, where α_i and α_j are parameters that can be thought of as representing participants' "preference." The model can then be expressed in the logit-linear form

$$\text{logit}\{P(i \text{ beats } j)\} = log\frac{\alpha_i}{\alpha_j} = \lambda_i - \lambda_j$$

where $\lambda_i = log\alpha_i$ for all i. Thus, assuming independence of all contests, the parameters $\{\lambda_i\}$ can be estimated by maximum likelihood. That is, it maximizes the likelihood of the observed data, which can be solved through generalized linear model. The observations required are the outcomes of ads comparison from participants. For example, summarizing these outcomes as w_{ij}, the number of times Ad i has beaten Ad j, we obtain the log-likelihood of the parameter $\{\alpha_i\}$:

$$L(\alpha) = \sum_i^n \sum_j^n \{w_{ij} log\alpha_i - w_{ij} log(\alpha_i + \alpha_j)\}$$

BradleyTerry2 package in R is implemented to facilitate the specification and model fitting.

Rescaling

The experimental study is multi-phase: we add new ad experiences (up to 20), test them, and add them to the ranking. In every phase we run surveys in multiple stages to be efficient with a limited sample size. In order to create consistent and meaningful metrics that accurately represent the relative preference of each ad experience, we rescale the estimates of *preference* $\{\alpha_i\}$ to a score interval between 1 and 5, where 1 is the LEAST preferred score and 5 is the MOST preferred. The two ad experiences ranked best and worst are pinned to these two scores as two anchors, so all the other experiences follow a score in between. It is defined to be the **Overall Rank Score**, with

interpretable meaning and consistency across studies.

One side note is that the score could fall out of the range of [1, 5] if a newly added ad experience is ranked worse or better than the two anchor ads.

PLS Regression - Overall Rank Score on 10 UX Metrics

In order to confirm that Overall Rank Score is a valid single metric, we would like to model its relationship with the UX metrics. The goal is to check if the relation is as expected. In hypothesis, some of the UX metrics should contribute more heavily than the others. For example, annoying should be highly correlated with the overall rank, while creepy not so much.

However, the pairwise correlation is very high in some of the metrics, such as annoying and distracting. See Figure 12 for the pairwise correlation. Directly fitting a multiple regression will cause the issue of multicollinearity. Therefore, Partial Least Squares (PLS) regression (Abdi, 2010) comes to play and it combines the technique of Principal Component Analysis (PCA) and multiple linear regression with the intention of dimension deduction. PLS finds components that explain the covariance between predictors and response (i.e., latent vectors). The components are a linear combination of 10 UX metrics (see Figure 13 below).

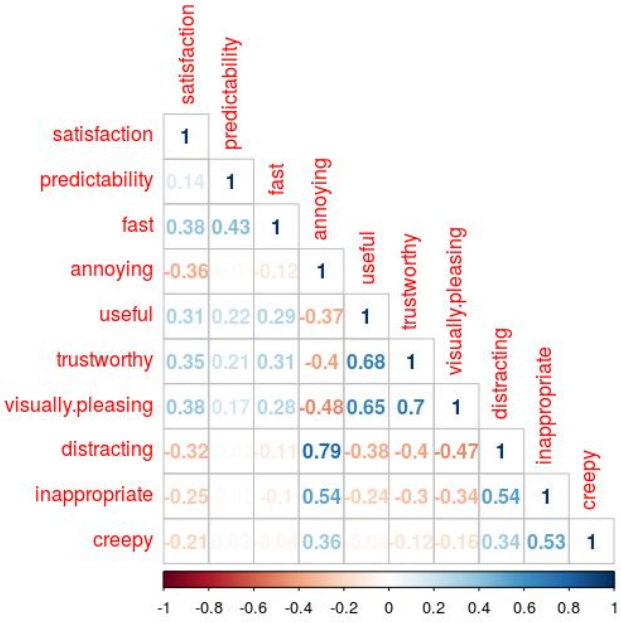


Figure 12. 10 UX metrics - pairwise correlation matrix plot

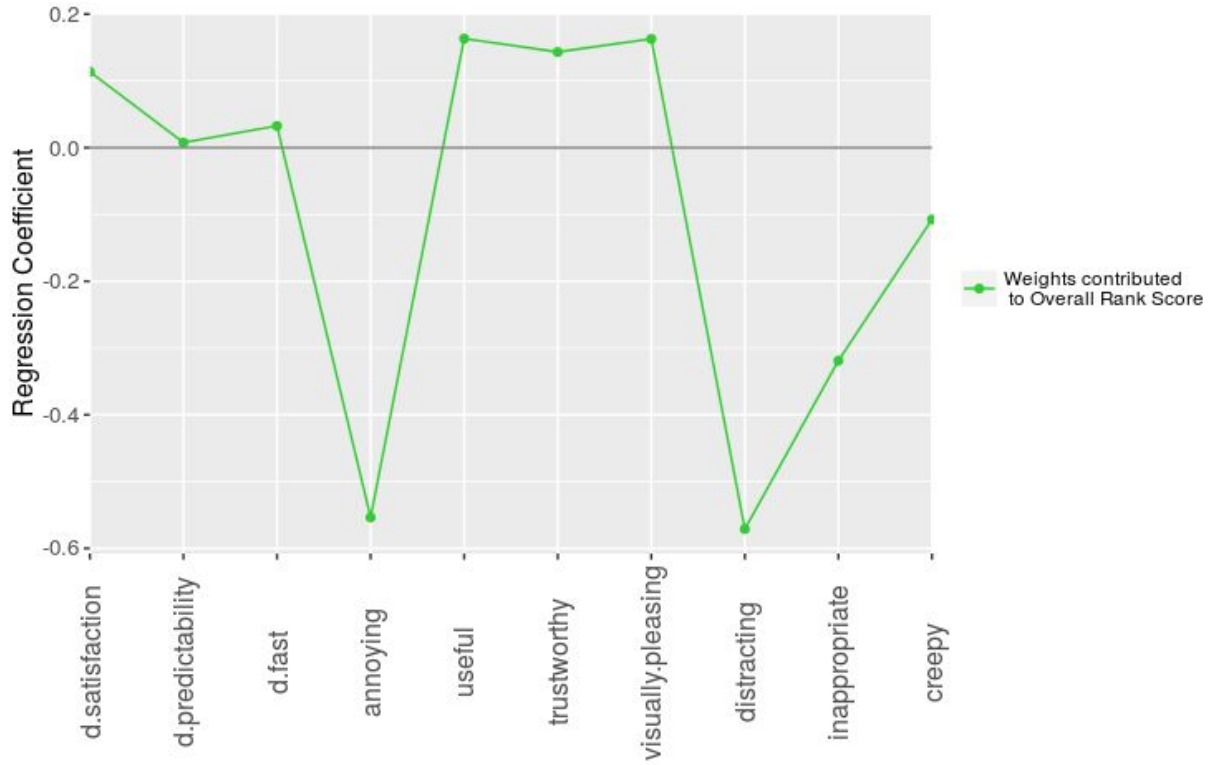


Figure 13. Weights of 10 UX metrics contributed to Overall Rank Score. (The weights are estimated regression coefficients from fitting PLS regression on Overall Rank Score over 10 UX metrics)

The figure above shows that Annoying and Distracting are two main negative contributors. Creepy and Difference in predictability are not contributing much, while all the others have moderate sensible contribution toward the overall rank score. This result confirms that the *Overall Rank Score* makes good connections with 10 UX metrics and is a good single metric to rely on. See the table below for the weights each UX metric contributed to the Overall Rank Score in PLS regression.

UX metrics	**PLS Regression Coefficient**
Difference in satisfaction	0.114
Difference in predictability	0.008
Difference in fast	0.033
annoying	-0.554
useful	0.163

trustworthy	0.143
visually pleasing	0.163
distracting	-0.571
inappropriate	-0.319
creepy	-0.107

Table 5: Weights (PLS Regression coefficients) of 10 UX metrics contributed to Overall Rank Score in the multi-ad mobile web experiment.

Demographic Distribution

We used Answers Research to recruit participants that are representative of the internet population. The four demographic variables we collected were: age, gender, income, and the status of employment. The demographic distributions looked good across two phases and aligned with Internet Population, except for gender being a bit skewed to Female in Phase II. See below.

Gender

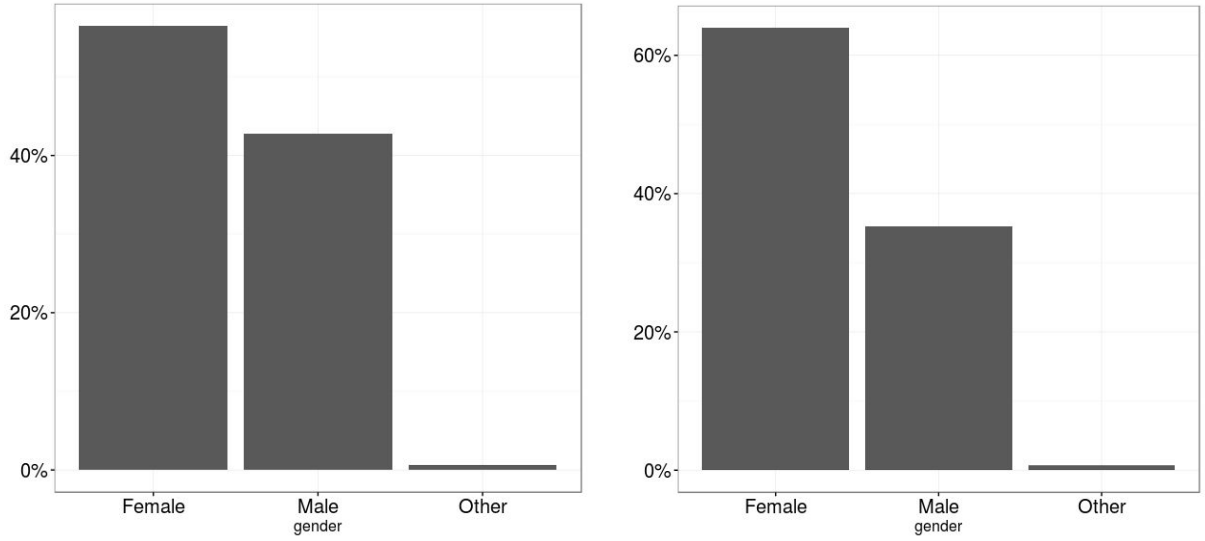


Figure 14. Gender distribution of Phase I (left) vs Phase I & II (right)

