An experimental methodology to measure consumer perceptions of ads in short-form video

Table of Contents

Abstract

Good ad experiences fund the open web: almost everything we enjoy every day about the web is wholly or partly funded by advertising. These experiences enable marketers to grow their businesses, and consumers to learn about new products, services, and deals. Bad ad experiences disrupt this balance by driving consumers to install ad blockers, leave the publisher's site, or take other actions that harm the open web's long-term viability.

Previous work, the Multi-Ad Study (see *An Experimental Methodology to Rank N Ad Experiences by Consumers' Perceptions*, Ad Experience Research Group, 2016), showed users clearly prefer some ad experiences over others in an article-reading environment, and that it is possible to measure their preferences. This paper describes the methodology and results from a new study to evaluate ad perceptions in a different online context: short-form[1] 2-2.5 minute videos in which watching the video content is the primary consumer experience (throughout the rest of this paper, we will refer to ads that appear before, during, or after the main video as *instream video ads).*

The new study used a similar experiential approach as the Multi-Ad Study, but was adapted so it could evaluate perceptions of 40 instream video ad experiences across mobile and desktop in short-form 2-2.5 minute publisher videos (work is underway to evaluate perceptions in short-form content videos around 8 minutes each). In this study, each participant watched 3 videos, each containing a different ad experience. Then, they rated their overall webpage experience and each ad on various dimensions. Finally, they ranked the three ads based on how much each ad interfered with their ability to watch the video.

Because a large number of ad experiences were studied, the same iterative sampling[2] approach was used as in the Multi-Ad Study to reduce the number of participants needed to evaluate all the ads. Then, the Bradley-Terry algorithm (Turner and Firth, 2012) was used to create a unified stack rank of ad experiences by perceived interference for each platform.

Our findings reveal clear differences among the ad experiences. The most interfering ads were generally those that make users wait before they can skip ads, are longer, or interrupt users while they watch a video. The least interfering ads tended to have the opposite characteristics.

Overall, this research indicates video publishers can reasonably use many ad experiences that are acceptable to consumers but should avoid ones that disrupt users' goals. It provides the foundation for international studies to establish global standards.

[1] "Video Advertising Glossary." iab Digital Video. Retrieved from http://dvglossary.www2.iab.com/#chapter-24

[2] "An Experimental Methodology to Rank N Ad Experiences by Consumers' Perceptions." *Coalition for Better Ads*. Retrieved from https://www.betterads.org/research/rankingpaper/

Introduction

Video publishers use ads to generate revenue and help fund the open web, which allows consumers to access valuable content. These ads also provide viewers with a valuable way to learn about businesses and products they might enjoy. However, viewers typically do not appreciate ads. In the context of videos, ads can interfere with viewers' ability to watch the video in several ways, such as by interrupting content, distracting from content, or generally annoying the viewer. Understanding viewers' perceptions of ad experiences can help publishers improve the user experience and ensure the ads they show do not drive users to take actions that harm the open web's long-term sustainability.

In this paper, a set of ad experiences are ranked based on viewers' perceptions of the degree to which the ads interfere with their ability to watch the main video, and a methodology is validated that will be used to create a global video standard. Given the multitude of video types, the focus is on evaluating ads in the context of one common online video genre: short-form general interest content between 2-2.5 minutes long that appears in a video player on a web page on a mobile or desktop device. This methodology is based on previous work, the Multi-Ad Study (see *An Experimental Methodology to Rank N Ad Experiences by Consumers' Perceptions*, Ad Experience Research Group, 2016), that ranks display ad experiences common on mobile and desktop web environments (e.g., popups, videos that automatically play with sound, etc.) in an article-reading context. This methodology was adapted to a video-watching context and employed to study users' perceptions of ad experiences on both mobile and desktop video content.

The rest of this paper is organized as follows: First, there is a summary of the methodology developed in the work mentioned above to rank display ad experiences, then a description our adaptations to this methodology to test ad perceptions in short-form video content. Finally, results are presented from testing 40 ad experiences with 5,616 participants.

Background

An Experimental Methodology to Rank N Ad Experiences by Consumers' Perceptions described the methodology used to evaluate users' perceptions of ad experiences in which reading an article online was the primary consumer experience. An ad experience refers to an experimental condition that combines multiple factors, including one or more ads and their content (i.e., ad creative) shown to participants in the context of a publisher site.

The methodology was based on two key principles: *adopting the user's perspective*, and measuring ads in the *context of common user activities online*.

The experiments used a within-subjects design in which participants read four articles (three articles were experimental conditions, each of which contained an ad experience; the control article did not), then answered 10 questions about their experience. The questions were split into two categories: overall webpage experience measures (satisfaction, predictability, page load speed) and ad-specific dimensions (annoyance, usefulness, trust, visual appearance,

distraction, inappropriateness, creepiness), if the article contained an ad. At the end of the study, participants stack-ranked the three ads by preference.

To minimize the number of participants, a staged process optimized which ad experiences to show to each participant, since randomly selecting three ads from the overall set would have required a larger sample size. After collecting preference rankings from many ad experience sets, the Bradley-Terry (B-T) Algorithm (Turner and Firth, 2012) was used to combine them into a unified ad preference ranking [see below charts, which depict the preference rankings on the mobile-web and desktop]. Using this approach, around 80 ad experiences were tested across the mobile-web and desktop.

Mobile Ad Experience Rankings

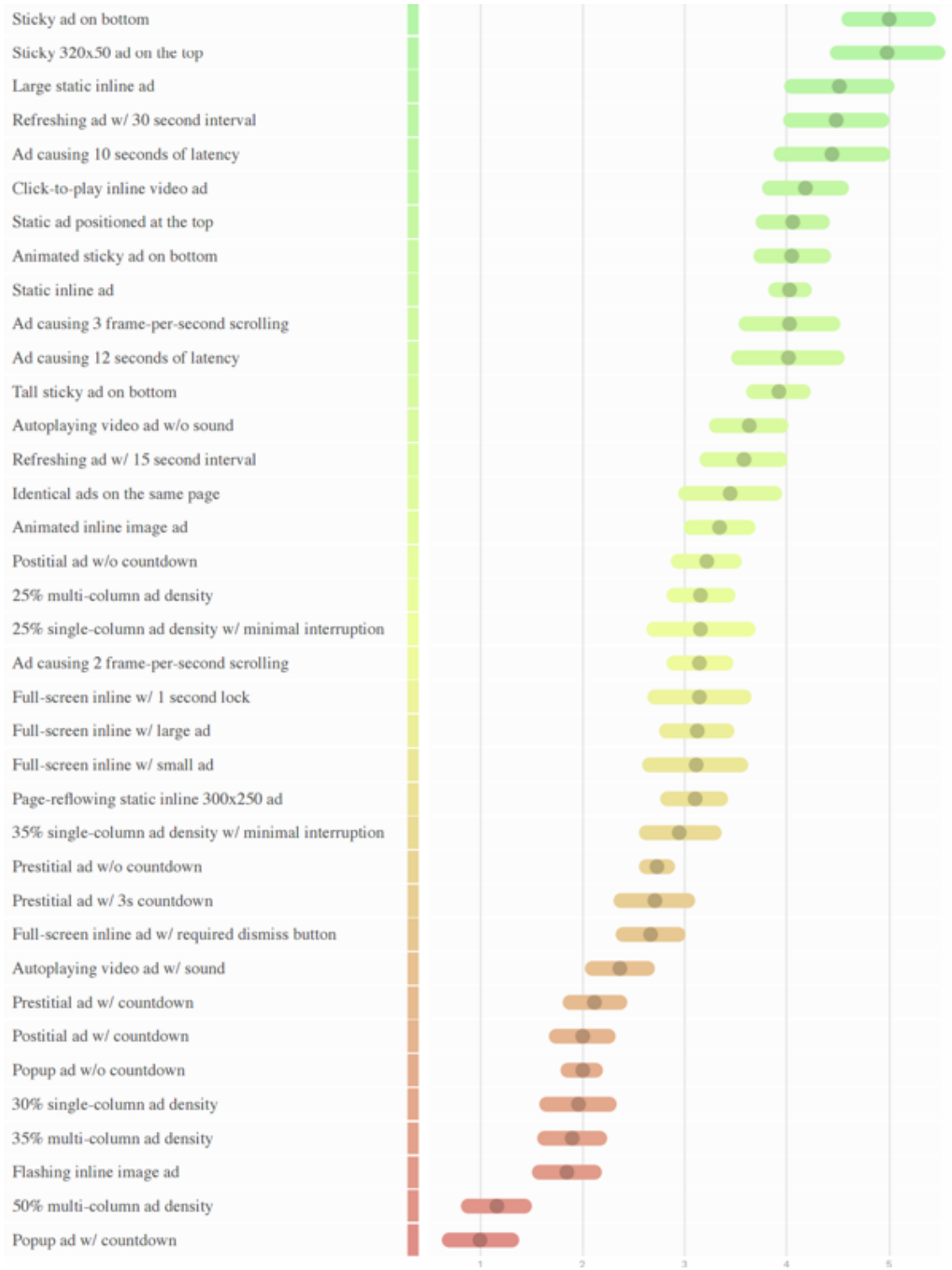


Figure 1. Overall Rank Score of the 37 mobile ad experiences, where 5 is the most favored experience and 1 is the least favored

Desktop Ad Experience Rankings

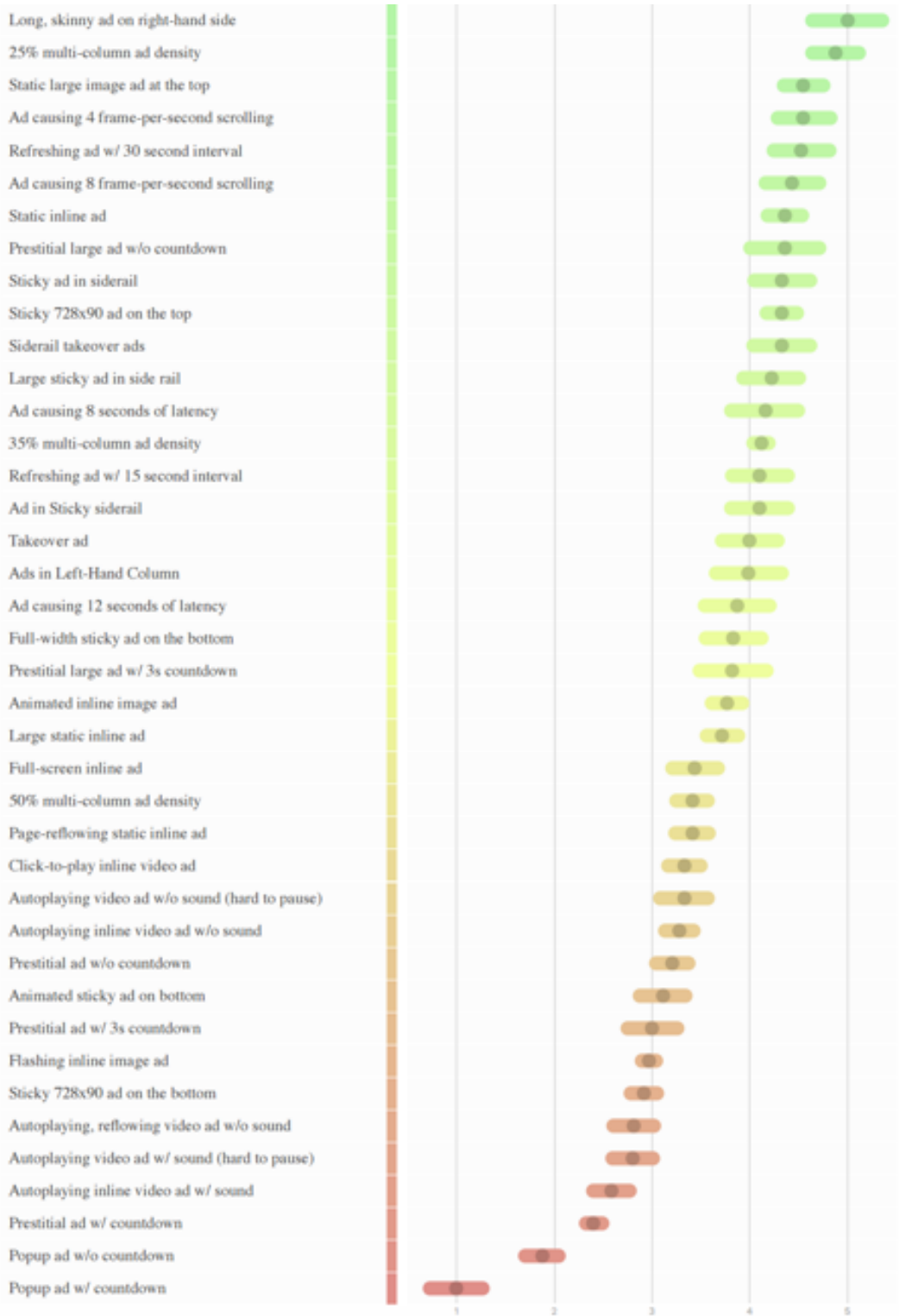


Figure 2. Overall Rank Score of the 40 desktop ad experiences, where 5 is the most favored experience and 1 is the least favored

The Multi-Ad approach showed it is possible to efficiently rank a large number of ad experiences by preference on different platforms. Moreover, it can be adapted to study other user contexts, such as watching videos online.

Experimental Overview

This paper describes two new experiments (mobile and desktop) used to evaluate ad experiences in an instream video environment, such as different length non-skippable pre-rolls, which play before the main video, or banners that are overlaid on the main video while it plays. The experiments used a similar within-subject methodology as the previous Multi-Ad work, with some adaptations that allow the evaluation of ad perceptions in an instream video environment.

Experimental Methodology

Identical mobile and desktop instream video experiments were run, using the same study flow and survey questions, although the screen layout was customized to fit each device environment. Experiments were separated by environment, so each participant only saw mobile or desktop ad experiences. Altogether, 40 ad experiences were evaluated (described in the next section) across mobile and desktop environments.

This section details the video methodology and the changes that were made in order to evaluate instream video ad experiences.

Overview: Experimental Design

Similar to our Multi-Ad experiments, a within-subjects design was used in which each participant watched 3 videos, each of which contained an ad, and then answered questions about their experience. To shorten the overall study length, the control (ad-less) experience was removed. After watching each video, participants answered questions about their general webpage experience and the ads. Finally, they ranked the 3 ads based on the degree to which each ad interfered with their ability to watch the video.

Main Adaptations to Study Instream Video Ad Experiences

While the overall study design closely mirrored that of the Multi-Ad experiments, the survey questions were tailored to evaluate ad attributes that are unique to instream video ads and parts of the protocol were changed to address issues observed during pilot studies.

Modifying the Survey Framework

In 2018, a series of qualitative and quantitative studies were run with US participants to identify which factors influence video ad perceptions. In the qualitative phase, 1,012 participants answered open-ended questions about video ad experiences, which revealed key themes in what influences perceptions. 12 adjectives were then selected from the set and 1,052 users were asked to rate the extent to which each adjective described online video ads. Finally, a factor analysis was run to determine which adjectives had similar ratings, which in turn showed

three underlying dimensions influence video ads perceptions: how *interruptive* the ad was, how much users *enjoyed* watching the ad, and how *relevant/worthwhile to watch* the ad was to them.

Creating the survey started with the three dimensions from the factor analysis described above (i.e., *distraction [interruption], enjoyment,* and *personal relevance*). Then, based on themes from additional qualitative video ads studies conducted in the US, five more ad-specific dimensions were added:

- Annoyance
- Whether the ad felt too personal
- The timing of the ad's appearance during the main video
- How long the ad felt
- How long participants felt they had to wait until they could skip the ad

Finally, two questions were added to evaluate how the ad impacted participants' broader webpage experience:

- Overall webpage satisfaction
- How predictable the webpage behavior felt

The final survey contained two general webpage experience questions and eight ad-specific questions (shown below), which were validated through cognitive pretests and piloted for effectiveness. All questions used fully labeled, vertical 5-point scales. A few questions were hidden that did not apply to certain ad experiences (for example, "how enjoyable to watch..." for overlays).

- How satisfied were you with the OVERALL EXPERIENCE watching the video on this web page? (from *Very satisfied* to *Very dissatisfied*)
- To what extent does the following statement describe the webpage?
    - "The web page did not surprise me with unexpected behaviors?" (from *Not at all* to *A great deal*)
- How ANNOYING was the ad? (from *Not at all* to *Extremely*)
- How ENJOYABLE was the ad to watch? (from *Extremely* to *Not at all*)
- How RELEVANT was the ad to you? (from *Extremely* to *Not at all*)
- To what extent does each of the following statements describe THE AD?
    - "The ad was distracting." (from *Not at all* to *A great deal)*
    - "The ad appeared at an unexpected point during the main video." (from *Not at all* to *A great deal)*
    - "The ad was too long." (from *Not at all* to *A great deal)*
    - "I had to wait a long time before I could get past the ad." (from *Not at all* to *A great deal)*
    - "The ad was too personal." (from *Not at all* to *A great deal)*

Finally, the construct used in the ranking exercise was changed to *ad interference* (e.g., "Which ad MOST / LEAST interfered with your ability to watch the video?"). *Ad interference* with the video-watching task was evaluated, rather than *webpage preferences* (from the Multi-Ad study),

since pilot studies using the latter resulted in noisy data. Qualitative studies indicated this was because the broader *webpage preferences* wording caused participants to consider different factors when completing the exercise (e.g., their interest in the main video content and the ad creative—i.e., the message or story the ad told, the ad's length, the site's visual appeal, etc.), depending on what was most salient to each person.