Age

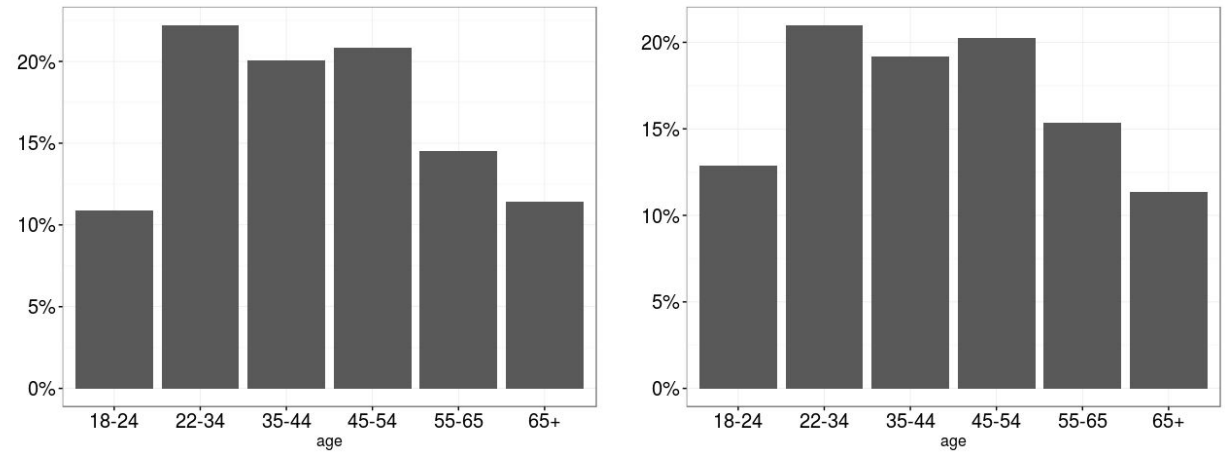


Figure 15. Age distribution of Phase I (left) vs Phase I & II (right)

Income

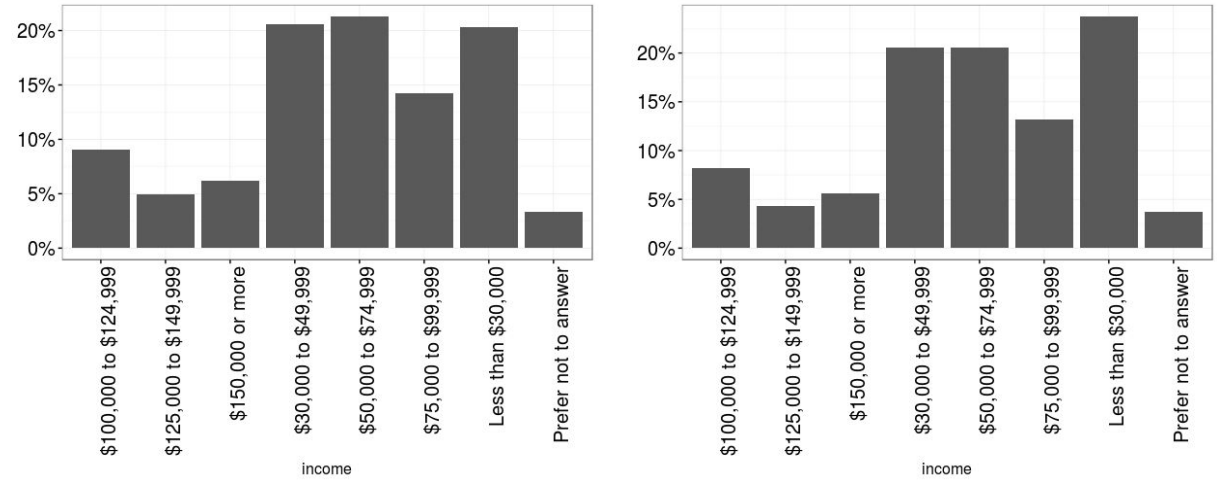


Figure 16. Income distribution of Phase I (left) vs Phase I & II (right)

Employment

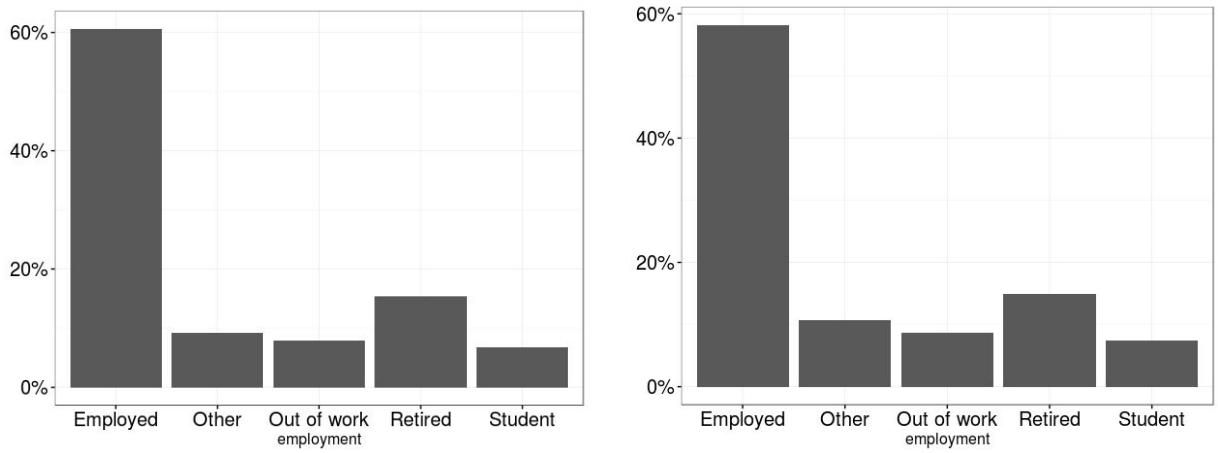


Figure 17. Employment distribution of Phase I (left) vs Phase I & II (right)

From the charts above, clearly Phase II data was somewhat skewed in gender. So we dug into the data to check whether gender has a significant impact on the result. First, we chose the most contributing UX metric - Annoying (to understand why, see Section PLS Regression coefficients, above) and it turned out that female and male participants do not rate differently in the Annoying score, nor in specific ad experiences. Then we got very similar coefficient estimates from fitting PLS regression by gender, as well as across phases. In essence, 10 UX metrics contribute to the overall score the same way across gender. Due to the nice property of the survey paradigm (i.e., asking for relative preference), the results based on relative rankings are less likely impacted even if there is any extreme rating.

Calculating Ad Density

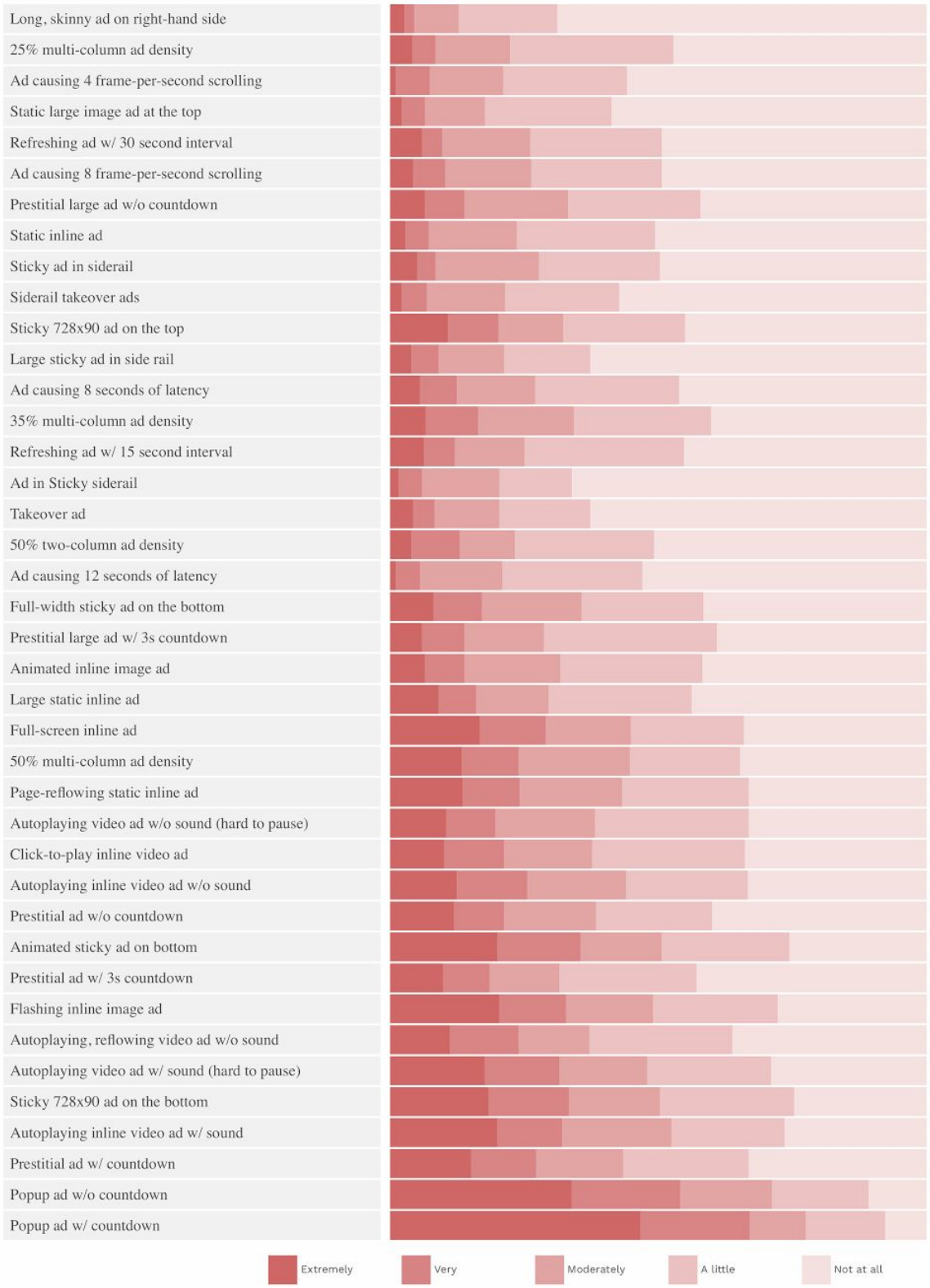
Determining the exact ratio of ad pixels to content pixels on a live webpage is complicated. Webpage rendering like font size and margin size depends on the device users visit the site with, and the calculation of ad pixel density depends on what counts as ad pixels and what counts as content pixels - for example whether you include margins and header bars as content pixels. For this study we did not include header bar or margin pixels, which appeared entirely to the side or below content as content pixels - meaning we included padding between paragraphs as content pixels, but not the padding between the side of a paragraph and the side of a device's screen. A 320x568 device like the iPhone 5 rendered 50%, 37% and 23% ad pixels for our 50%, 35% and 25% experiences respectively.