Modifying the Study Protocol

In addition to modifying survey questions, the study protocol was changed to address issues that arose during pilot studies.

To minimize participant fatigue and drop-off, the study was shortened by removing the control video. Participants already spent around 7 minutes watching the three main videos, and the control does not affect the stankrank (the stackrank ranks *ad experiences* based on the degree to which each interferes with participants' ability to watch the main video).

Furthermore, some participants in qualitative study pretests thought they were required to watch the entire ad as part of the experiment, regardless of their real-world behaviors. To correct this misperception, a phrase was added to the study overview hinting that they could skip ads if they normally do so.

Experiment Staging

Similar to the Multi-Ad experiments, the experiences were studied in three stages per platform, which maximized the ability to detect differences among the ad experiences using the fewest participants. Each successive stage refined the stack-rank from the previous stage.

Stage 1: Creating the initial stack-rank

Participants viewed three videos, each containing one ad experience selected from the full set, meaning each participant viewed three total ad experiences. The three were chosen by ensuring that, across the full set, the number of pairwise comparisons was balanced. For example, 6 second non-skippable pre-rolls appeared with 6 second non-skippable mid-rolls just as often as they did with popups. The number of participants was chosen such that each pair of ads was evaluated by roughly 16 participants, to achieve sufficient confidence of the differences in perceptions of the two ads. Given each participant's ranking of the three ads they saw, the Bradley-Terry algorithm was used to aggregate the rankings across all participants to create an initial stack-rank based on perceived interference (see Appendix).

Stage 2: Refining the stack-rank

The stack-rank generated in Stage 1 generally showed differences in perceived interference rank scores among the highest and lowest ranked experiences, but those that fell close to each other were often more difficult to differentiate since their confidence intervals overlapped substantially. In Stage 2, the sets of ads participants saw were therefore restricted to only include those that were within five ranks of each other. For example, a participant could see a

set of ads that were ranked 3, 4, and 7 but not ones that were ranked 3, 4, and 16. The data were then combined from Stage 2 with that from Stage 1 to create an updated stack-rank.

Stage 3: Further refining the stack-rank

In Stage 3, the stack-rank from Stage 2 was further refined by comparing experiences whose ranks were within three places of each other. For example, a participant could see a set of ads that were ranked 3, 4, and 6 but not ones that were ranked 3, 4, and 7.

In each stage, 640 mobile and 1,232 desktop participants were used. The total number of desktop participants was higher because more desktop ad experiences were tested.

Using the Bradley-Terry (B-T) Algorithm

The Bradley-Terry (B-T) algorithm (Turner and Firth, 2012) was used to combine each participant's ranking of the three ads they saw into a stack-rank of all ad experiences. The B-T algorithm is a statistical method that takes pairwise comparisons between a set of items and estimates the latent "ability" parameters for each item. In the context of this study, the ability of an ad refers to its perceived interference; the greater the ability, the smaller the perceived interference. The estimates of perceived interference were used to rank the ads, and the confidence intervals allowed us determination of whether the differences between ads were statistically significant.

Study Participants

Answers Research, a market research firm, was used to recruit 1,920 mobile users and 3,696 desktop users in the US (studies replicating the research in additional countries are underway but are not included in this whitepaper). More desktop participants were used because 6 additional ad experiences were tested on desktop (different overlay variations) that are uncommon or unused on mobile. Answers Research was chosen because of their ability to reach a large, representative sample of US internet users.

The company emailed individuals who had signed up to participate in studies and who matched target criteria (i.e., device type and broad demographic traits matching that of the US internet population). Each email included a link to the study and how long it would take.

A breakdown of participant demographics appears in the Appendix.

Material

Overview

The experiments varied the ad experiences while using a fixed set of publisher videos, video ad creatives, and survey questions. Each participant saw each publisher video, ad experience, and ad creative once. We used a set of publisher videos and ad creatives that were similarly interesting and counterbalanced the publisher videos and ad creatives across the experimental conditions, so any differences in ad perceptions could be attributed to the ad experience. The

stimuli were identical on mobile and desktop, except we dynamically resized the webpage layout to fit common mobile and desktop screen sizes.

Ad Creatives

Video ad creatives were selected from three real brands, after pre-testing the creatives to confirm they were suitable. First, brands were chosen from different product verticals that were relevant to a broad demographic (e.g., deodorant, laundry detergent, soft drink), then survey pre-tests were run to confirm the creatives were similarly interesting. This was important to minimize the possibility that a specific ad creative might bias users' perceptions of an ad experience, since the goal was to identify *ad formats and experiences* that generally interfere with participants' video-watching experience.

In total, 13 video ad creatives were used: 3 brands, each with 4 different length ads (i.e., 6, 15, 30, and 60 sec), plus an extra 15 second creative in a fourth brand to test the "ad pod" experience (i.e., two consecutive 15 sec non-skippable ads). Six second ads were created by shortening and editing existing creatives to convey a coherent message within a short timeframe. For the longer ads (i.e., 15, 30, 60 second variations), the advertisers' original creatives were mostly used. The only modifications were shortening them by removing a few seconds from either the start or the end of the original creative.

In addition to the video ad creatives, static overlays ads were also created in various industry standard banner sizes (e.g., 728x90 on desktop) to use in the desktop experiments. They were designed to match each of the three video ads' aesthetics and advertising campaigns.

Figure 3. Examples of three video ad creatives (from left to right): video ad, skippable ad, and large overlay ad.

Main Videos

To maximize user engagement and minimize the potential for noise, videos were chosen about general topics that fulfilled two “interest” criteria. First, each video had to meet an “interestingness” threshold so participants would be engaged during the study. Second, all three videos had to receive similar “interesting” ratings. Pretests indicated the videos met both criteria: around 80% of respondents rated each video as “Moderately”, “Very”, or “Extremely” interesting.

It was also important to confirm that the main videos were generally more interesting than the ads, which is the typical relationship in real-world scenarios.

Comprehension, Overall Experience, and Ad-specific Questions

After watching each video, participants answered a comprehension question to confirm they paid attention to the main video. The reasonably high comprehension rate (87% across all three videos and stages for both desktop and mobile) suggested they were attentive during each video.

After each video, participants answered overall experience and ad-specific questions. The overall experience questions covered overall satisfaction and web page behavior predictability. The ad-specific questions covered annoyance, distraction, enjoyability, relevance, creepiness, the timing of the ad’s appearance, ad length, and wait time before the ad became skippable. To ensure participants paid attention while answering the questions, attention-check questions were included after the first and third videos (e.g., *“We sometimes include questions in our study to ensure respondents are paying close attention. For this question, select Never as the answer regardless of how you would normally answer it. How often do you check your email?“*). Responses were filtered out from participants who answered either of these questions incorrectly.

Final Ranking Survey

After watching all three videos and filling out the surveys, participants completed an exercise in which they ranked the ad experiences by which most and least interfered with their ability to watch the video. The exercise showed screenshots of the three ads to refresh their memory. Responses were filtered out from participants who chose the same ad experience as the most and least interfering.

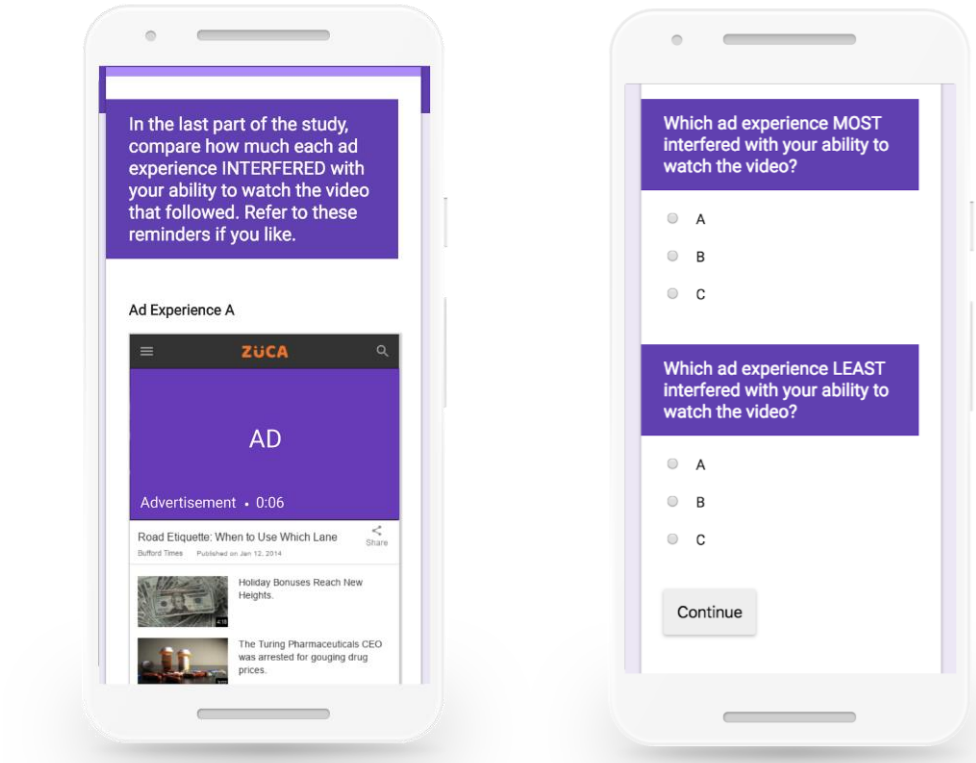


Figure 4. The top and bottom of the final ranking survey in the instream video experiment.

After the ranking exercise, there was an “audio check” question that required participants’ device audio to be turned on for them to be able to answer it. An audio recording was played that told participants the answer to the question, then asked them to select it from a multiple choice set, confirming their sound was on. Responses were filtered out from those who chose the wrong answer. This was important because audio is a critical part of the video ad experience.

Results and Discussion

Overview

This section first presents the overall ranking results, followed by the important user experience (UX) metric scores. The ranking shows how much the ads interfere with users’ video-watching experience, while the UX metrics help us understand why ads are ranked in a certain way (for example, does an ad rank poorly because it is too long or because it requires users to wait a while before they can skip it?). These metrics are also a good measure of the broader user experience because they evaluate different video ad attributes.

Overall Ranking Chart

Each ad experience has a rank score estimate with a 95% confidence interval. In order to

normalize perceived interference, a score of 1 was assigned to the most interfering ad experience and a score of 5 to the least interfering ad experience.

For ads in which the error bars overlap, the rank scores are similar. This means that perceived interference among the ads do not significantly differ.

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

Mobile Results

Mid-rolls (different lengths and skip delay times) and 60 sec non-skippable pre-rolls rank as the worst ad experiences (bottom 25% of the 16 tested ads). A 6 sec non-skippable pre-roll and a popup that appears before the video starts, inside the video player, are the best ad experiences. The 6 sec non-skippable pre-roll is significantly better than the equivalent mid-roll variation, which differs only by the timing of when the ad was shown during the video. The graph also reveals other insights about which characteristics make an ad rank better or worse.

Instream Video Mobile Ad Experience Rankings

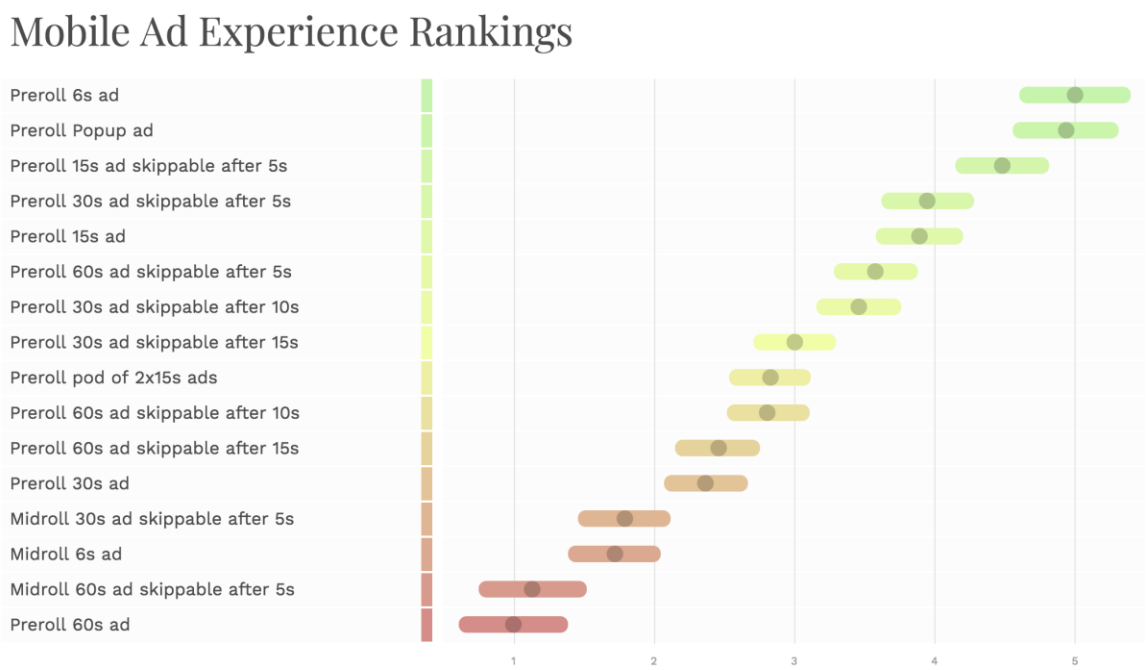


Figure 5. Rank scores of the 16 mobile ad experiences, based on perceived interference with the video-watching experience and scaled from 1 (interferes most) to 5 (interferes least).

Making the user wait to skip makes an ad experience rank worse.
Ad experiences with longer “skip delay times,” which is the duration of the ad users had to watch before they could proceed to the publisher video, generally rank worse within each pre-roll length (e.g., 15, 30, 60 sec). For 15 sec pre-rolls, we tested unskippable and 5 sec skip

delay times. For the longer 30 and 60 sec pre-rolls, unskippable and 5 sec skip delays were tested along with longer 10 and 15 sec skip delay times. The findings match expectations.

Long ads are especially bad.
Among the 6, 15, 30, and 60 sec non-skippable pre-rolls, each successively longer ad ranks significantly worse than all shorter variations, which relates to the above finding. Furthermore, the 60 sec non-skippable pre-roll is especially bad compared to the same length pre-roll with shorter skip delay times. This is reflected by the larger range in rank scores between non-skippable pre-rolls and those that were skippable after 5, 10, or 15 sec, which is larger for 60 sec ads than for 30 sec ads.

Timing matters: ads that appear while the video plays are worse.
This pattern can be observed among the different ad experience categories we tested (i.e., mid-rolls, pre-rolls, popup). For example, when comparing each of the three mid-roll ads to its pre-roll equivalent (e.g., 30 sec non-skippable mid-roll vs. 30 sec non-skippable pre-roll), the mid-roll always ranks significantly worse. On the other hand, the popup ranked as one of the best ads, since it appears before the video starts. These patterns are intuitive, since the ads that play while participants watch videos disrupt their focus.

Ad length, "unexpected appearance," and wait time are most predictive of rating differences across all ad experiences.

Of the 10 metrics that were collected, a Partial Least Squares Regression analysis showed an ad's length, unexpected appearance during the video, and wait time most predict an ad experience's rank score. The next most predictive metrics are annoyance, distraction, and the ad creative's enjoyability (the last three have roughly equal predictive power). See Appendix for more details.

Ad length, "Unexpected appearance," Wait time, Annoyance, Distraction, and Ad creative enjoyability
The six figures below order the ad experiences by their position in the overall rank (from the least interfering to the most interfering). The stacked bars represent the distribution of participant ratings on a 1-5 scale for each ad metric.

Ad length directionally aligns with the overall interference ranking (i.e. longer ads generally rank worse). The 6 second non-skippable mid-roll does not follow this general pattern.

There are also two interesting findings related to ad length. Among mid-rolls, participants generally feel each is longer than its pre-roll equivalent (e.g., the 6 second mid-roll is perceived to be longer than the 6 second pre-roll), suggesting an identical format with a more obtrusive placement can make the ad "feel longer". In addition, participants generally believe a pod, which consists of two consecutive non-skippable 15 sec ads before the main video, is shorter than one 30 sec non-skippable pre-roll ad, even though total ad duration is the same in both cases

(Figure 6). Perhaps this is because the two ads in the pod break up the amount of time participants have to focus on a single ad.

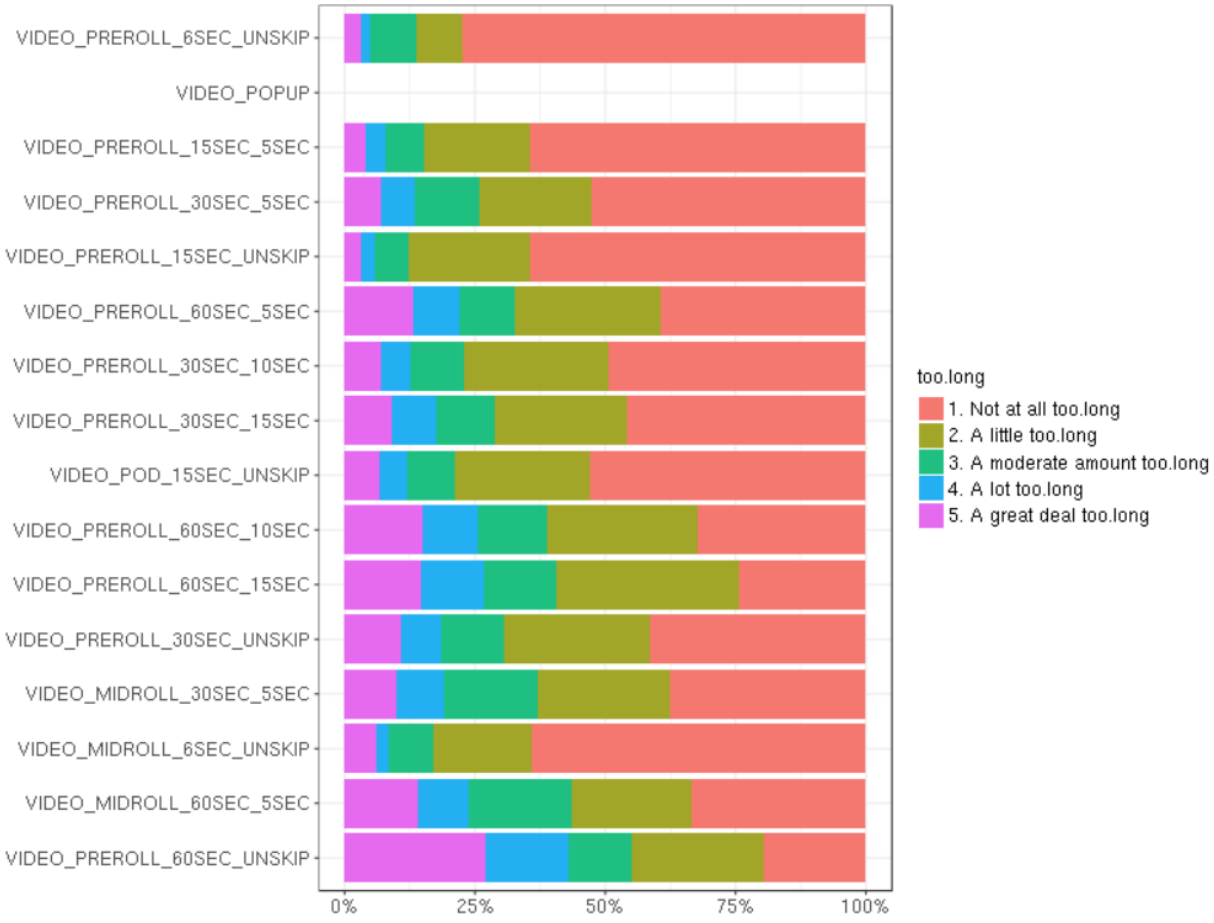


Figure 6.

The **"unexpected appearance"** metric is noticeably worse for the three mid-rolls near the bottom. All other ads (various pre-rolls, a popup) are perceived more favorably and have similar ratings. These patterns are intuitive, since mid-rolls appear during the video (Figure 7).

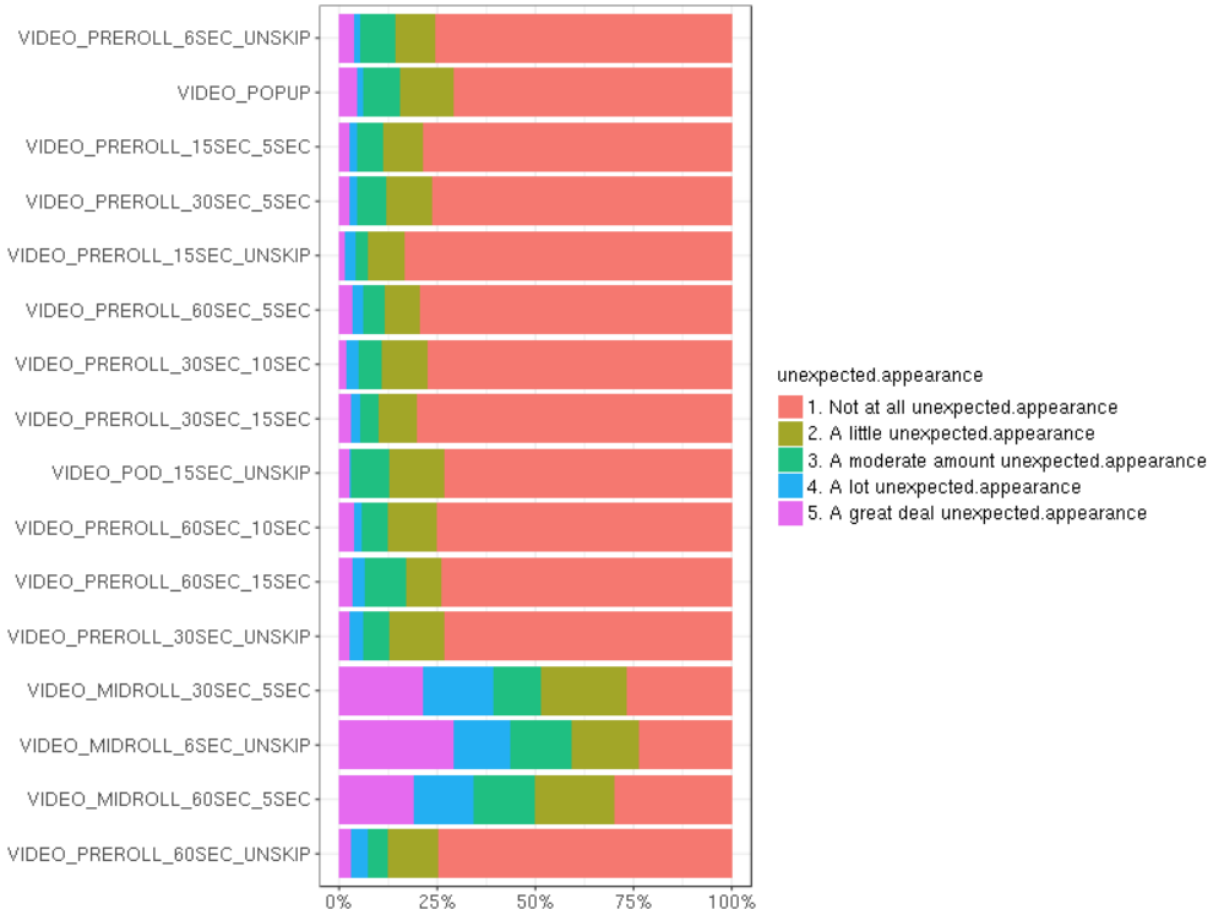


Figure 7. Unexpected appearance

"Wait time" aligns reasonably well with the overall interference ranking. The 60 second non-skippable pre-rolls (one of the worst ranking ads) receive especially low ratings. In general, perceived wait time increases when participants have to watch long, non-skippable ads (e.g., 30, 60 sec) or when they have to wait before they can skip long ads (e.g., waiting 15 sec before they can skip a 30 or 60 sec ad) (Figure 8).

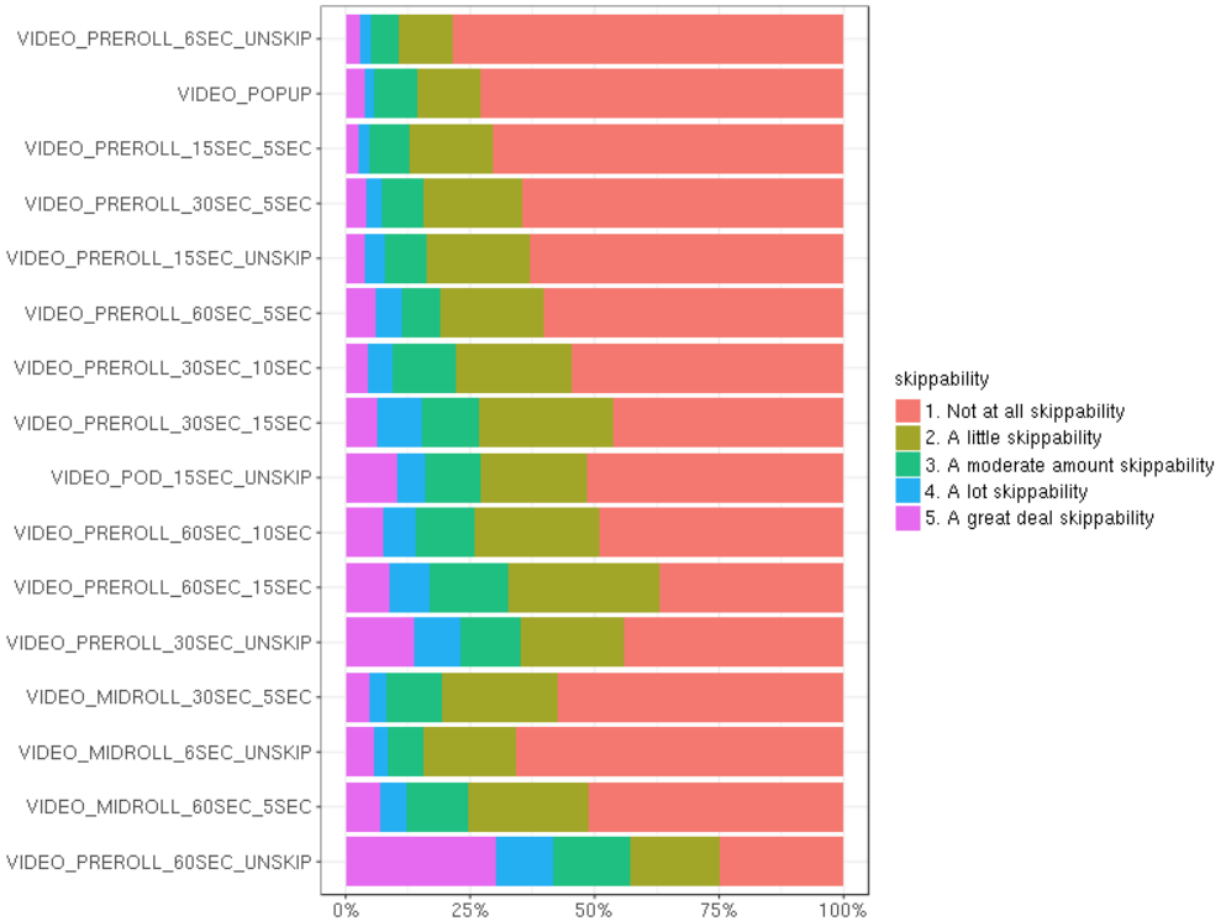


Figure 8. Skippability

Annoyance and **distraction** show similar patterns. Both metrics are slightly correlated with the overall interference ranking. The three mid-rolls and the 60 sec pre-roll at the bottom receive slightly lower ratings than all other ads. In addition, each mid-roll receives worse ratings than its pre-roll equivalent, especially among shorter ads (e.g., 6 sec non-skippable mid-rolls vs. pre-rolls).