As such, for the experiences labeled "50% Ad Density," "35% Ad Density," and "25% Ad Density," please consider those figures as a careful estimate; the actual ad density that any user saw could have varied by a few percent based their device size.

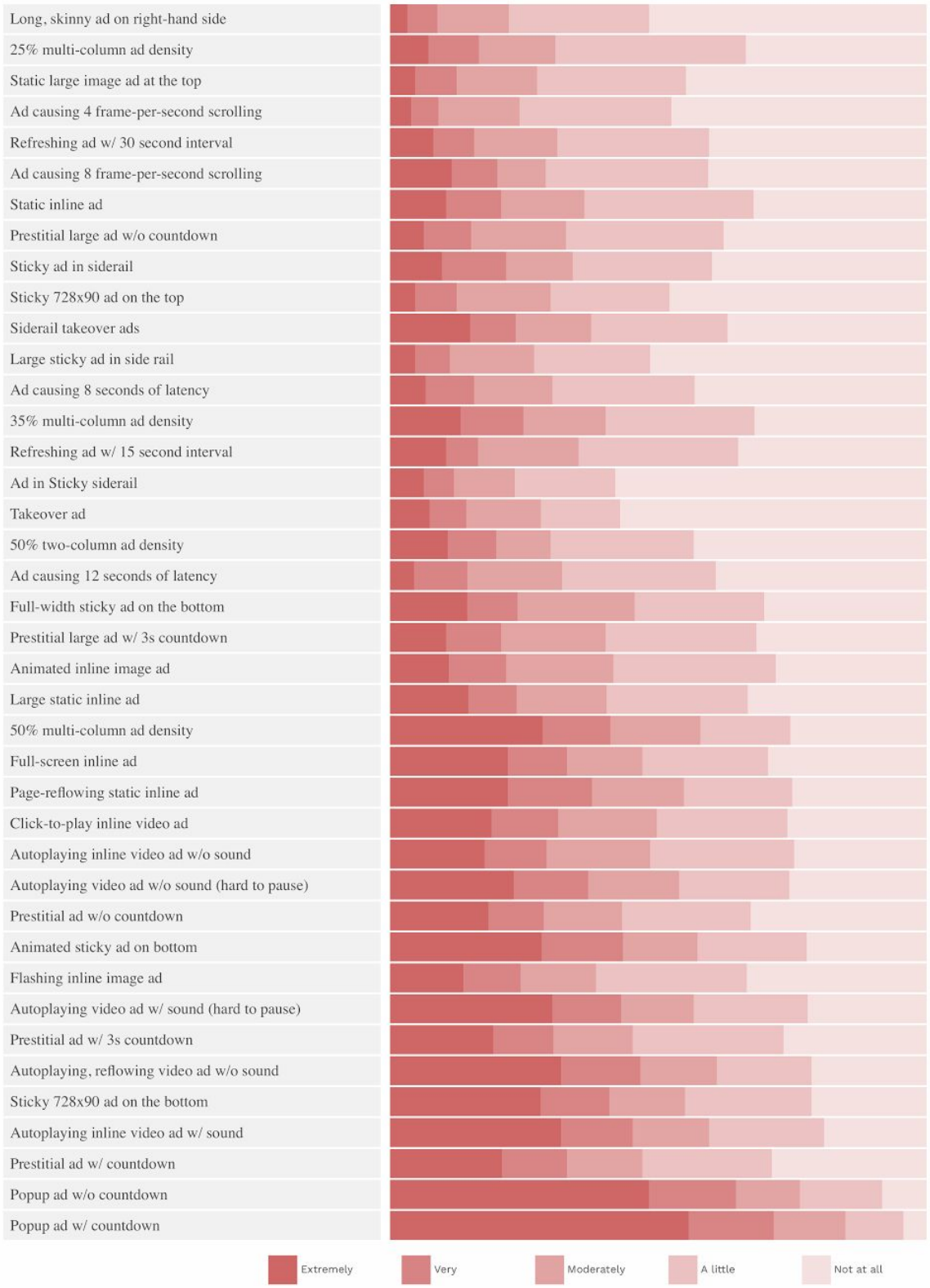
Full Results

Desktop Ad Experience Rankings

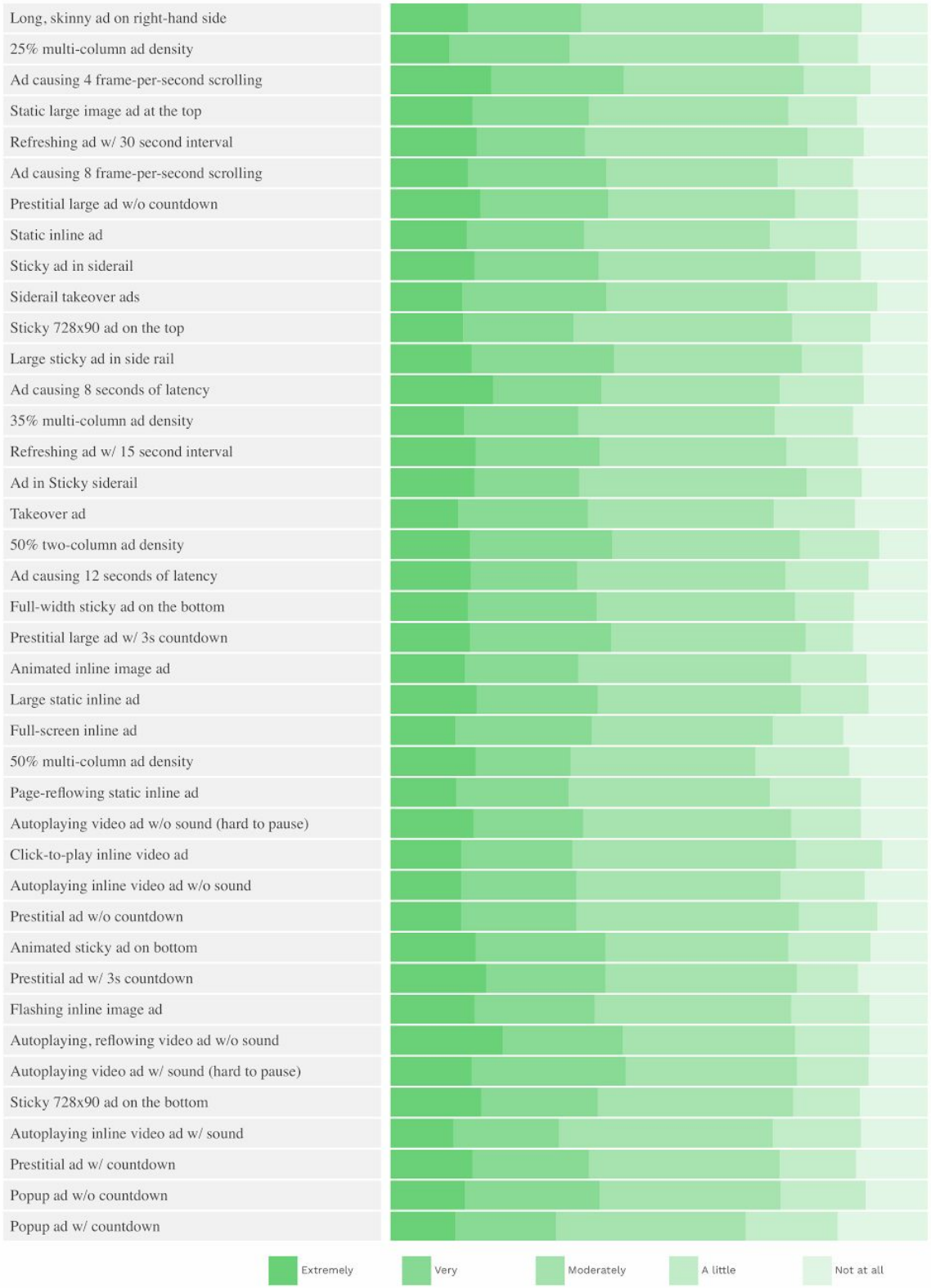
DESKTOP - Annoying



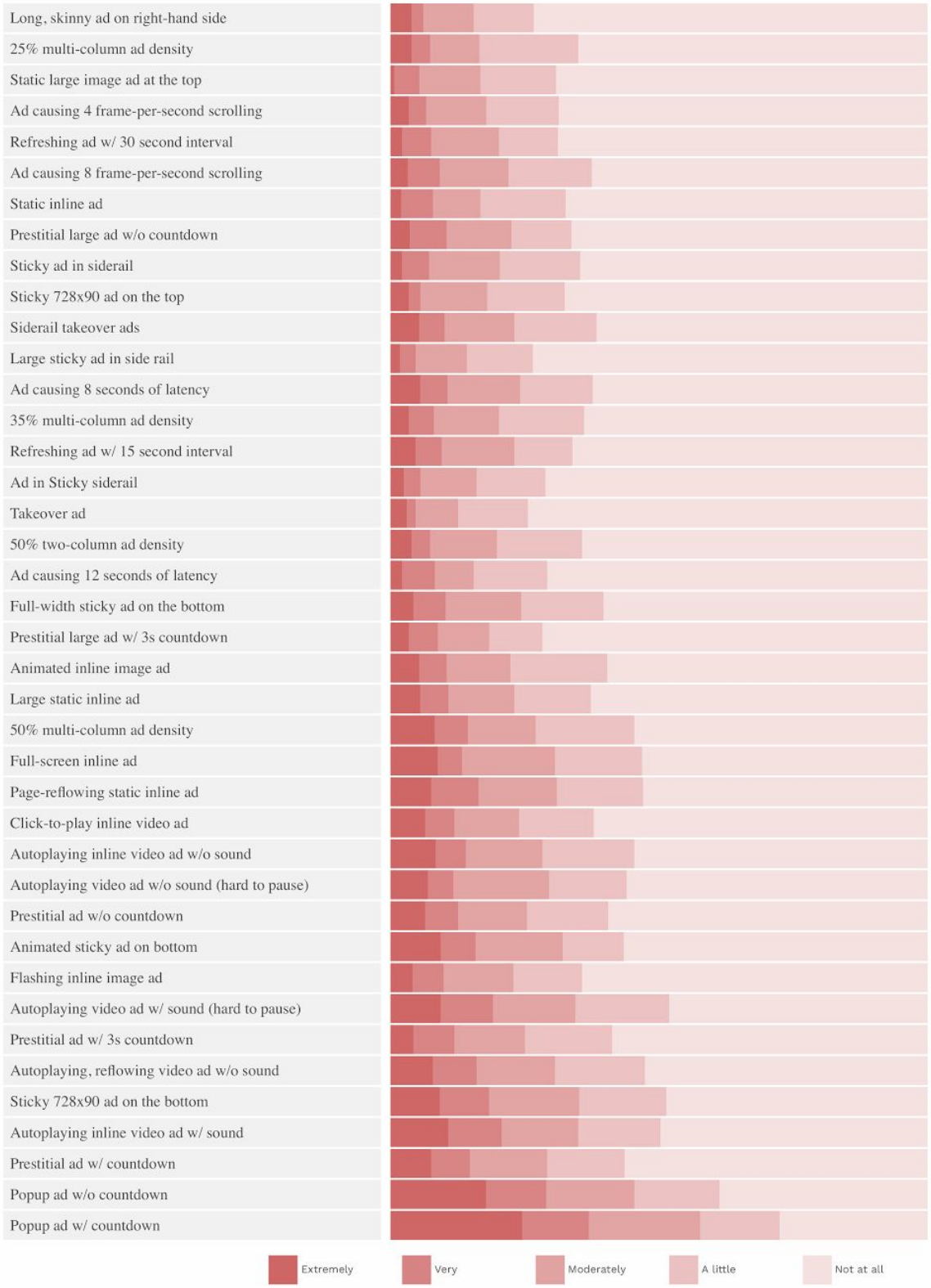
DESKTOP - Distracting