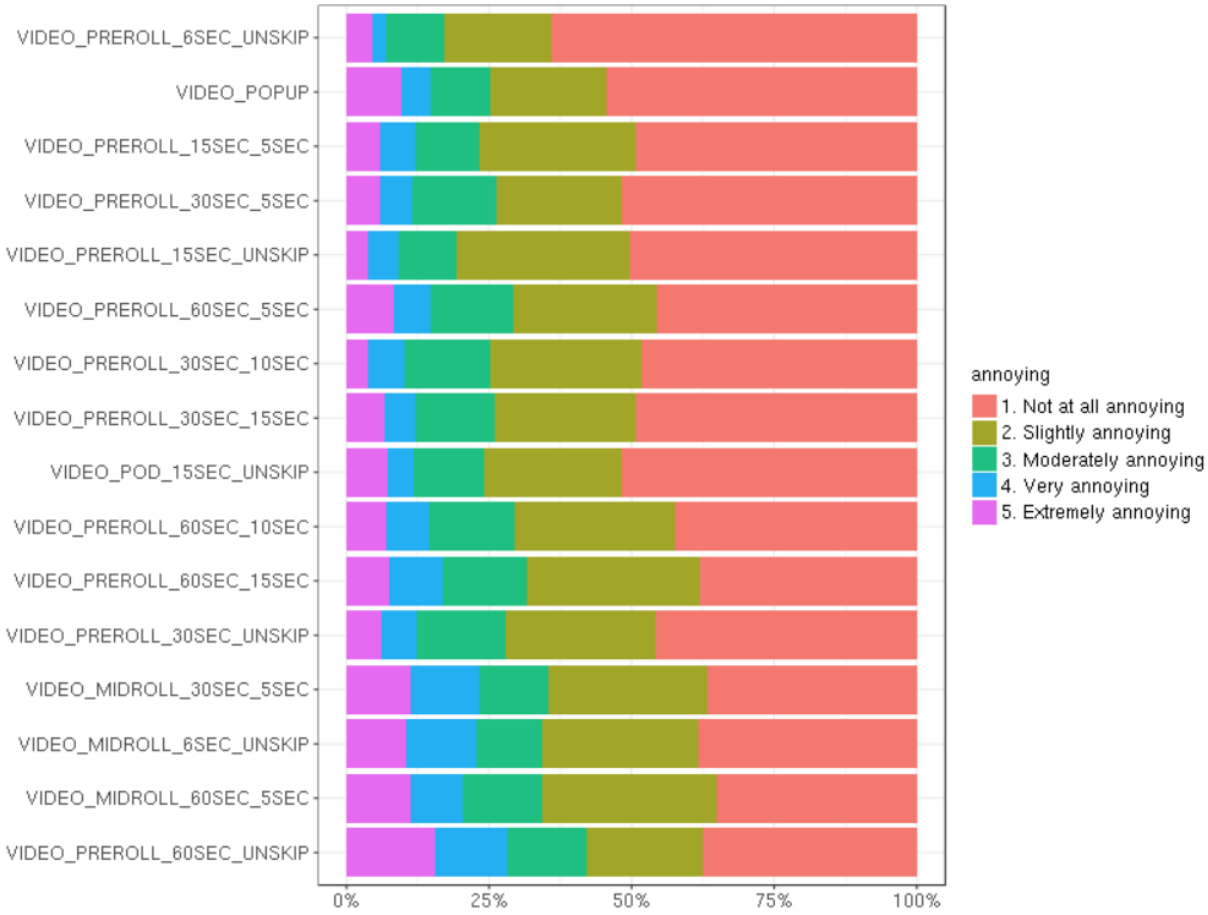


Figure 9. Annoyance on mobile

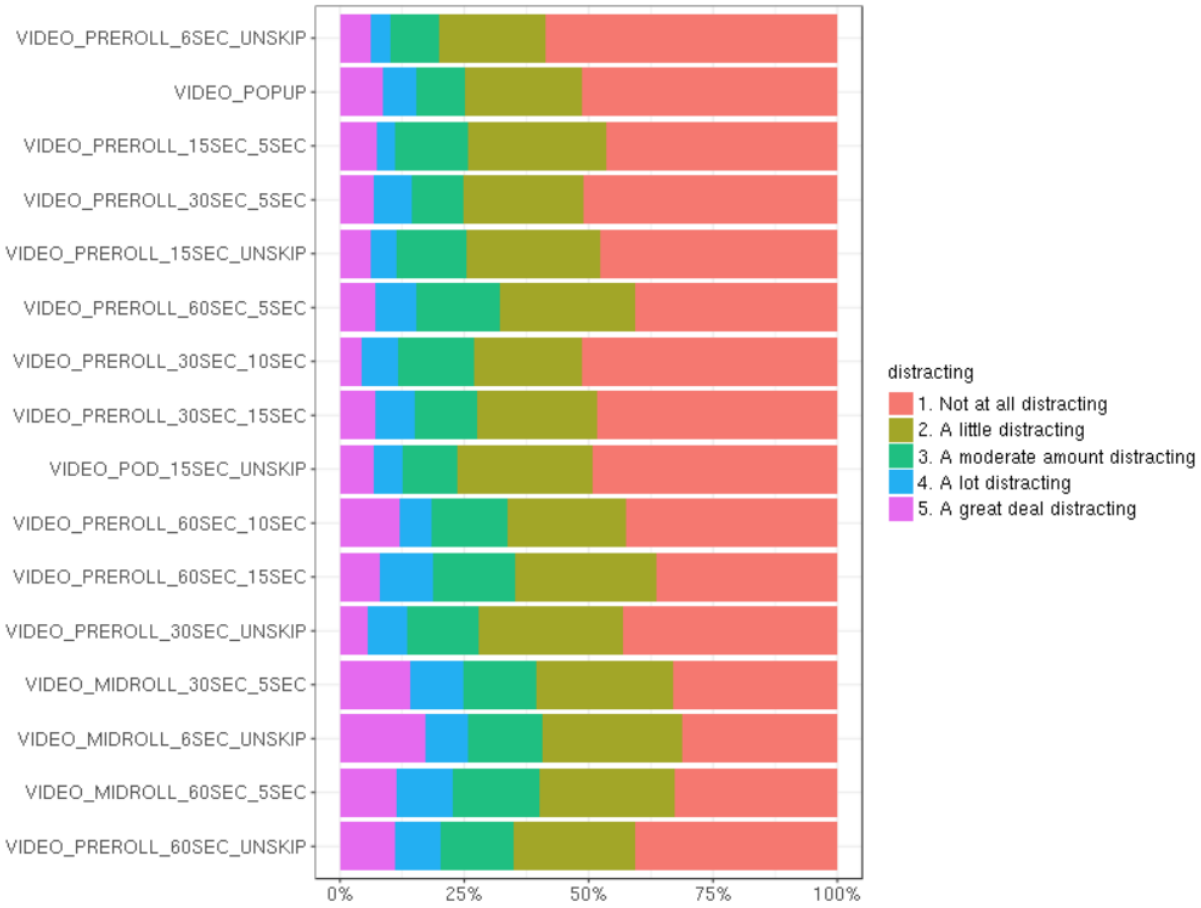


Figure 10. Distraction on mobile

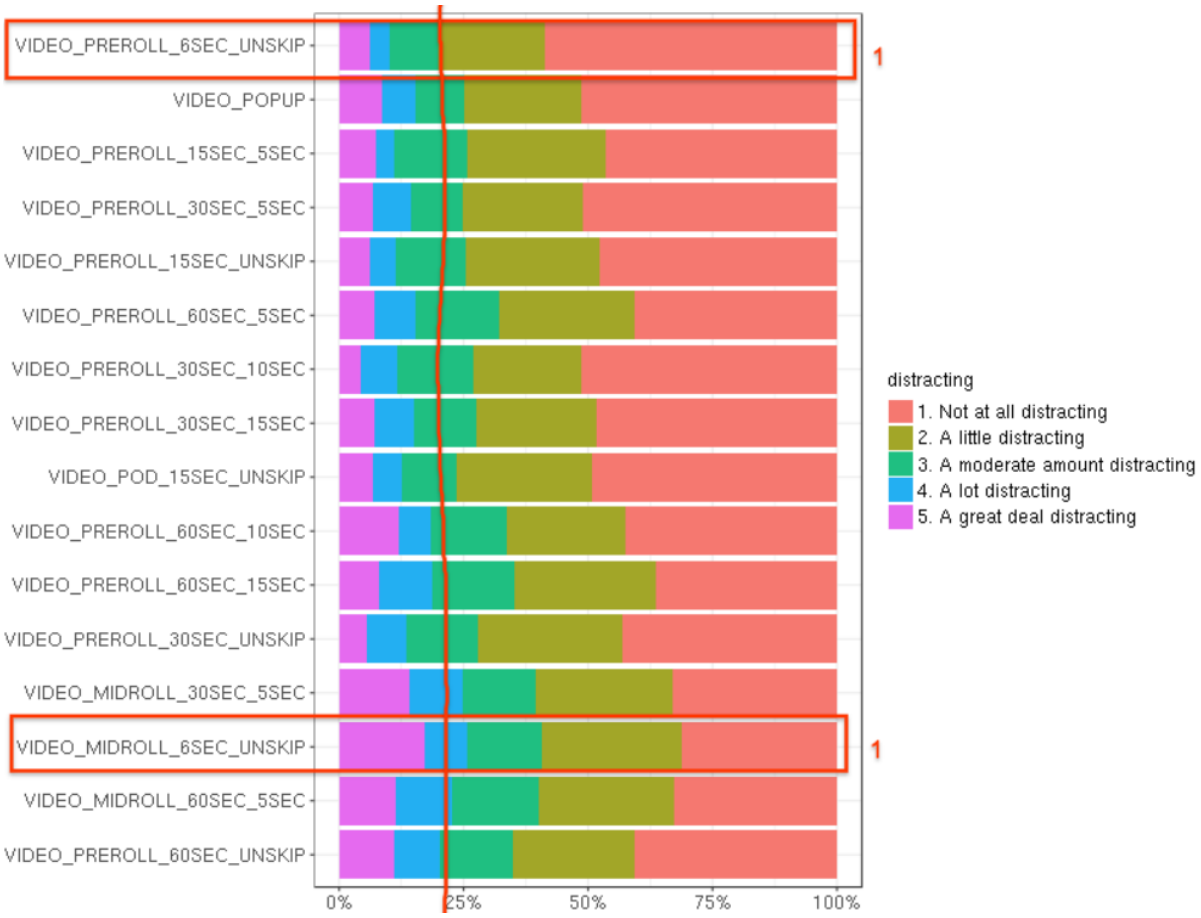


Figure 11. 6 sec mid-roll vs pre-roll

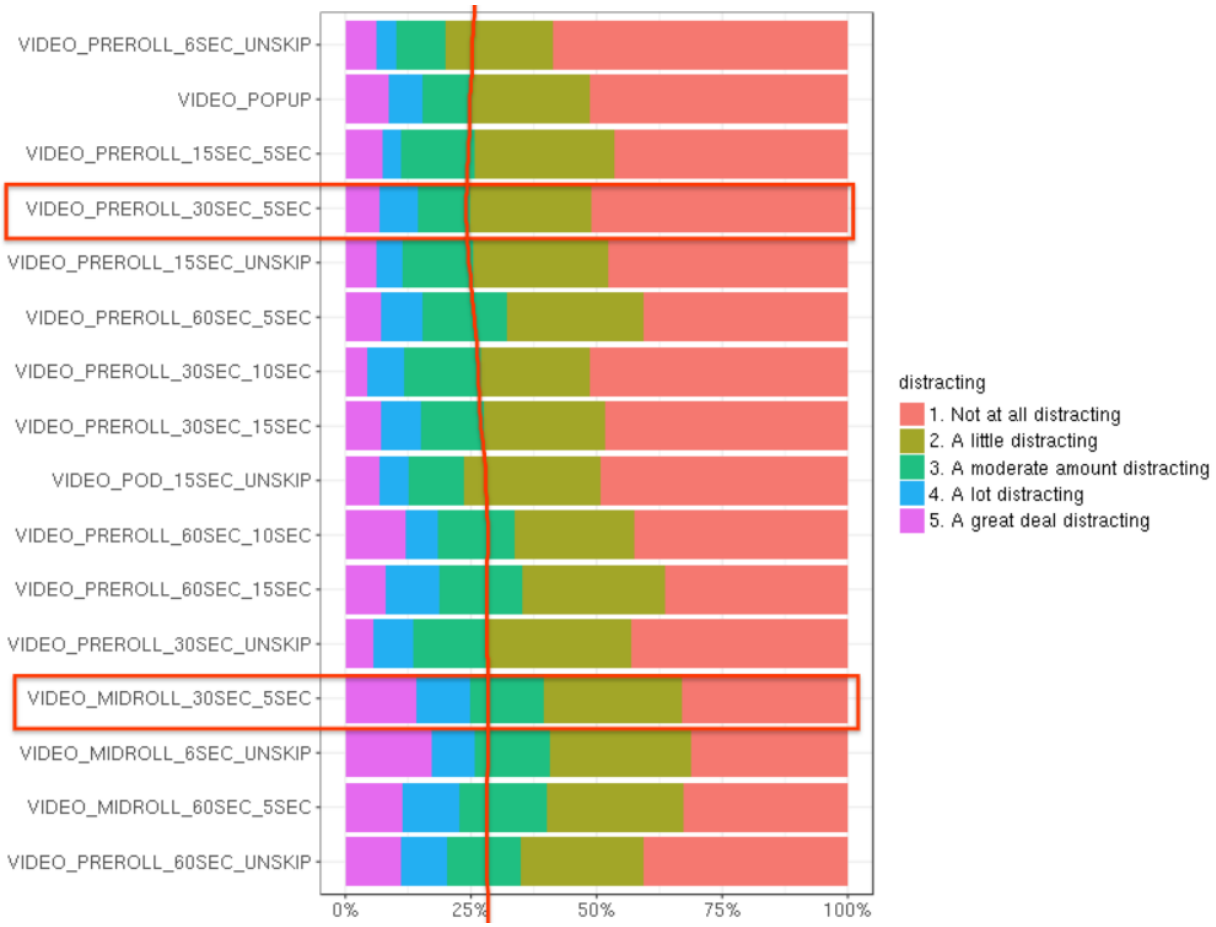


Figure 12. 30 sec mid-roll vs. pre-roll

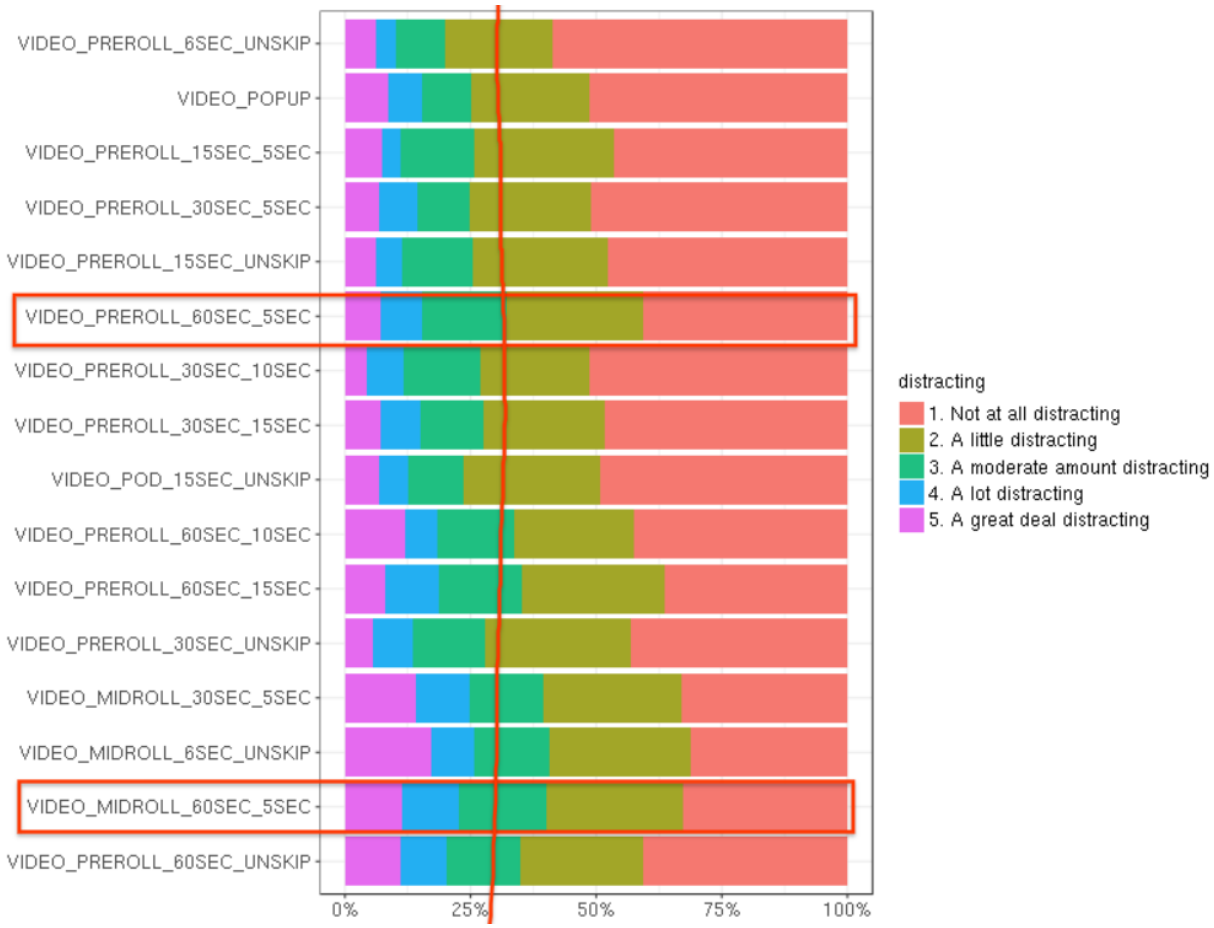


Figure 13. 60 sec mid-roll vs. pre-roll

Ad creative enjoyability is the final metric that relates to the overall interference ranking. Although it is fairly consistent across the ad experiences, the 6 sec non-skippable mid-roll and the 60x sec non-skippable pre-roll rate slightly lower than all other ads. The 6 sec non-skippable mid-roll also rates notably lower than its pre-roll equivalent, suggesting a short, identical format with a more obtrusive placement can negatively impact how much a user enjoys watching it—for reasons similar to the above (Figure 9).

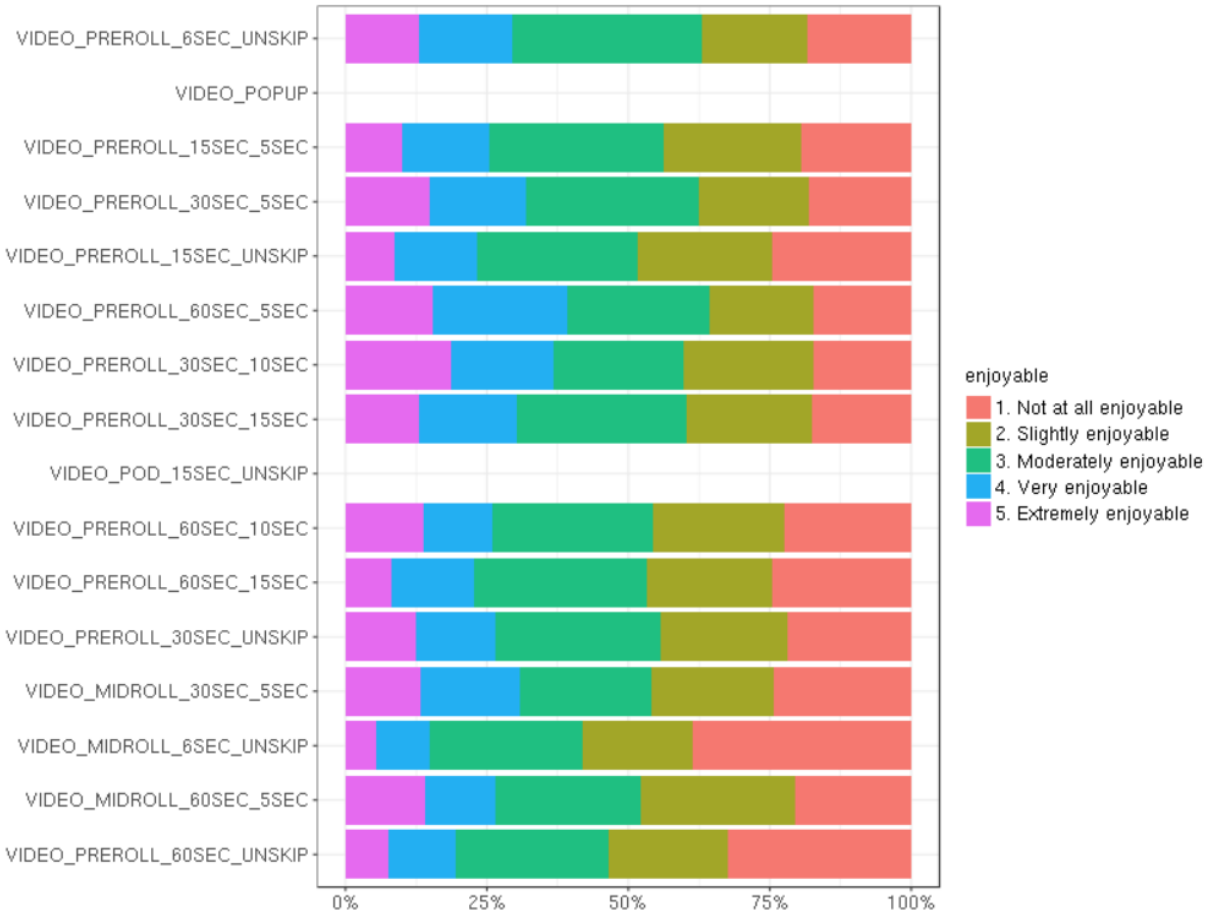


Figure 14. Mobile enjoyability

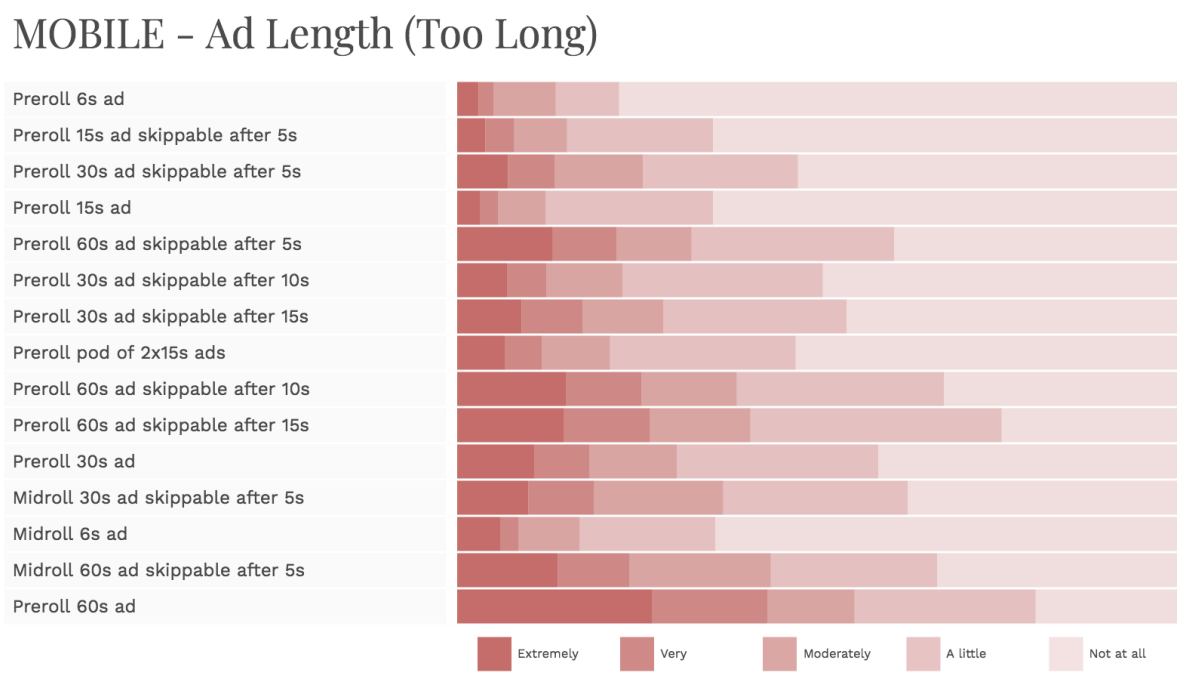


Figure 15. Ad length rating distribution for each ad experience in the instream video mobile web experiment.

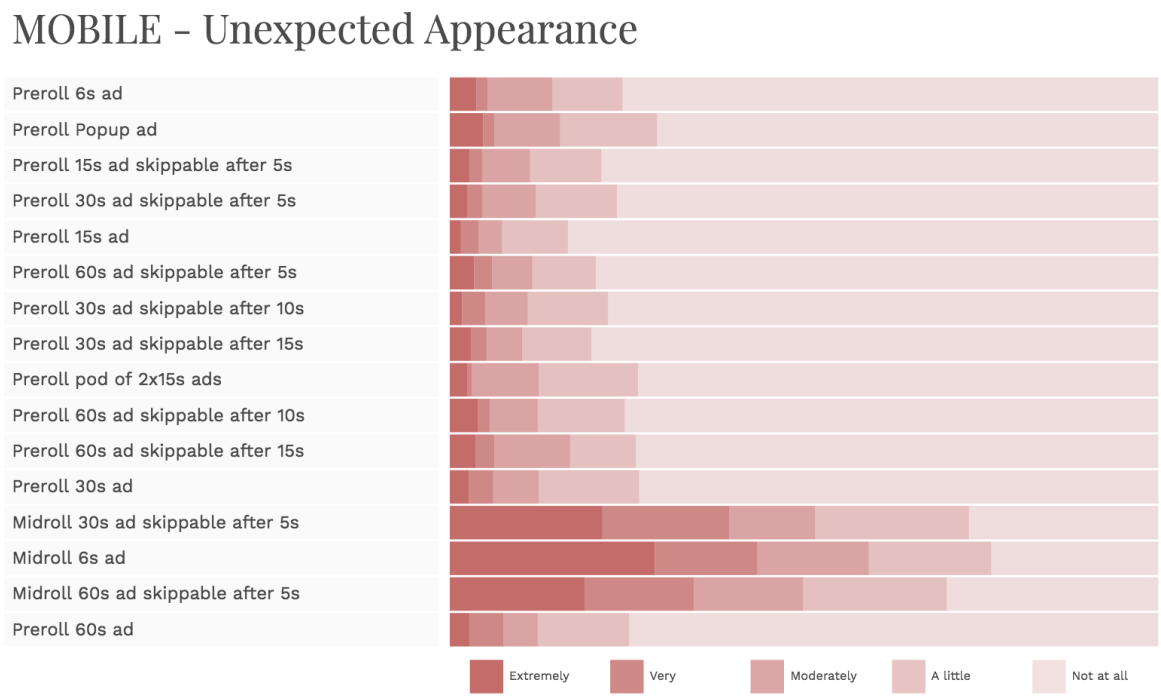


Figure 16. "Unexpected appearance" rating distribution for each ad experience in the instream video mobile web experiment.

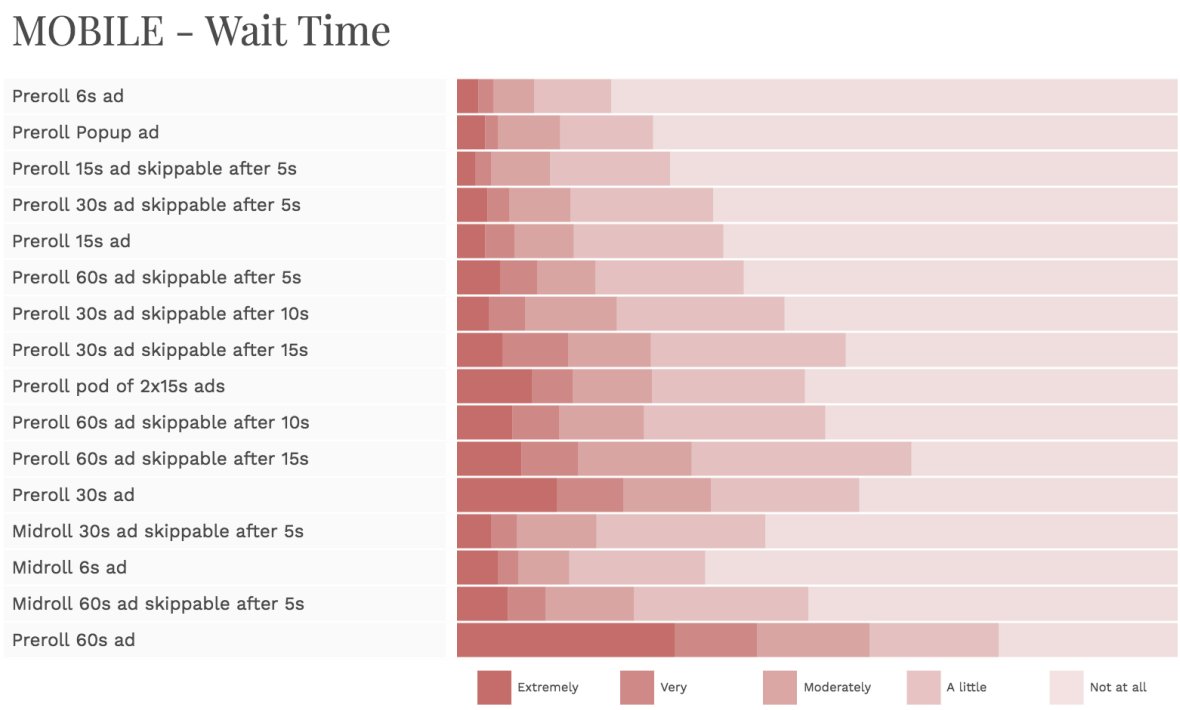


Figure 17. "Wait time" rating distribution for each ad experience in the instream video mobile web experiment.

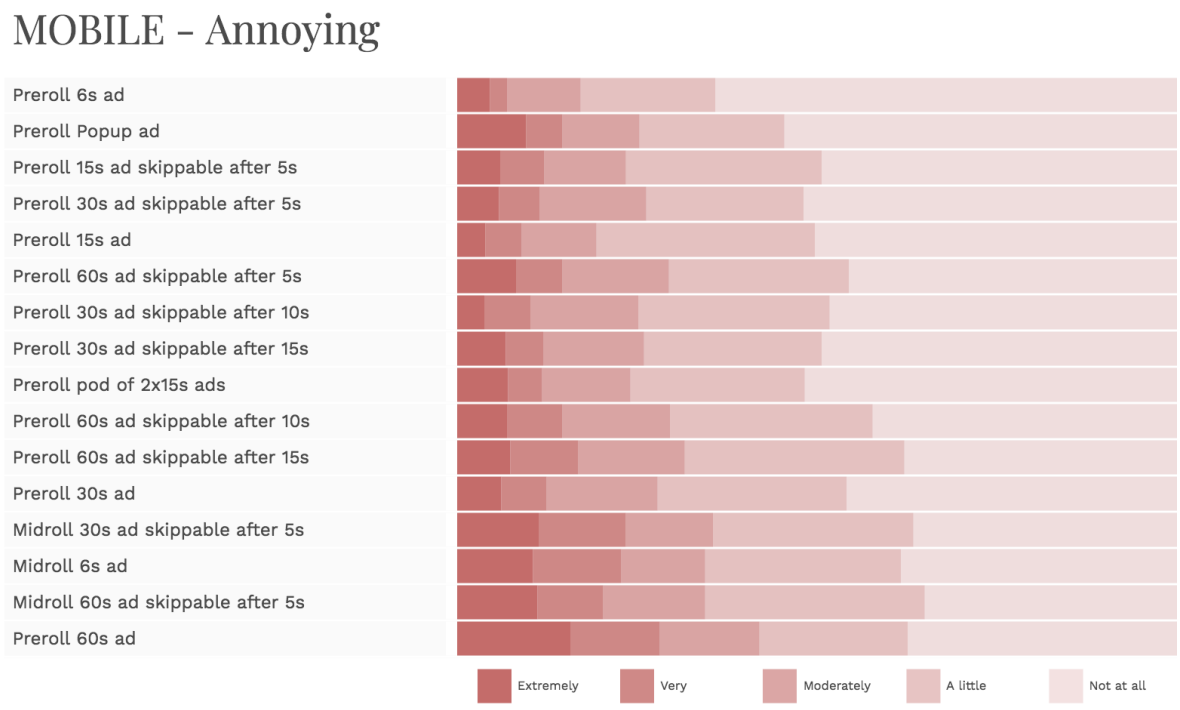


Figure 18. "Annoyance" rating distribution for each ad experience in the instream video mobile web experiment.

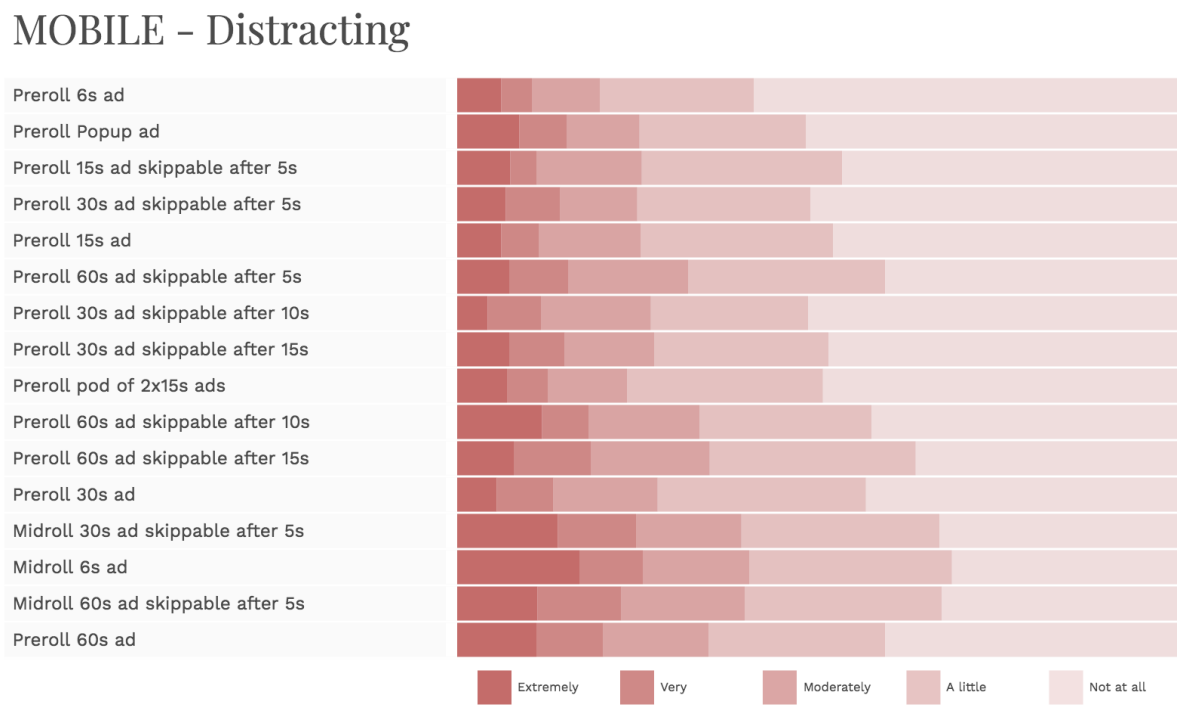


Figure 19. "Distraction" rating distribution for each ad experience in the instream video mobile web experiment.

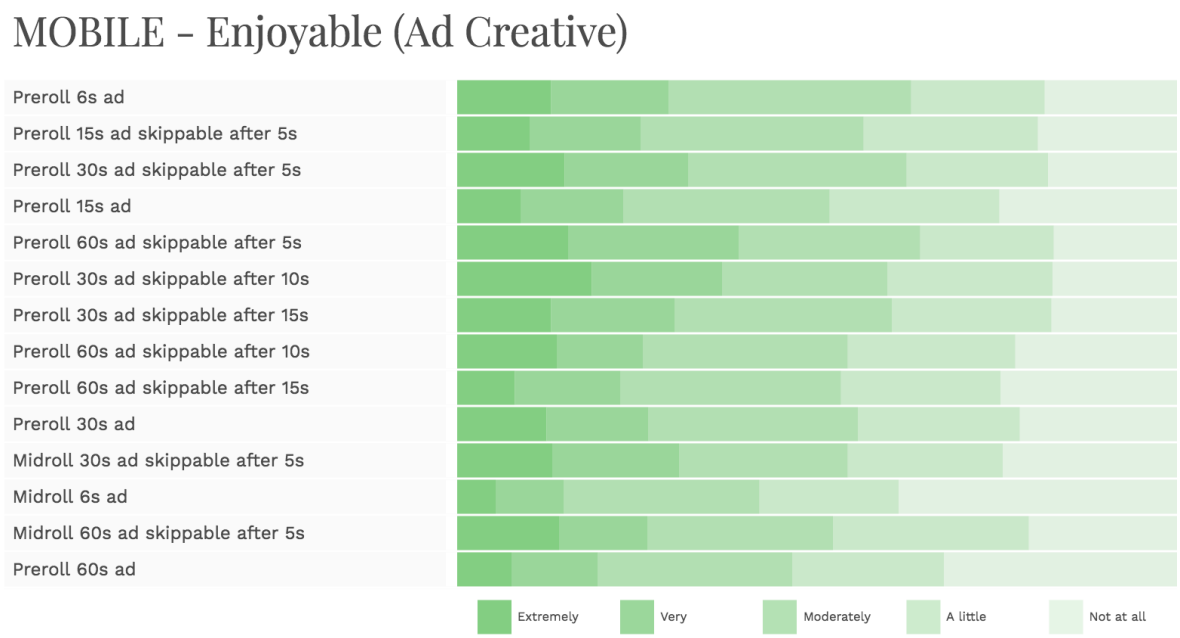


Figure 20. "Ad creative enjoyability" rating distribution for each ad experience in the instream video mobile web experiment.

The Other Metrics: Satisfying, Relevant, Predictable, Personal
For the user experience metrics of Satisfying, Relevant, Predictable, Personal, most ads generally rate similarly to each other (the graphs visualizing these metrics can be found in the appendix). Relevance is the only metric that shows some variability. The 6 sec non-skippable mid-roll receives slightly lower scores compared to the other ads, and to its pre-roll equivalent. This is interesting because one might expect relevance to relate mainly to the ad creative, not to the format. Similar to previous explanations, the mid-roll's more interruptive placement may make it harder for participants to focus on the ad creative and, therefore, determine whether it is personally relevant to them. [graph]

Desktop Results

The "worst" ad experiences (bottom ~25%) are mid-rolls (different lengths and skip delay times), a large overlay that appears during the video, and long pre-rolls that are non-skippable or have long skip delay times (e.g., skippable after 15 sec). Of the ads that were tested on both platforms, the ones that are in the bottom 25% on mobile are also in the bottom ~25% on desktop (i.e., the three mid-rolls and the 60 sec non-skippable pre-roll), though in a slightly different order.

The best ad experience is an overlay that appears after the video ended, by a large margin. This ad is significantly less interfering than the overlay variations that appear before or during the main video. The graph also reveals other insights about which characteristics make an ad rank better or worse, which are very similar to the mobile findings. [graph]

Instream Video Desktop Ad Experience Rankings

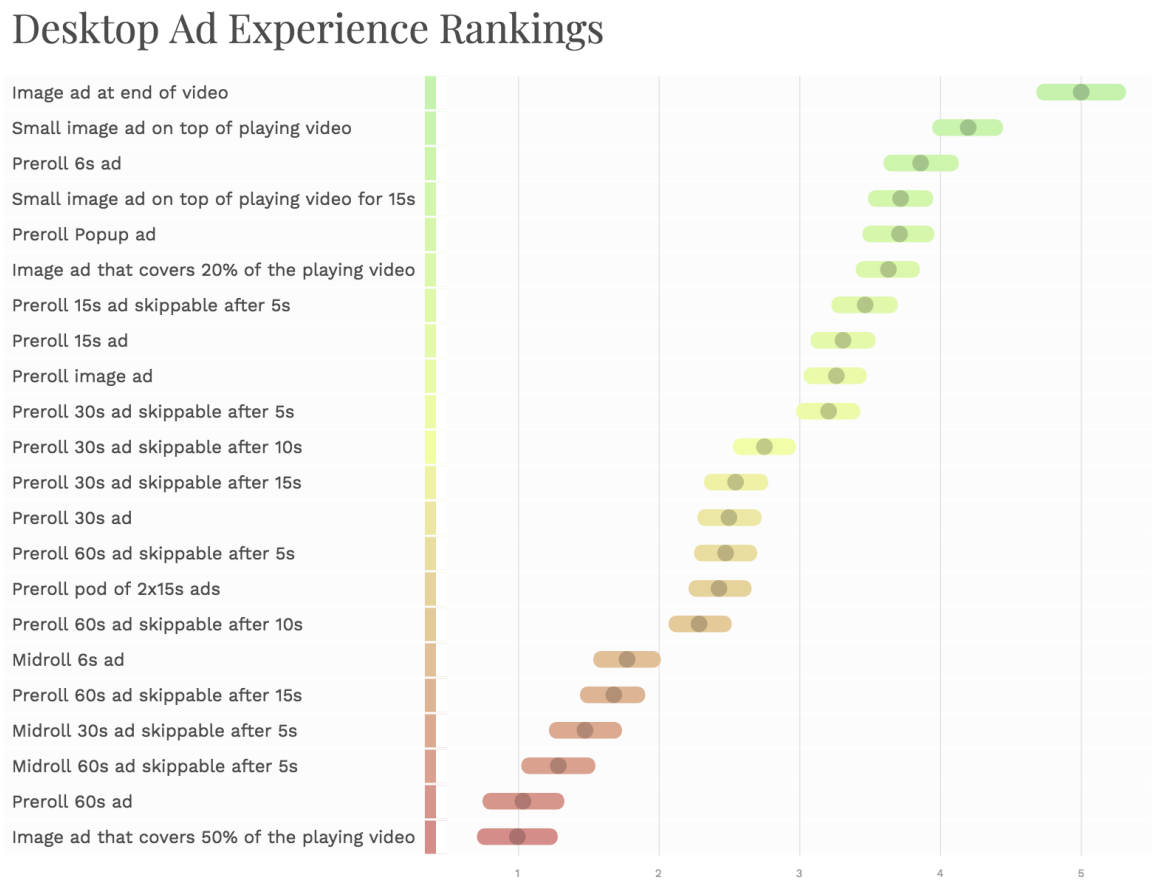


Figure 21. Rank scores of the 24 desktop ad experiences, based on perceived interference with the video-watching experience and scaled from 1 (interferes most) to 5 (interferes least).