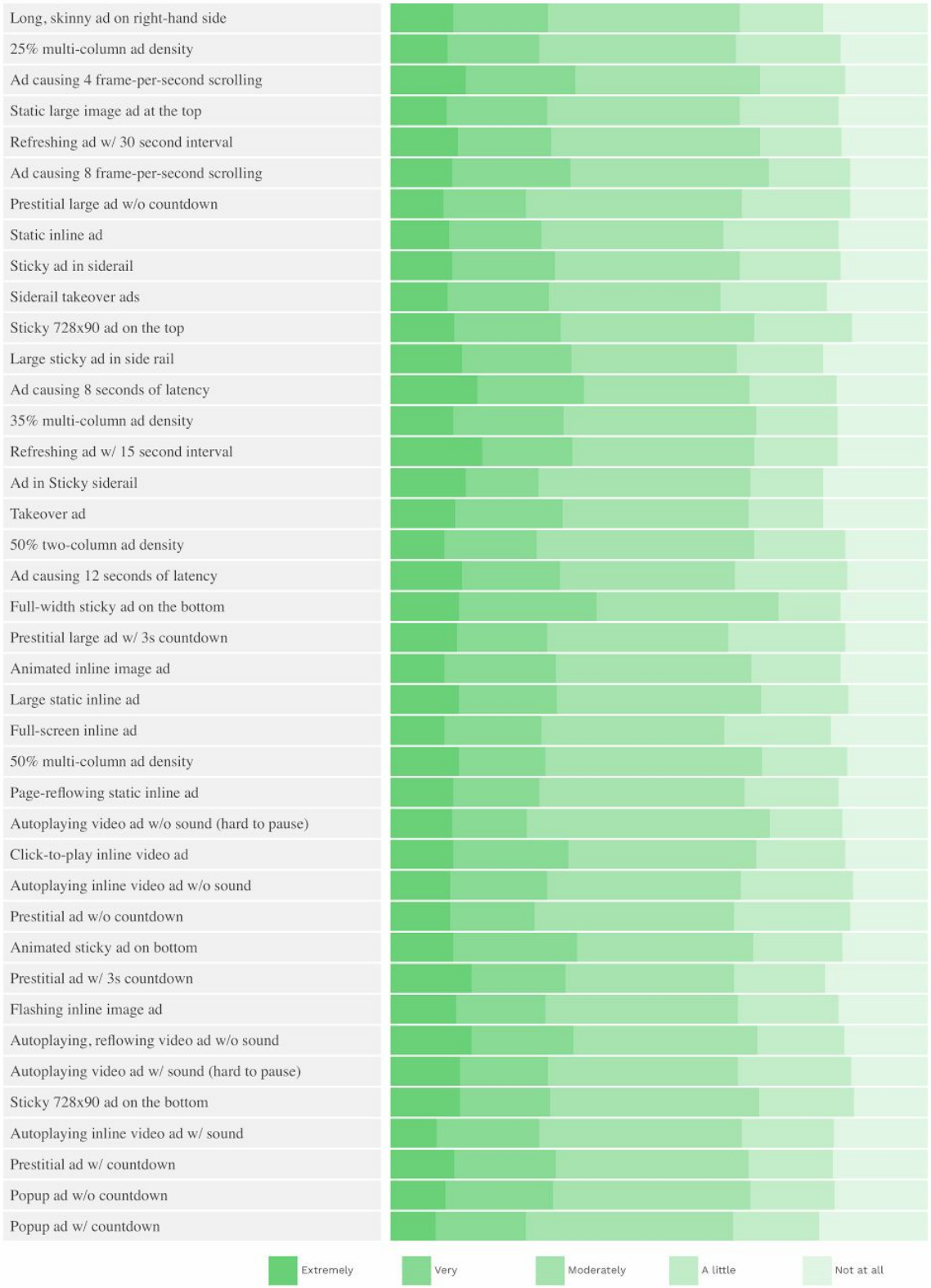
DESKTOP - Fast



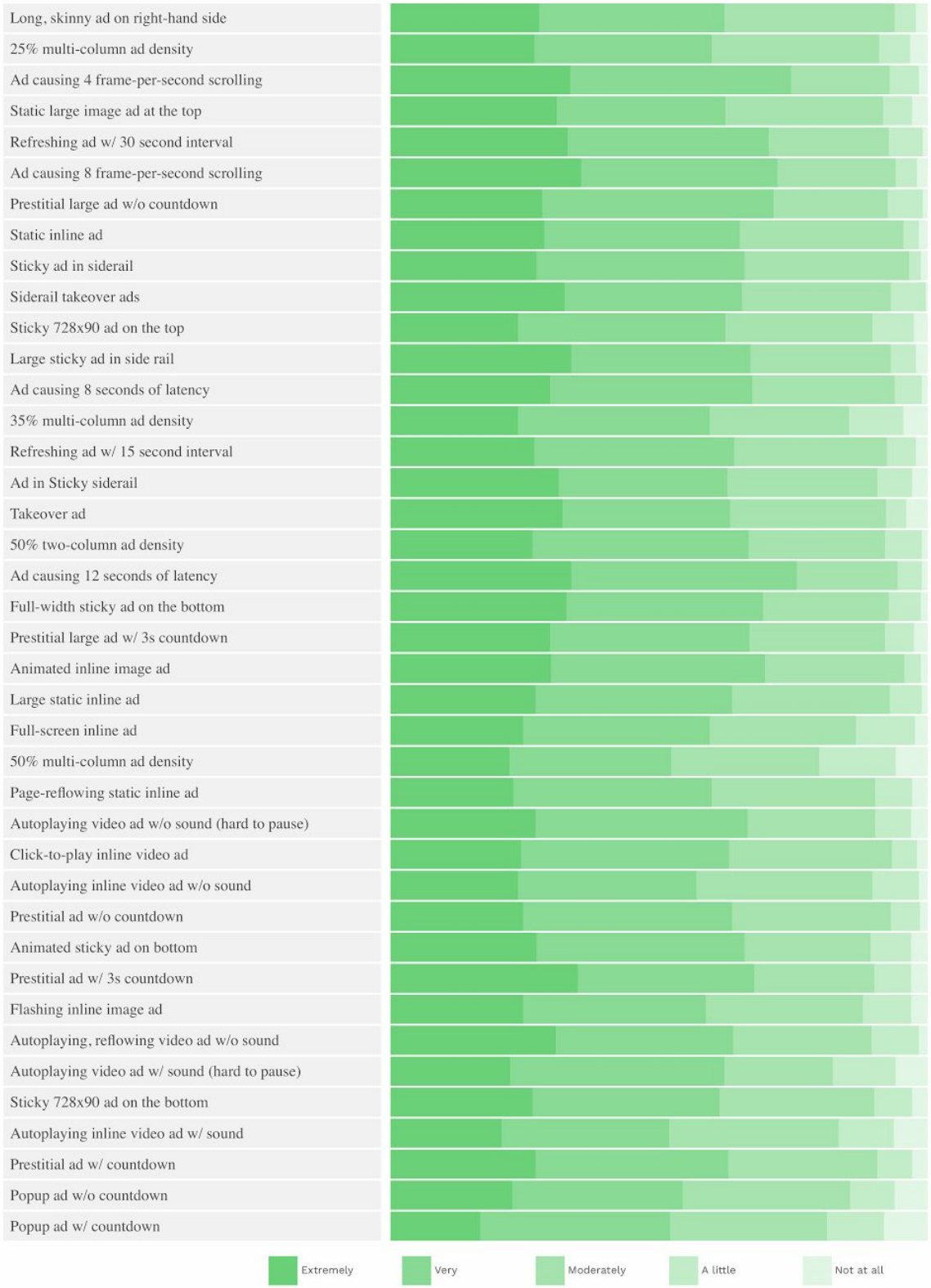
DESKTOP - Inappropriate



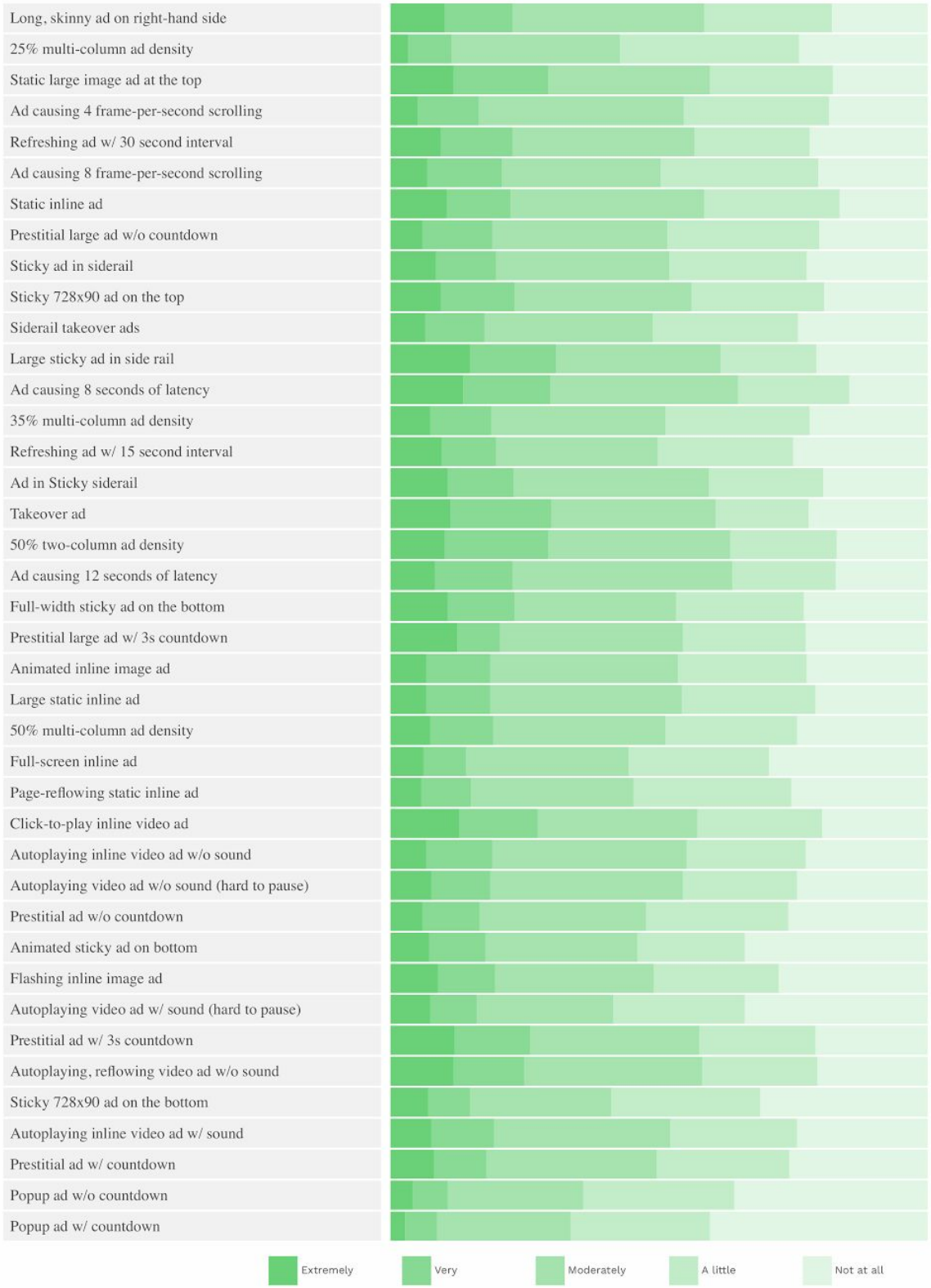
DESKTOP - Predictability



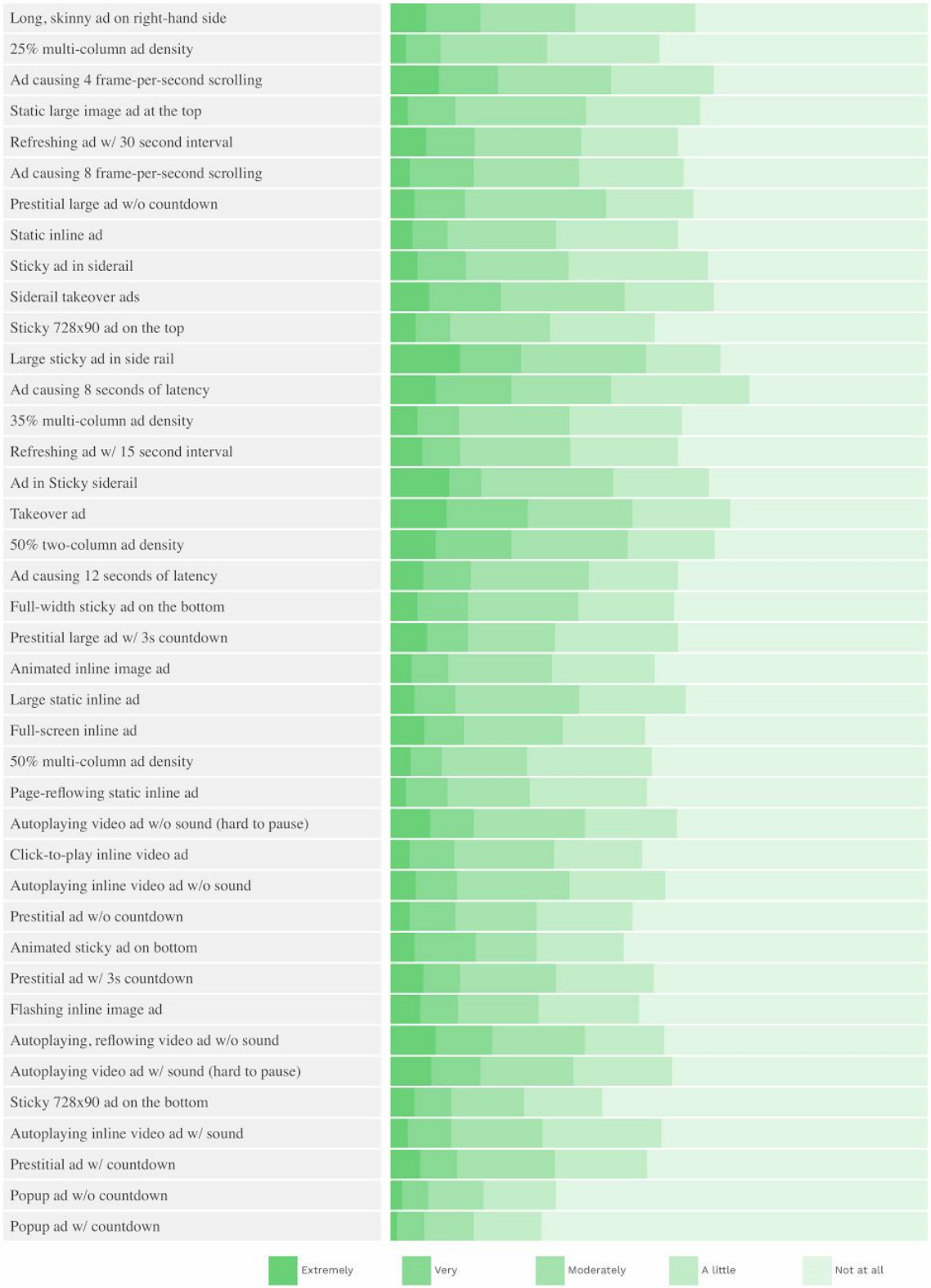
DESKTOP - Satisfaction



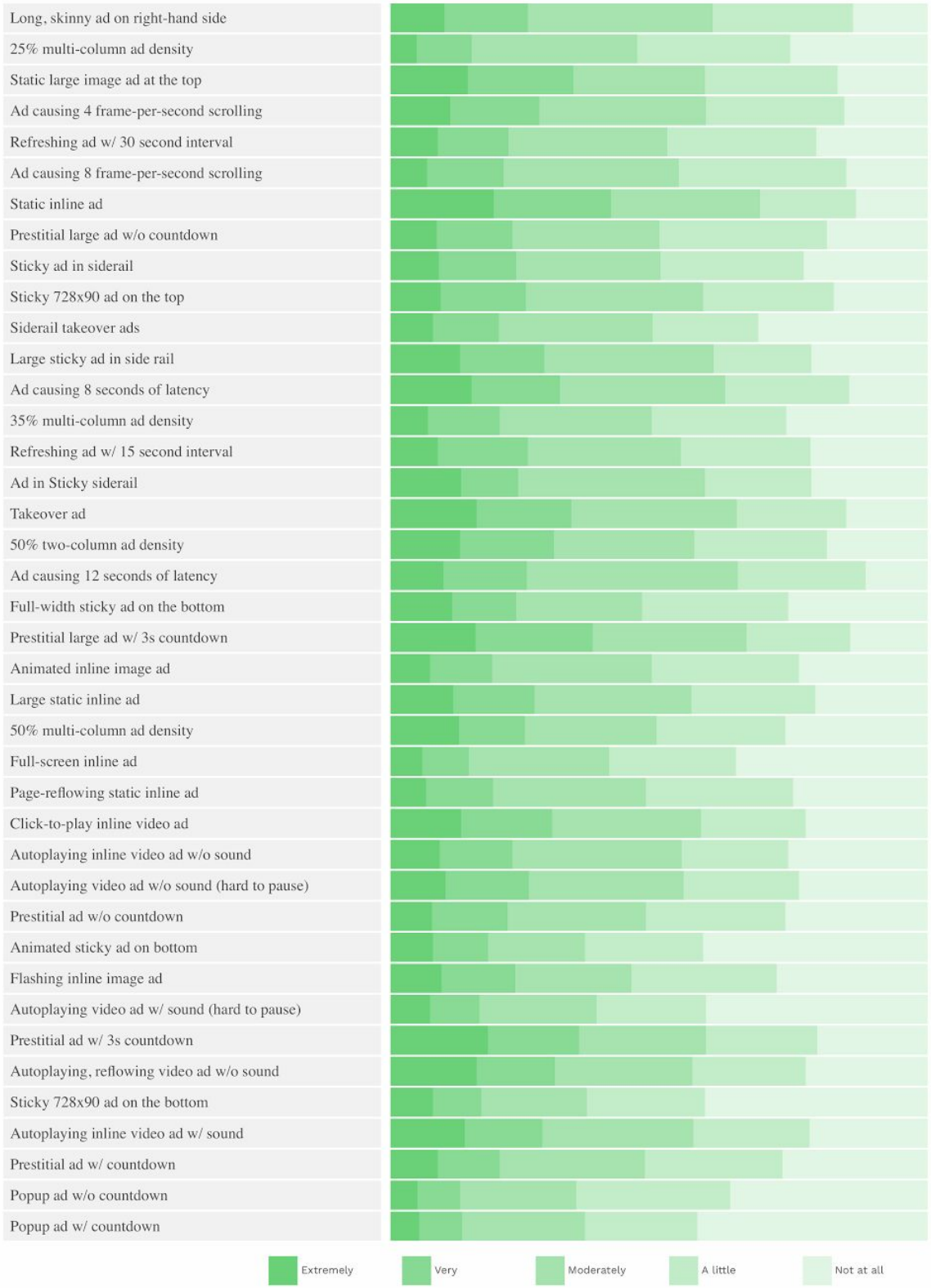
DESKTOP - Trustworthy



DESKTOP - Useful

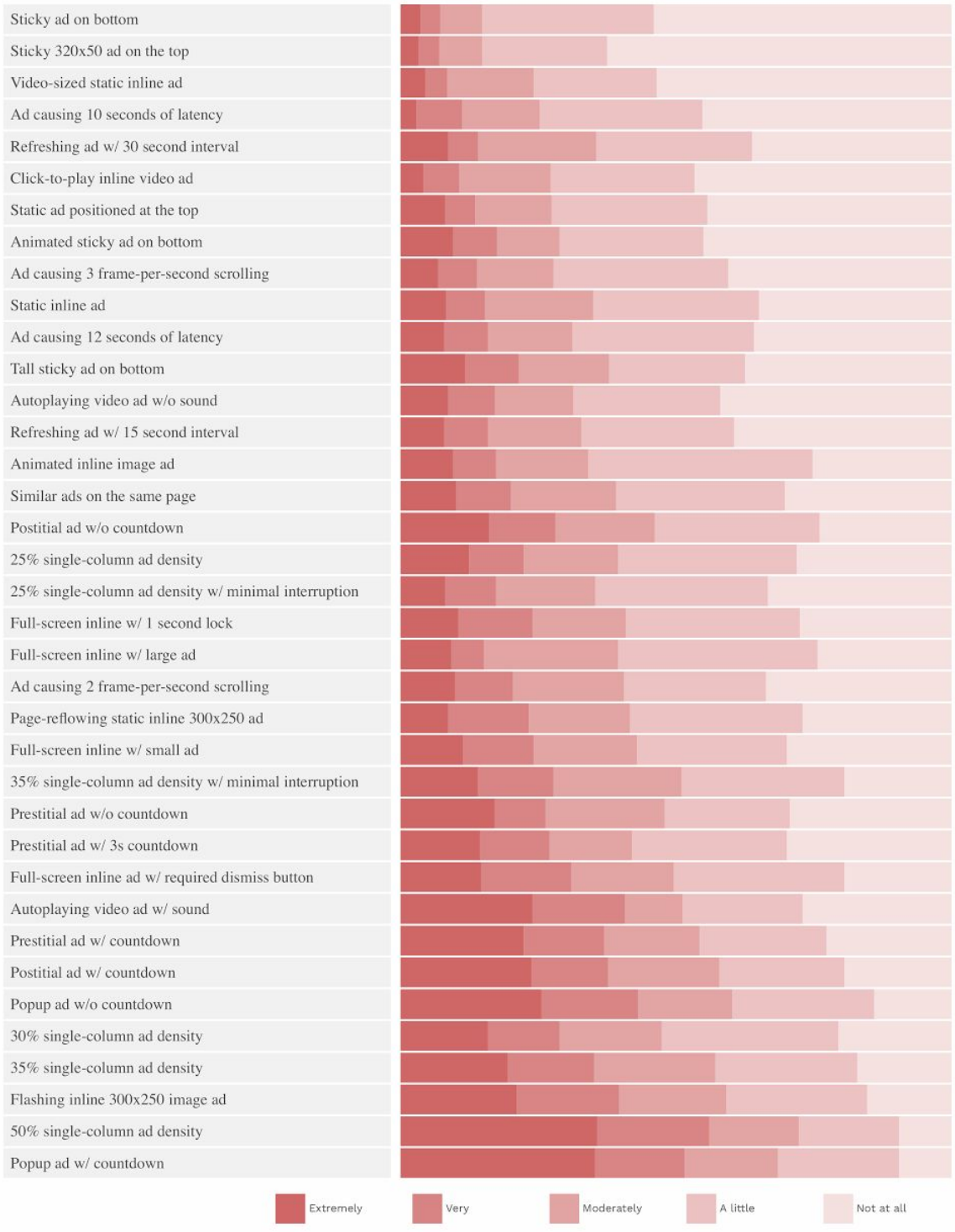


DESKTOP - Visually Pleasing

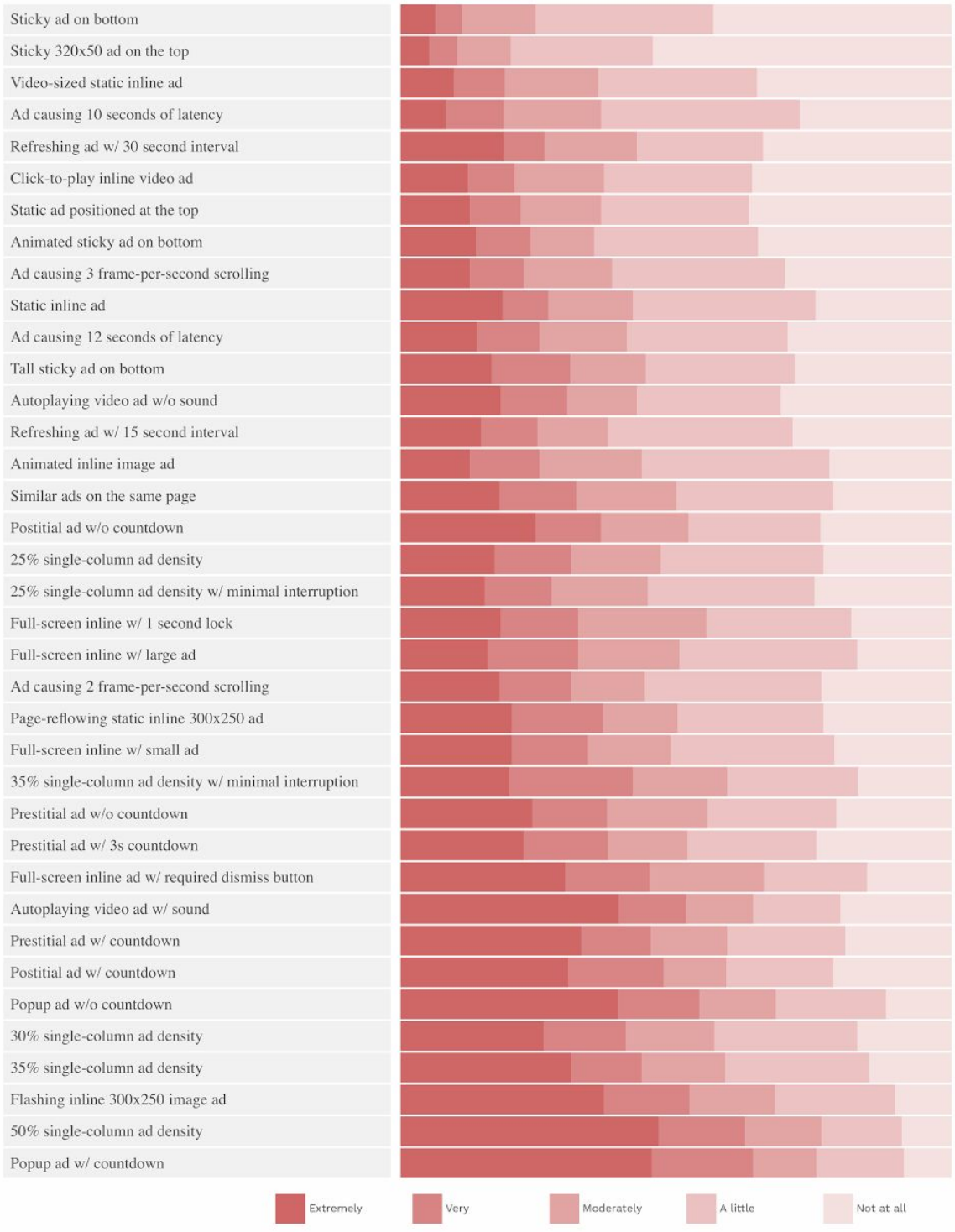


MOBILE - Creepy