Longer wait times make an ad experience rank worse.
As with mobile, desktop ads with longer skip delay times generally rank worse within each pre-roll length, particularly with 60 sec ads.

Longer ads also rank worse than shorter ads.
All other things being equal, the ranking of an ad is worse as its length increases.

Timing matters: ads that appear while the video plays ("mid-roll") are generally worse.
Similar to mobile, most of the worst desktop ads are ones that appear while the video plays, and each mid-roll is always significantly worse than its equivalent pre-roll format. One exception is the smaller overlays that appear during the video (i.e., the second and fourth ads from the top in Figure 12). Desktop study participants perceive them relatively favorably, likely because their small size and unobtrusive placement at the bottom of the video enable participants to keep watching the main video without much distraction.

Most overlays rank well.

Four out of the six overlays rank well (among the top ~25% of ads), likely because their placement after the video ends (e.g., end of video) or smaller size and placement (e.g., bottom 1/3 of the video frame) make them less interruptive while participants watch the video.

Ad length, "unexpected appearance," and wait time are most predictive of rating differences across all ad experiences.
Similar to mobile, the length of a desktop ad is the primary predictor of its rating, followed by unexpected appearance, and wait time (the latter two having roughly equal predictive power). The next three metrics are also the same as on mobile (i.e., annoyance, distraction, and enjoyability), though their importance differs slightly across the two platforms.

Ad length, "Unexpected appearance," Wait Time, Annoyance, Distraction, and Enjoyability Metrics
The six figures below order the ad experiences by their position in the overall rank (from the least interfering to the most interfering). The stacked bars represent the distribution of participant ratings on a 1-5 scale for each ad metric.

The **ad length** metric directionally aligns with the overall interference ranking (i.e. longer ads tend to rank worse), though the 6 second non-skippable mid-roll does not follow the general pattern. As with mobile, the mid-roll's interruptive placement likely makes it rank poorly.

Furthermore, the same interesting findings observed in the mobile experiments also arise in desktop. For example, participants generally perceive the pod with two consecutive 15 sec non-skippable pre-rolls as being shorter than a 30 sec non-skippable pre-roll, and that each mid-roll is longer than its pre-roll equivalent. [graph]

"Unexpected appearance" shows fairly large rating differences among different categories of ads (e.g., pre-rolls, mid-rolls, overlays). As on mobile, mid-rolls and the large overlay that covers 50% of the video while it plays have the worst ratings. The other five overlays receive much better ratings than the worst ads. Finally, pre-rolls generally receive the best ratings, regardless of ad length or wait time. [graph]

"Wait time" generally aligns with the overall interference ranking. The 60 second non-skippable pre-rolls again receive the worst ratings, followed by 60 second pre-rolls that are skippable after 15 sec. As with mobile, perceived wait time increases for long, non-skippable ads (e.g., 30, 60 sec), or long ads that require participants to wait a while before they can skip it (e.g., waiting 15 sec before they can skip a 30 or 60 sec ad). [graph]

Annoyance is the final metric that mirrors the overall interference ranking. The large overlay that appears during the video and covers 50% of the video player has the worst ratings overall. Interestingly, the 60 sec non-skippable pre-roll has better ratings than this ad, even though both have a similar rank score. This is likely because the pre-roll disrupts users less while they watch the video since it appears before, rather than during, the video. [graph]

Distraction shows a similar pattern to annoyance, with the trends described above present in the distraction ratings. [graph]

Similar to mobile, **enjoyability** ratings are fairly similar across all ad experiences, except for the 6 sec non-skippable mid-roll. [graph]

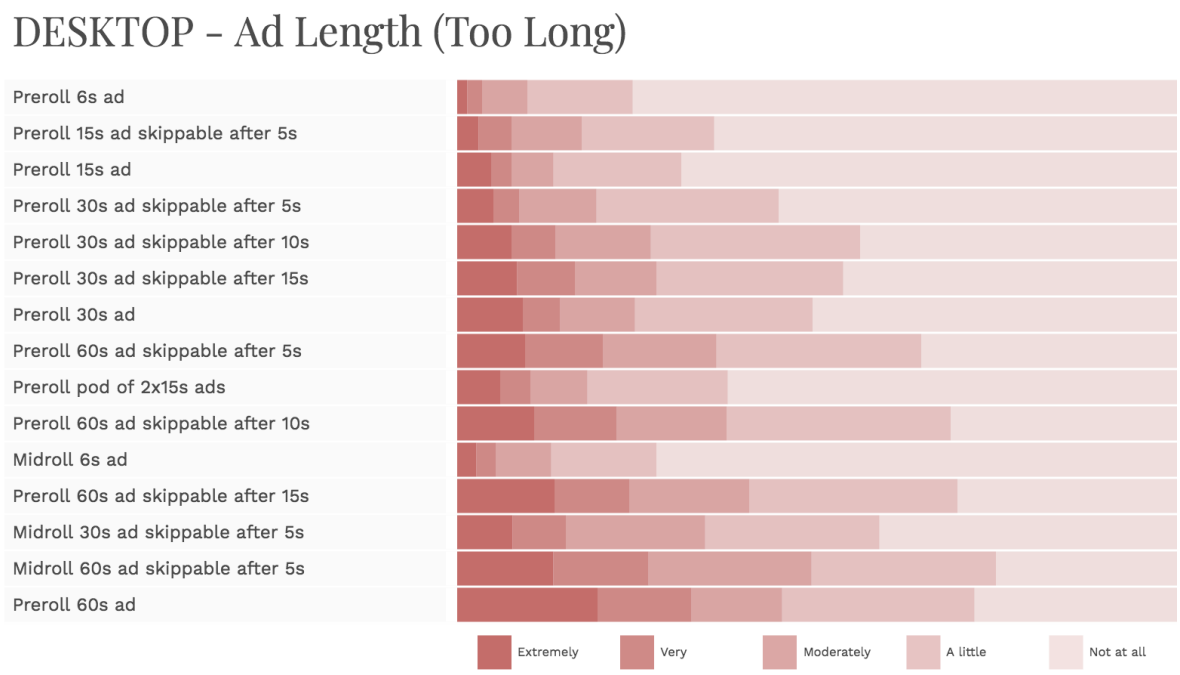


Figure 22. Ad length rating distribution for each ad experience in the instream video desktop web experiment.

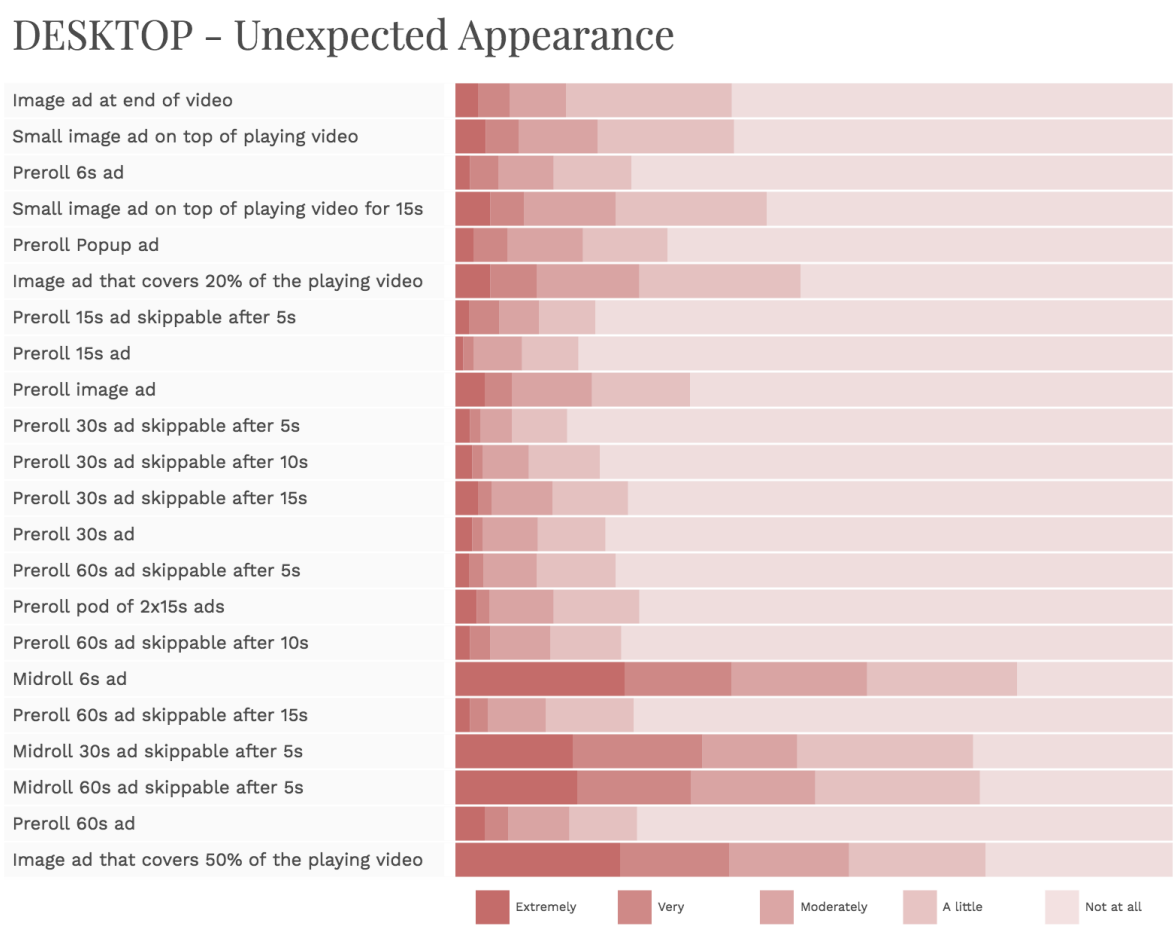


Figure 23. "Unexpected appearance" rating distribution for each ad experience in the instream video desktop web experiment.

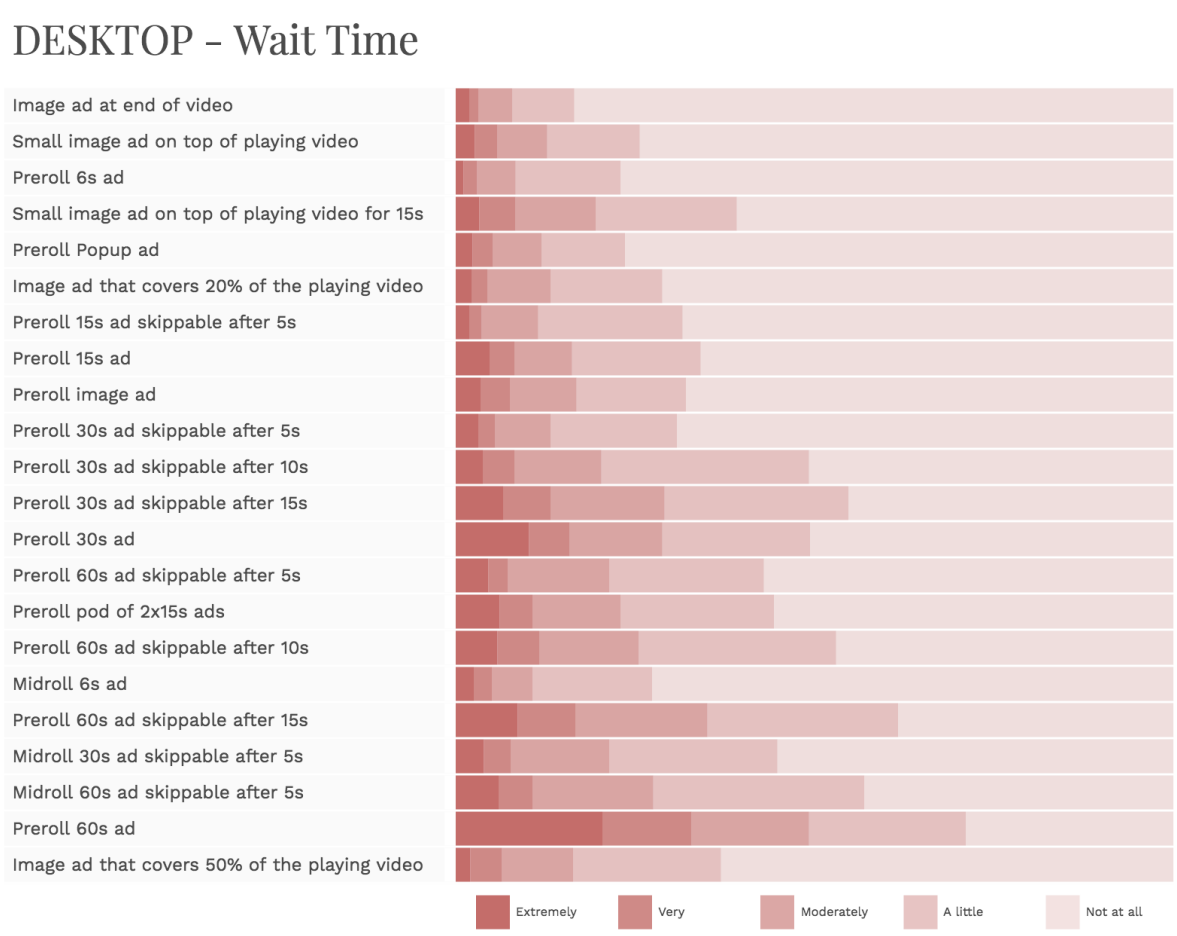


Figure 24. "Wait time" rating distribution for each ad experience in the instream video desktop web experiment.

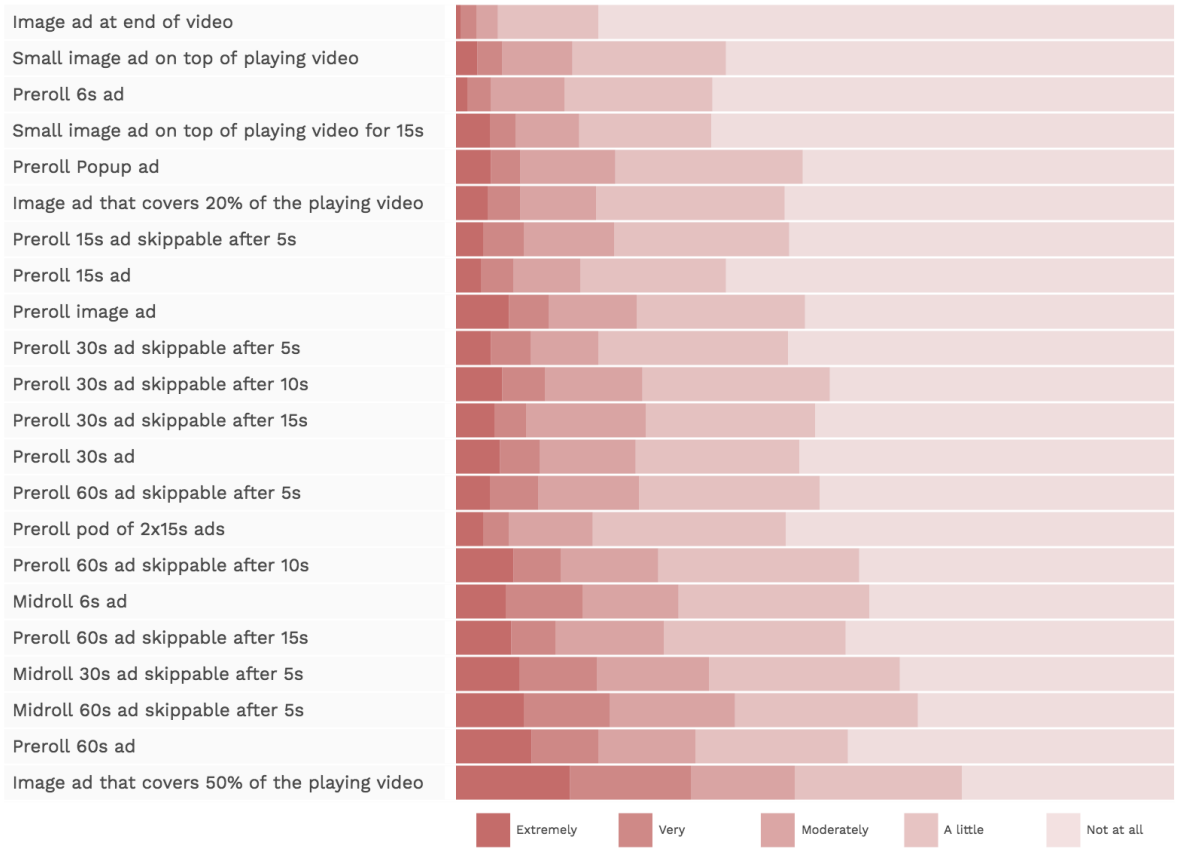


Figure 25. "Annoyance" rating distribution for each ad experience in the instream video desktop web experiment.

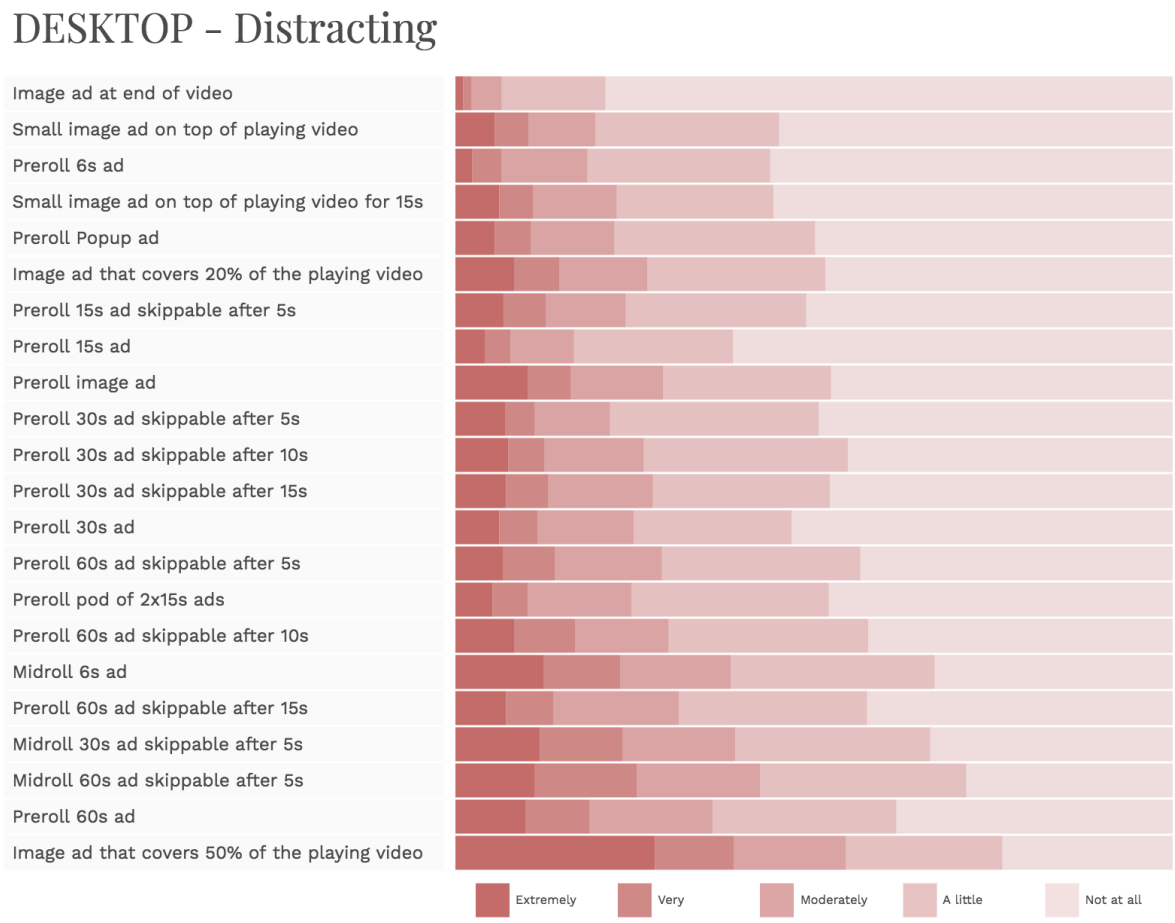


Figure 26. "Distraction" rating distribution for each ad experience in the instream video mobile web experiment.

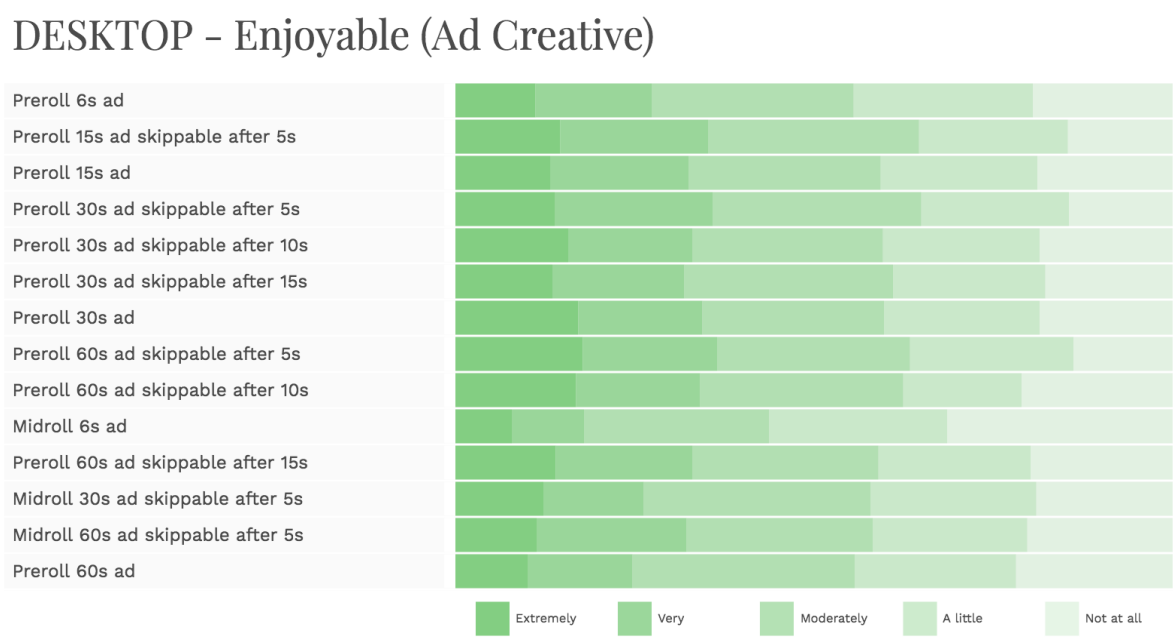


Figure 27. "Ad creative enjoyability" rating distribution for each ad experience in the instream video mobile web experiment.

The Other Metrics: Satisfying, Relevant, Predictable, Personal
For the other user experience metrics of Satisfying, Relevant, Predictable, Personal, most ads generally rate similarly to each other (the graphs visualizing these metrics can be found in the appendix).

Comparing Mobile and Desktop

Ad experiences that were common to mobile and desktop generally ranked similarly on both platforms. For example, the three mid-rolls and the 60 second non-skippable pre-roll are located in the bottom ~25% of ads in both environments. There are other consistent patterns across mobile and desktop, such as participants preferring each successively shorter non-skippable pre-roll to the next longer version.

Conclusions

This paper describes how the Multi-Ad Study methodology was adapted to evaluate perceptions of 40 instream video ad experiences in short-form 2-2.5 minute videos on the mobile and desktop web. Using this approach, ad experiences were efficiently ranked by perceived video-watching interference using an algorithm to optimize which experiences were shown to each participant and the Bradley-Terry model to estimate the overall ranking.

Results suggest certain ad experiences interfere more than others with users' ability to watch short 2 to 2.5 minute videos. Generally, the most interfering ads across mobile and desktop were ones that made users wait before they could watch the video (e.g., long, non-skippable

ads or ads with long skip delays) or appeared during the video. Additional analysis exploring the relationship between individual ad dimensions and the stack-rank confirmed perceived ad length, the timing of the ad's appearance, and wait time most explained variance in the stack-rank.

Results also reveal interesting insights about how obtrusive ad placements can negatively impact user perceptions. For example, users felt each mid-roll was longer than its equivalent pre-roll format and found the shorter 6 sec mid-roll less enjoyable to watch than the same length pre-roll - demonstrating how ill-timed placements can affect broader ad sentiments.

The approach described in this paper can be effectively used to study user perceptions of common or future instream video ad experiences. As with the Multi-Ad study, it can discriminate between good and bad ad experiences across many ads, in the context of a natural online activity (i.e., watching a video). Moreover, it can incorporate new ad experiences into the stack-rank in response to industry innovation, and can be adapted to study other environments such as longer videos, while maintaining the same core design.

Next Steps

There are plans to test additional countries, user contexts, and instream video ad experiences to inform the development of Better Ads Standards.

1. **Expand studies to select geographical regions**
   Study additional countries in key regions to determine whether video ad perceptions align well enough to justify creating a global instream video better ads standard. For pragmatic reasons, the intent is to study a small set of countries, since prior Coalition research indicates users' perceptions of common mobile and desktop web ad experiences in 18 countries align reasonably well.
2. **Test different short-form video lengths**
   Since perceptions of certain ad experiences may change in different length videos (perhaps if the ad length to video length ratio substantially changes), the study framework will be modified to evaluate longer video content of up to 10 min.
3. **Test more ad experiences on mobile and desktop**
   Experiences that are common in the video ads industry will be prioritized. Possible candidates include shorter or longer length video ads, or mid-rolls containing notification countdowns before the ad plays.
4. **Determine the initial instream video Better Ads Standards**
   The Coalition will use a similar approach as when adopting the desktop and mobile web Better Ads Standards (see Determining a Better Ads Standard Based on User Experience Data, Coalition for Better Ads, 2017). First, results from the experiments described in this paper will be used to identify a set of ad experiences that most interfere with users' ability to watch the main video. Then, a separate validation study will be run to see whether the most interfering ad experiences are more likely to result in unfavorable business outcomes, such as users' self-report likelihood of installing an ad blocker. If results from the instream video ad experiments and validation study align, there would be sufficient evidence to recommend an instream video ads standard.

References

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

Coalition for Better Ads (2017). *Determining a Better Ads Standard Based on User Experience Data. From betterads.org/research/standardpaper/.*

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 watching the video on this web page?
Your overall experience includes anything that may have enhanced or interfered with your ability to watch each video.

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

To what extent do you agree with the following statement?

“The web page did not surprise me with unexpected behaviors."

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

Ad Specific Questions

We showed a screenshot of the ad participants just saw, in the context of the main video, to refresh their memory of the ad.

We would like to ask you about an ad you may have noticed in Video A.

[Participant sees screenshot of the exposed ad experience]

How ANNOYING was the ad?

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

How ENJOYABLE was the ad to watch?*

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

**Participants did not see this question for overlay, pod, and popup ad experiences*

How RELEVANT was the ad to you?

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

To what extent does each of the following statements describe THE AD?

The ad was distracting.

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

The ad appeared at an unexpected point during the main video.

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

The ad was too long.*

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

**Participants did not see this question for the overlay and popup ad experiences.*

I had to wait a long time before I could get past the ad.

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

The ad was too personal.

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

Ad Ranking Exercise

In the last part of the study, compare how much each ad experience INTERFERED with your ability to watch the video that followed.

Refer to these reminders if you like.

Ad Experience A
[Participant sees screenshot of the exposed ad experience]

Ad Experience B
[Participant sees screenshot of the exposed ad experience]

Ad Experience C
[Participant sees screenshot of the exposed ad experience]

Which ad experience MOST interfered with your ability to watch the video?

- Ad A
- Ad B
- Ad C

Which ad experience LEAST interfered with your ability to watch the video?

- Ad A
- Ad B
- Ad C

Ad Experiences We Tested

Mobile Instream Video Experiences

Mobile Instream Video Experiences Tested in a Video Playing Site Context		
Name	**Description**	**Link**
Pre-roll 6s ad	6 second pre-roll ad that appears before the video plays. The ad is not skippable.	http://poetic-glass-136423.appspot.com/experience?exp=VIDEO_PREROLL_6SEC_UNSKIP&advertiser_id=1&article_id=121
Pre-roll 15s ad	15 second pre-roll ad that appears before the video plays. The ad is not skippable.	http://poetic-glass-136423.appspot.com/experience?exp=VIDEO_PREROLL_15SEC_UNSKIP&advertiser_id=1&article_id=121
Pre-roll 30s ad	30 second pre-roll ad that appears before the video plays. The ad is not skippable.	http://poetic-glass-136423.appspot.com/experience?exp=VIDEO_PREROLL_30SEC_UNSKIP&advertiser_id=1&article_id=121
Pre-roll 60s ad	60 second pre-roll ad that appears before the video plays. The ad is not skippable.	http://poetic-glass-136423.appspot.com/experience?exp=VIDEO_PREROLL_60SEC_UNSKIP&advertiser_id=1&article_id=121
Pre-roll 15s ad skippable after 5s	15 second pre-roll ad that appears before the video plays, and is skippable after 5s	http://poetic-glass-136423.appspot.com/experience?exp=VIDEO_PREROLL_15SEC_5SEC&advertiser_id=1&article_id=121
Pre-roll 30s ad skippable after 5s	30 second pre-roll ad that appears before the video plays, and is skippable after 5s	http://poetic-glass-136423.appspot.com/experience?exp=VIDEO_PREROLL_30SEC_5SEC&advertiser_id=1&article_id=121
Pre-roll 60s ad skippable after 5s	60 second pre-roll ad that appears before the video plays, and is skippable after 5s	http://poetic-glass-136423.appspot.com/experience?exp=VIDEO_PREROLL_60SEC_5SEC&advertiser_id=1&article_id=121
Pre-roll 30s ad skippable after 10s	30 second pre-roll ad that appears before the video plays, and is skippable after 10 sec	http://poetic-glass-136423.appspot.com/experience?exp=VIDEO_PREROLL_30SEC_10SEC&advertiser_id=1&article_id=121
Pre-roll 30s ad skippable after 15s	30 second pre-roll ad that appears before the video plays, and is skippable after 15 sec	http://poetic-glass-136423.appspot.com/experience?exp=VIDEO_PREROLL_30SEC_15SEC&advertiser_id=1&article_id=121

Pre-roll 60s ad skippable after 10s	60 second pre-roll ad that appears before the video plays, and is skippable after 10 sec	http://poetic-glass-136423.appspot.com/experience?exp=VIDEO_PREROLL_60SEC_10SEC&advertiser_id=1&article_id=121
Pre-roll 60s ad skippable after 15s	60 second pre-roll ad that appears before the video, and is skippable after 15 sec	http://poetic-glass-136423.appspot.com/experience?exp=VIDEO_PREROLL_60SEC_15SEC&advertiser_id=1&article_id=121
Mid-roll 6s ad	6 second mid-roll that appears during the video	http://poetic-glass-136423.appspot.com/experience?exp=VIDEO_MIDROLL_6SEC_UNSKIP&advertiser_id=1&article_id=121
Mid-roll 30s ad skippable after 5s	30 second mid-roll ad that appears during the video, and is skippable after 5 sec	http://poetic-glass-136423.appspot.com/experience?exp=VIDEO_MIDROLL_30SEC_5SEC&advertiser_id=1&article_id=121
Mid-roll 60s ad skippable after 5s	60 second mid-roll ad that appears during the video, and is skippable after 5 sec	http://poetic-glass-136423.appspot.com/experience?exp=VIDEO_MIDROLL_60SEC_5SEC&advertiser_id=1&article_id=121
Pre-roll Popup ad	A popup (responsively sized to be 50% and 85% of the device's width and height, respectively) appears in the center of the video player after the page loads. There is a clearly visible dismiss button. The video starts after the popup is dismissed.	http://poetic-glass-136423.appspot.com/experience?exp=VIDEO_POPUP&advertiser_id=1&article_id=121
Pre-roll pod of 2x15s ads	Short ads combined into a single "pod" that play consecutively (2 x 15s non-skippable ads)	http://poetic-glass-136423.appspot.com/experience?exp=VIDEO_POD_15SEC_UNSKIP&advertiser_id=1&article_id=121

Desktop Instream Video Experiences

Desktop Instream Video Experiences Tested in a Video Playing Site Context		
Name	**Description**	**Link**
Pre-roll 6s ad	6 second pre-roll ad that appears before the video plays. The ad is not skippable.	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_VIDEO_PREROLL_6SEC_UNSKIP&advertiser_id=1&article_id=121

Pre-roll 15s ad	15 second pre-roll ad that appears before the video plays. The ad is not skippable.	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_VIDEO_PREROLL_15SEC_UNSKIP&advertiser_id=1&article_id=121
Pre-roll 30s ad	30 second pre-roll ad that appears before the video plays. The ad is not skippable.	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_VIDEO_PREROLL_30SEC_UNSKIP&advertiser_id=1&article_id=121
Pre-roll 60s ad	60 second pre-roll ad that appears before the video plays. The ad is not skippable.	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_VIDEO_PREROLL_60SEC_UNSKIP&advertiser_id=1&article_id=121
Pre-roll 15s ad skippable after 5s	15 second pre-roll ad that appears before the video plays, and is skippable after 5s	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_VIDEO_PREROLL_15SEC_5SEC&advertiser_id=1&article_id=121
Pre-roll 30s ad skippable after 5s	30 second pre-roll ad that appears before the video plays, and is skippable after 5s	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_VIDEO_PREROLL_30SEC_5SEC&advertiser_id=1&article_id=121
Pre-roll 60s ad skippable after 5s	60 second pre-roll ad that appears before the video plays, and is skippable after 5s	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_VIDEO_PREROLL_60SEC_5SEC&advertiser_id=1&article_id=121
Pre-roll 30s ad skippable after 10s	30 second pre-roll ad that appears before the video plays, and is skippable after 10 sec	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_VIDEO_PREROLL_30SEC_10SEC&advertiser_id=1&article_id=121
Pre-roll 30s ad skippable after 15s	30 second pre-roll ad that appears before the video plays, and is skippable after 15 sec	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_VIDEO_PREROLL_30SEC_15SEC&advertiser_id=1&article_id=121
Pre-roll 60s ad skippable after 10s	60 second pre-roll ad that appears before the video plays, and is skippable after 10 sec	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_VIDEO_PREROLL_60SEC_10SEC&advertiser_id=1&article_id=121