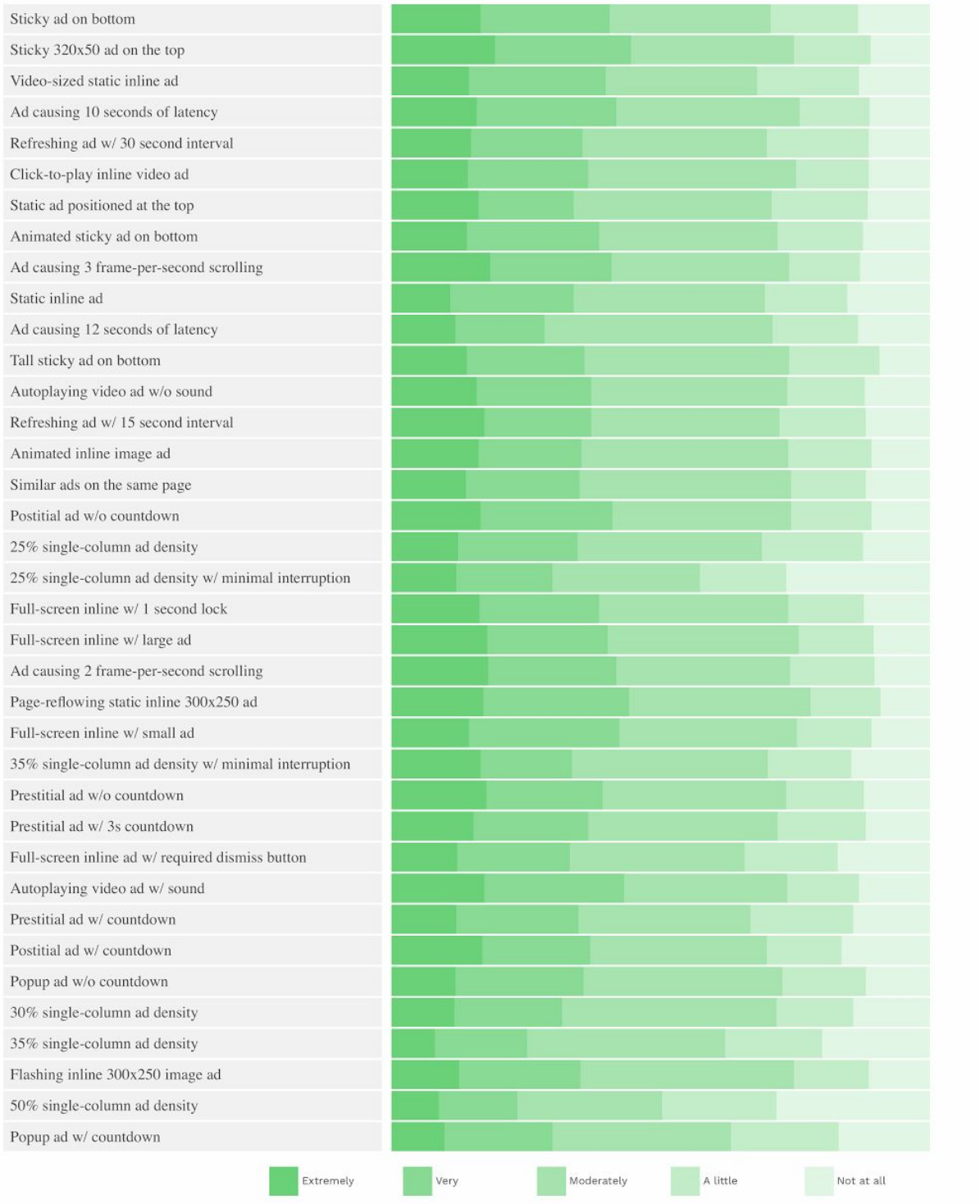
MOBILE - Annoying



MOBILE - Distracting

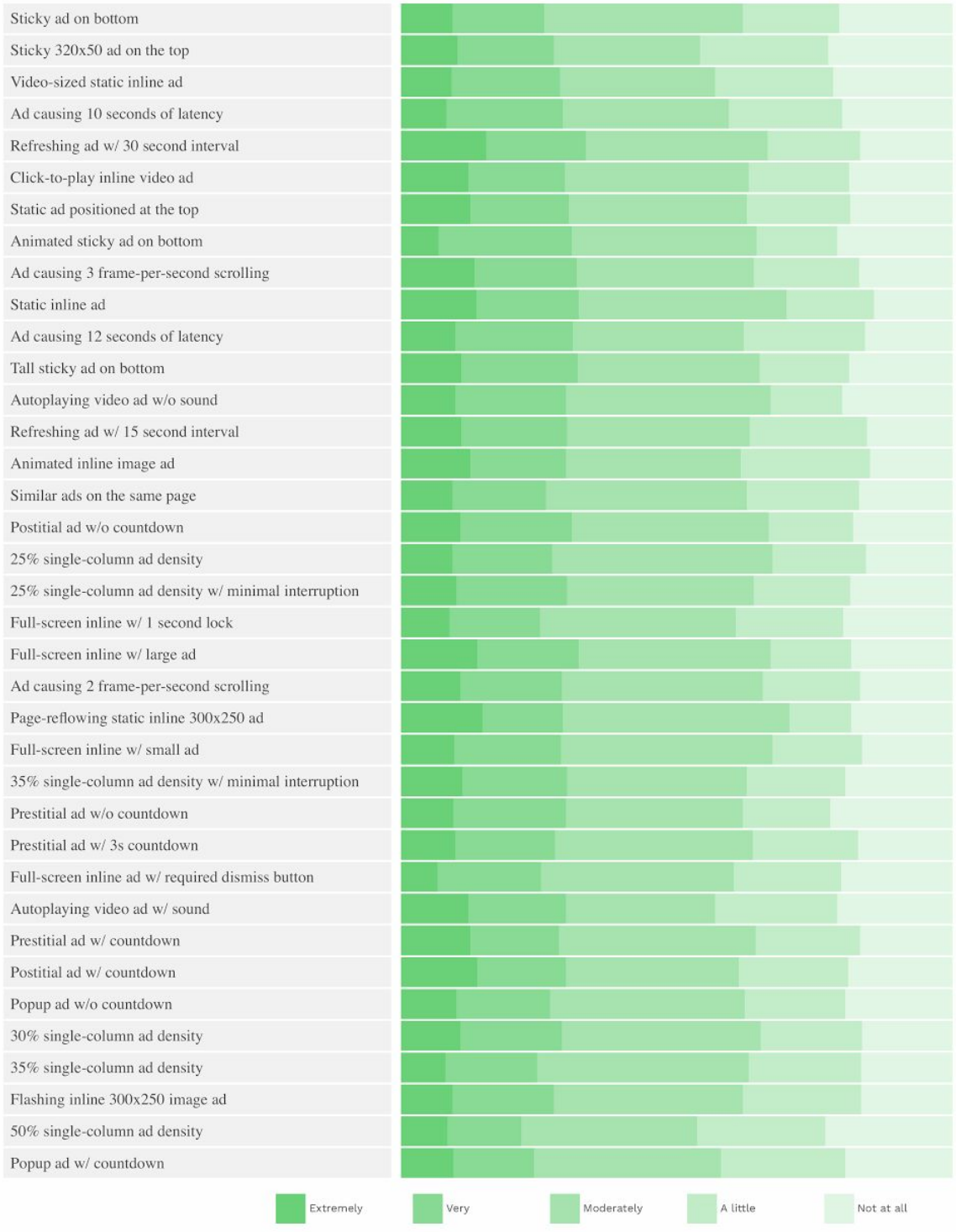


MOBILE - Fast

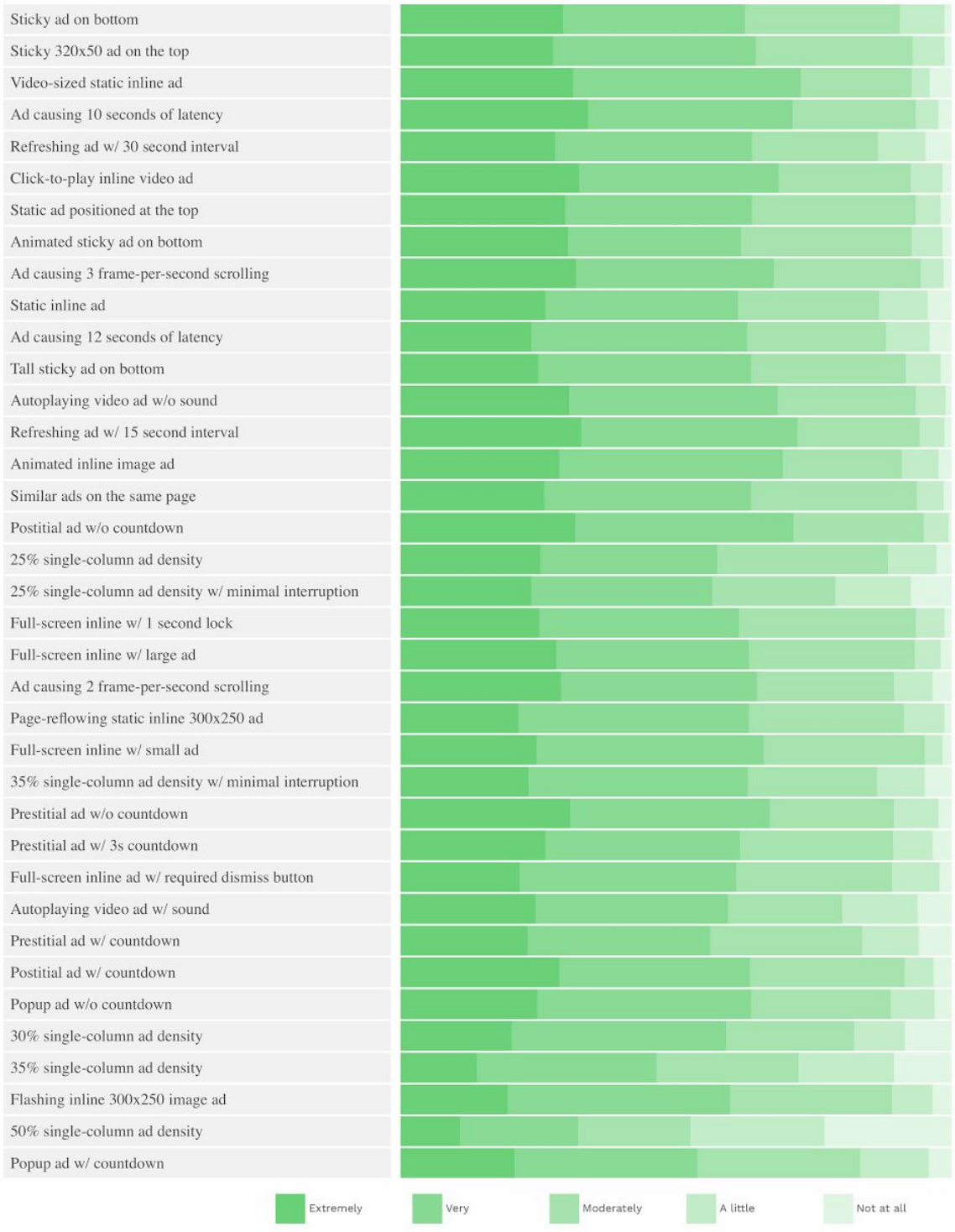

Sticky ad on bottom
Sticky 320x50 ad on the top
Video-sized static inline ad
Ad causing 10 seconds of latency
Refreshing ad w/ 30 second interval
Click-to-play inline video ad
Static ad positioned at the top
Animated sticky ad on bottom
Ad causing 3 frame-per-second scrolling
Static inline ad
Ad causing 12 seconds of latency
Tall sticky ad on bottom
Autoplaying video ad w/o sound
Refreshing ad w/ 15 second interval
Animated inline image ad
Similar ads on the same page
Postitial ad w/o countdown
25% single-column ad density
25% single-column ad density w/ minimal interruption
Full-screen inline w/ 1 second lock
Full-screen inline w/ large ad
Ad causing 2 frame-per-second scrolling
Page-reflowing static inline 300x250 ad
Full-screen inline w/ small ad
35% single-column ad density w/ minimal interruption
Prestitial ad w/o countdown
Prestitial ad w/ 3s countdown
Full-screen inline ad w/ required dismiss button
Autoplaying video ad w/ sound
Prestitial ad w/ countdown
Postitial ad w/ countdown
Popup ad w/o countdown
30% single-column ad density
35% single-column ad density
Flashing inline 300x250 image ad
50% single-column ad density
Popup ad w/ countdown
Extremely
Very
Moderately
A little
Not at all

MOBILE - Inappropriate

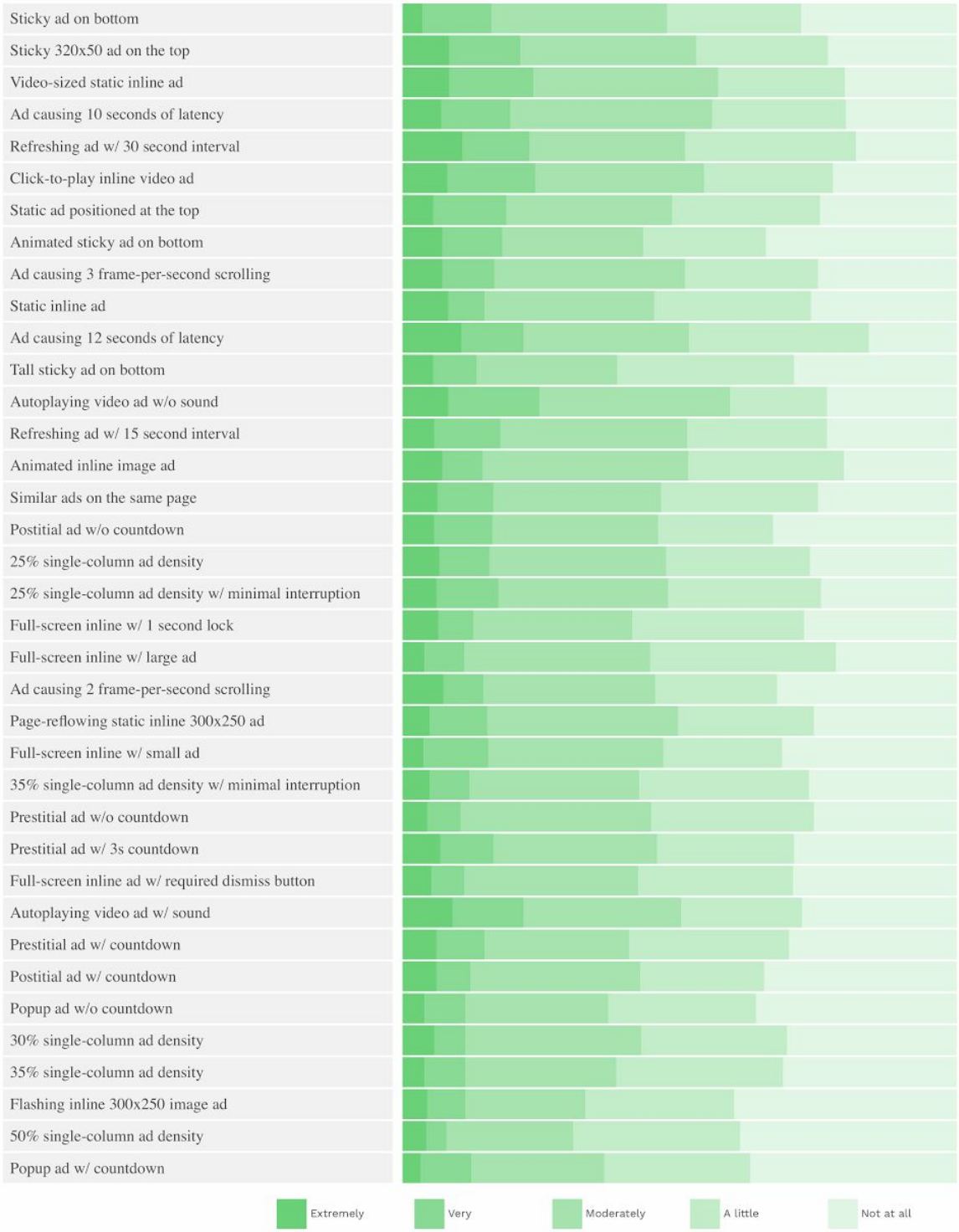
MOBILE - Predictability



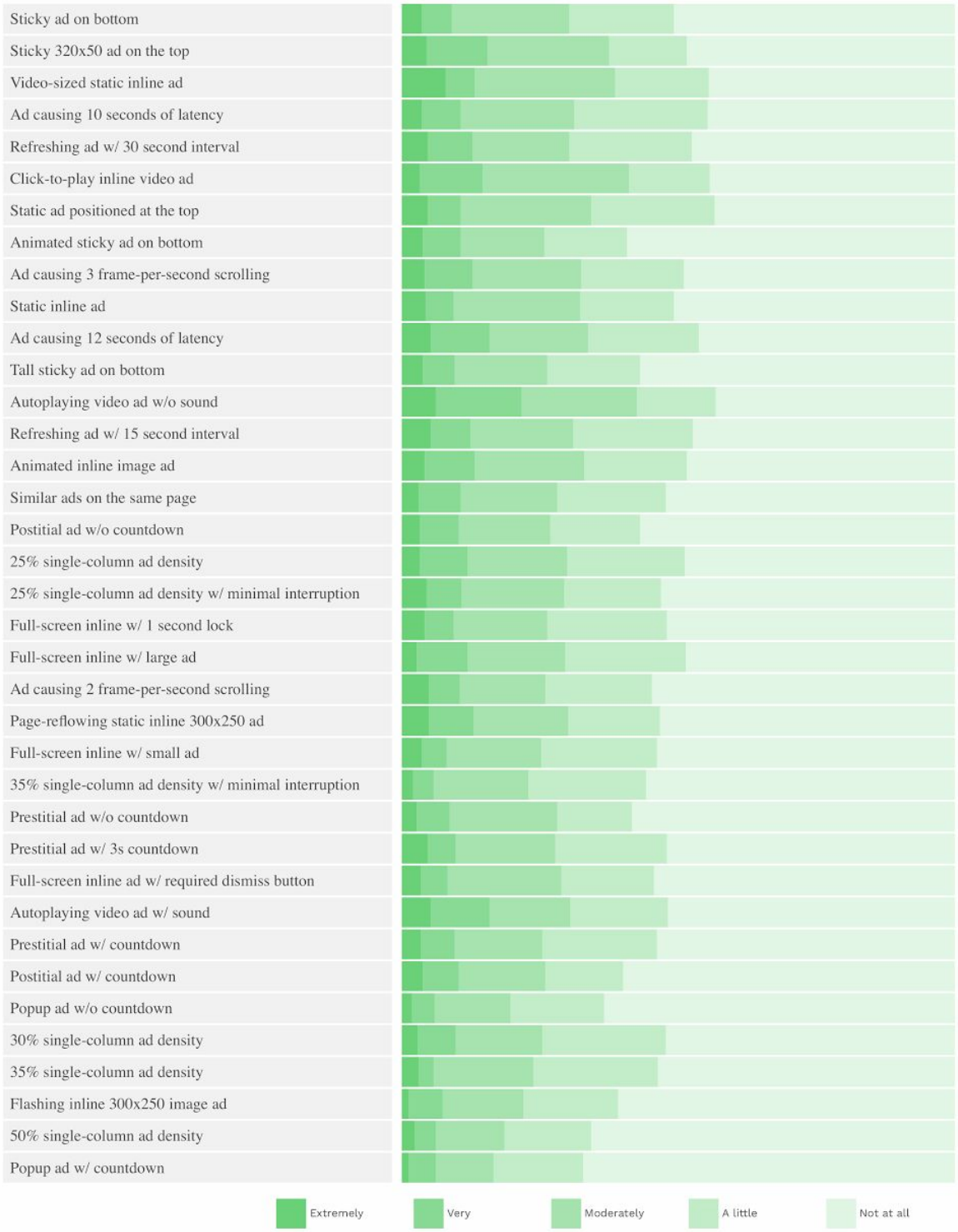
MOBILE - Satisfaction



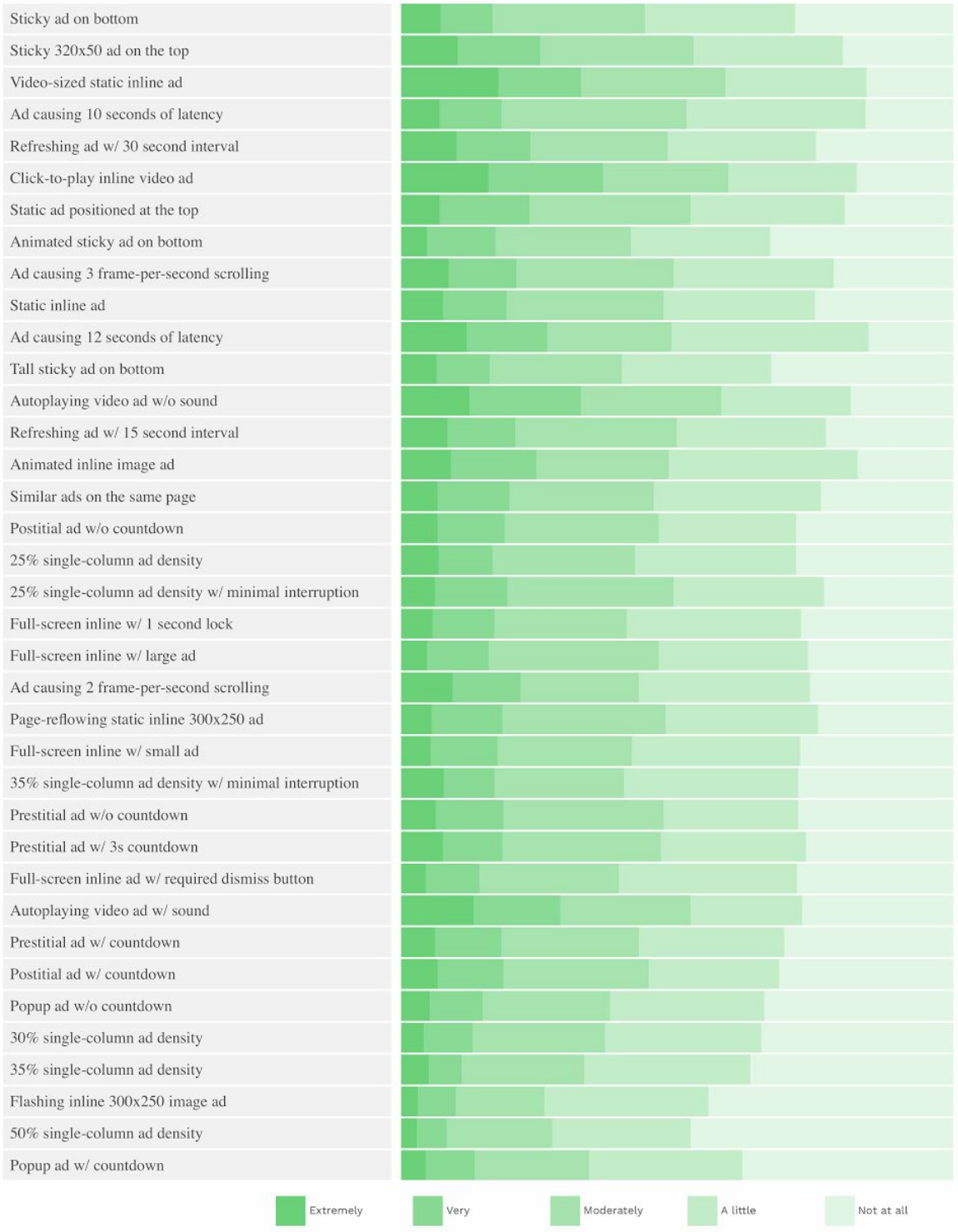
MOBILE - Trustworthy



MOBILE - Useful



MOBILE - Visually Pleasing



Results Table