Pre-roll 60s ad skippable after 15s	60 second pre-roll ad that appears before the video, and is skippable after 15 sec	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_VIDEO_PREROLL_60SEC_15SEC&advertiser_id=1&article_id=121
Mid-roll 6s ad	6 second mid-roll that appears during the video	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_VIDEO_MIDROLL_6SEC_UNSKIP&advertiser_id=1&article_id=121
Mid-roll 30s ad skippable after 5s	30 second mid-roll ad that appears during the video, and is skippable after 5 sec	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_VIDEO_MIDROLL_30SEC_5SEC&advertiser_id=1&article_id=121
Mid-roll 60s ad skippable after 5s	60 second mid-roll ad that appears during the video, and is skippable after 5 sec	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_VIDEO_MIDROLL_60SEC_5SEC&advertiser_id=1&article_id=121
Pre-roll image ad	Before the video plays, an overlay ad is shown in the middle of the video window with a skip button. The Skip button must be pressed to continue to the video. The ad is 728x90.	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_VIDEO_OVERLAY_START&advertiser_id=1&article_id=121
Image ad that covers 20% of the playing video	While the video plays, a static overlay 728x90 image ad shows in the bottom 1/3 of the video frame with a clear dismiss button in the top right. The ad width is 85% of the player width. The ad is shown at 0:10 and disappears at 0:40.	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_VIDEO_OVERLAY&advertiser_id=1&article_id=121
Image ad that covers 50% of the playing video	While the video plays, an overlay ad shows on top of the video with a clear dismiss button in the top right (covers 50% of video and is placed in the bottom part of the video player).	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_VIDEO_OVERLAY_50&advertiser_id=1&article_id=121
Image ad at end of video	After the video plays, a static image 728x90 image ad is displayed in the video window with a black background. The ad is not dismissable.	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_VIDEO_OVERLAY_END&advertiser_id=1&article_id=121

Pre-roll Popup ad	A popup (responsively sized to be 40% and 70% of the device's width and height, respectively) appears in the center of the video player after the page loads. There is a clearly visible dismiss button. The video starts after the popup is dismissed.	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_VIDEO_POPUP&advertiser_id=1&article_id=121
Pre-roll pod of 2x15s ads	Short ads combined into a single "pod" that play consecutively (2 x 15s non-skippable ads)	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_VIDEO_POD_15SEC_UNSKIP&advertiser_id=1&article_id=121
Small image ad on top of playing video	Overlay which is smaller in size (i.e., 468x60).	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_VIDEO_OVERLAY_SMALL&advertiser_id=1&article_id=121
Small image ad on top of playing video for 15s	Overlay which is smaller in size (i.e., 468x60px), and disappears after a certain period of time (i.e., 15 sec)	http://poetic-glass-136423.appspot.com/experience?exp=DESKTOP_VIDEO_OVERLAY_SMALL_15SEC&advertiser_id=1&article_id=121

Estimating the Rank Score Using the Bradley-Terry Algorithm

The data that was collected are all individual-level ranks of three ad combinations, out of about 20 ad experiences (16 in mobile, 24 in desktop). The total number of ad experiences is not directly relevant to the analysis. The participants' rankings of the three ads were aggregated into a set of paired comparisons. The Bradley-Terry model was then applied to this data and estimated a latent ability parameter for each ad, where ability means how much the ad is perceived to interfere with participants' video-watching experience.

Bradley-Terry Model

In this application, ads are in a "contest" judged by a participant. It is assumed 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' perceived video-watching interference (larger values of α mean less interference). 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 $\{\alpha_i\}$:

$$L(\alpha) = \sum_{i}^{n} \sum_{j}^{n} \{w_{ij} log\alpha_i - w_{ij} log(\alpha_i + \alpha_j)\}$$

We fit the model and obtain parameter estimates using the BradleyTerry2 package in R.

Rescaling

Surveys were run in multiple stages to be efficient with a limited sample size. In order to create consistent and meaningful metrics that accurately represent the relative *interference* of each ad experience, the estimates of perceived *interference* $\{\lambda_i\}$were rescored to an interval between 1 and 5, where 1 is the MOST interfering score and 5 is the LEAST interfering score. This rescaling was used to ensure interpretable results and consistency across studies.

PLS Regression - Rank Score on UX Metrics

In order to check that the rank score is a valid single metric for measuring ad preferences, its relationship with the collected UX metrics is modeled. The goal is to understand how much of the information contained in the UX metrics is captured by the rank score and how much each UX metric contributes to it. This would confirm that the rank score is a good one-dimensional measure of ad preferences.

However, the pairwise correlation between some metrics is very high. For example, annoyance and distraction have a correlation of 0.74 (Figure 19). The multicollinearity between the UX metrics means that directly fitting a multiple regression with the UX metrics as predictors will be problematic. Therefore, the Partial Least Squares (PLS) regression (Abdi 2010) was used, which combines the ideas of Principal Component Analysis (PCA) and multiple linear regression with the intention of dimension reduction. PLS finds components (i.e., latent vectors) that explain the covariance between predictors and response. Each of the components is a linear combination of the 10 UX metrics (see Figure 20 below).

(Note: the rank score is measured at the ad level, while the UX metrics are measured at the participant level. The mean of each UX metric rating (on a 1-5 scale) for each ad is calculated and these means are used as the explanatory variables in PLS.)

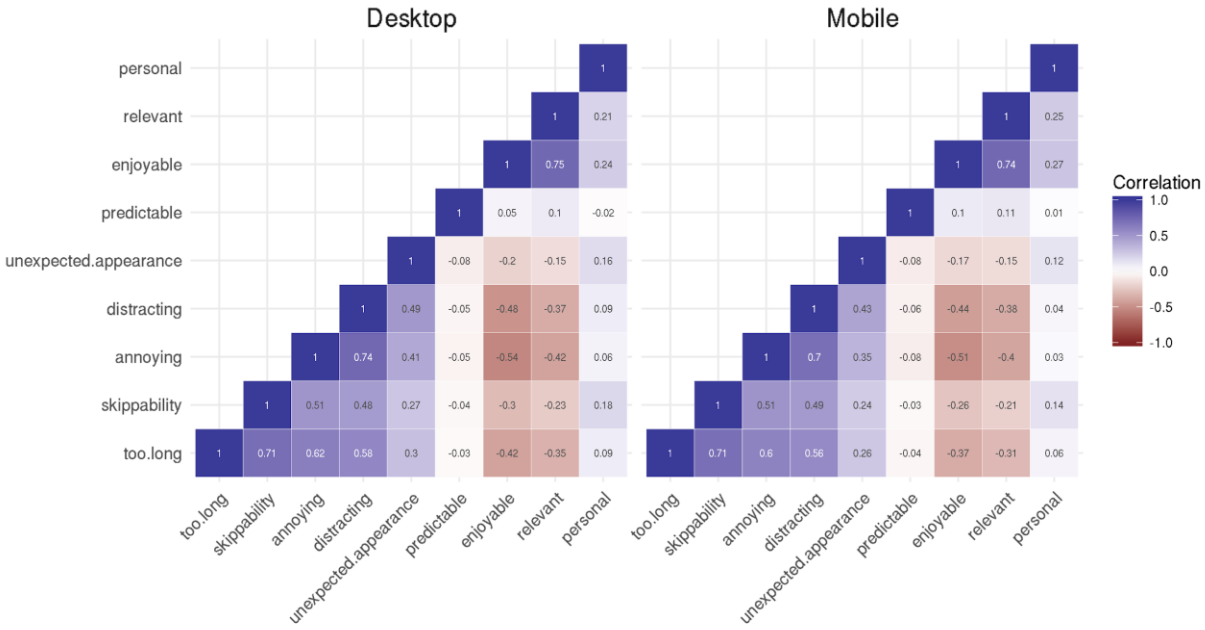


Figure 28: 10 UX metrics - pairwise correlation matrix plot

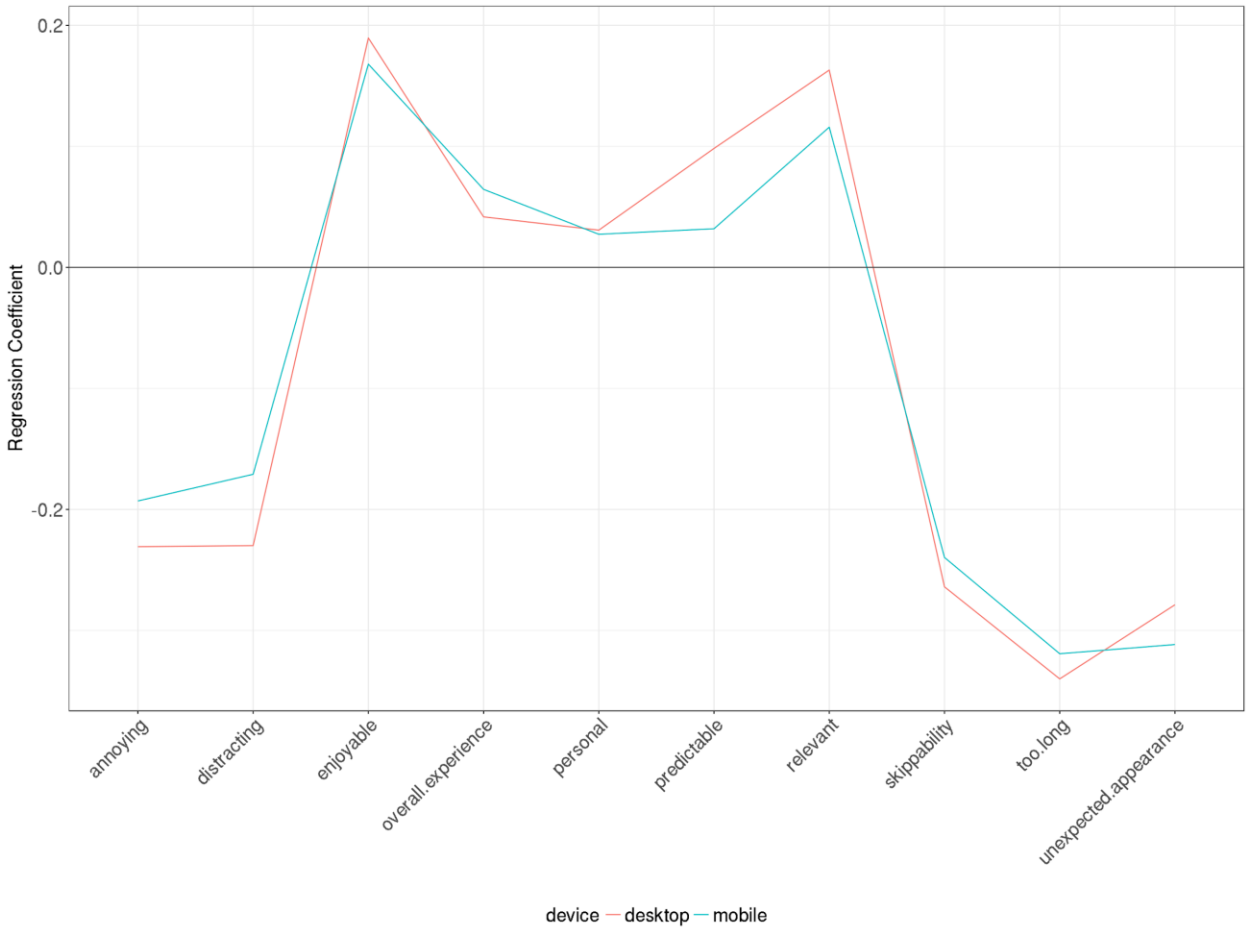
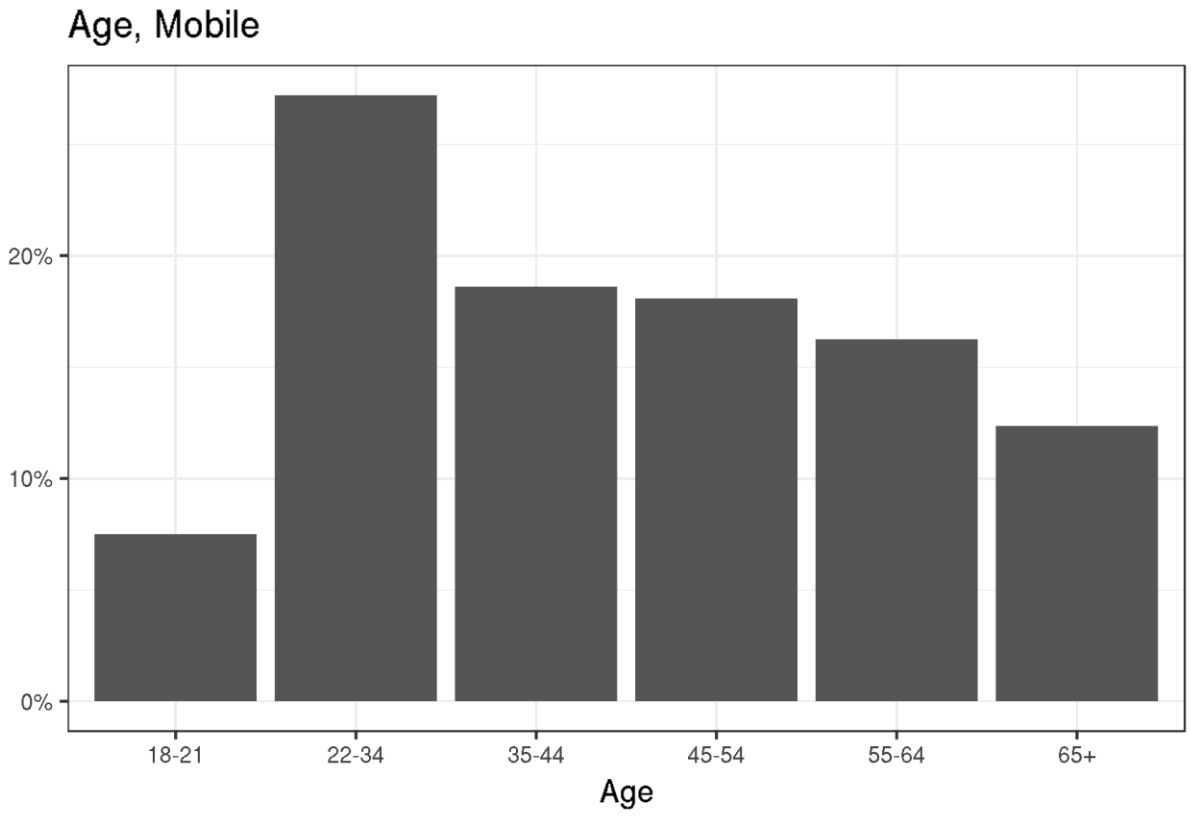
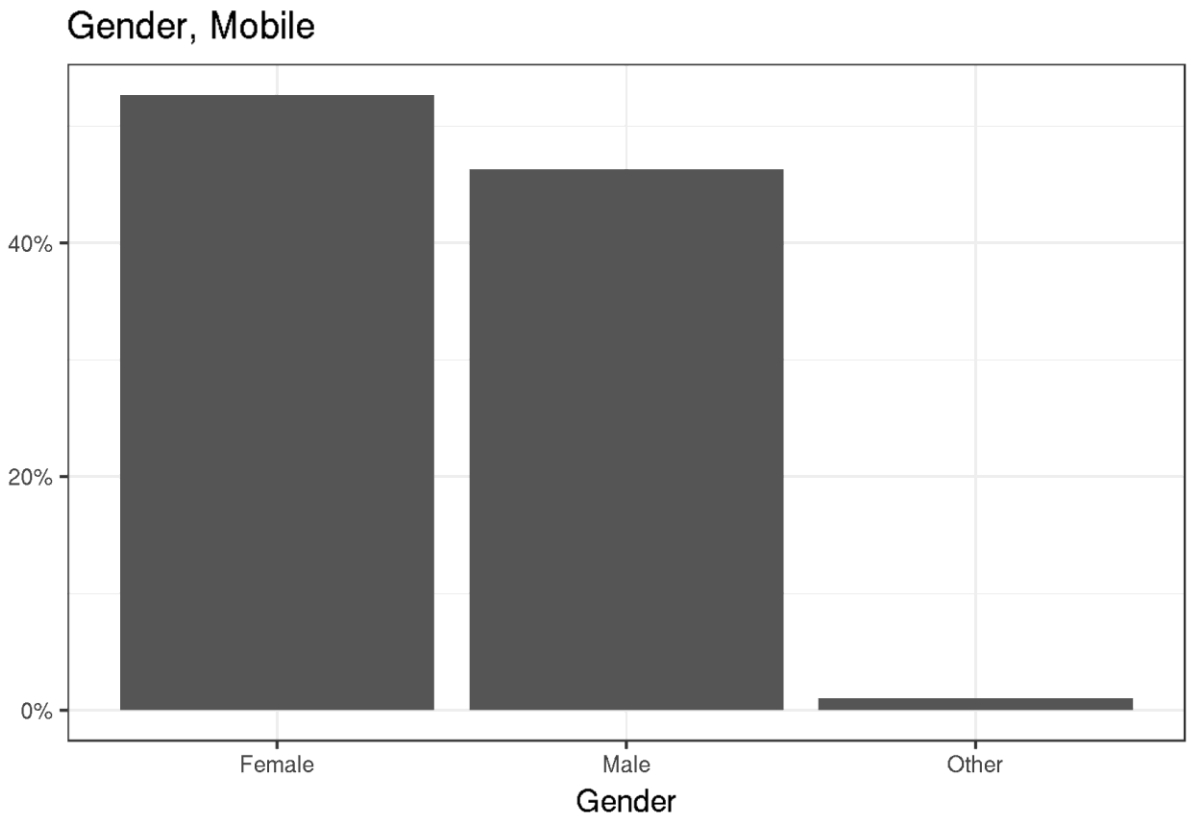


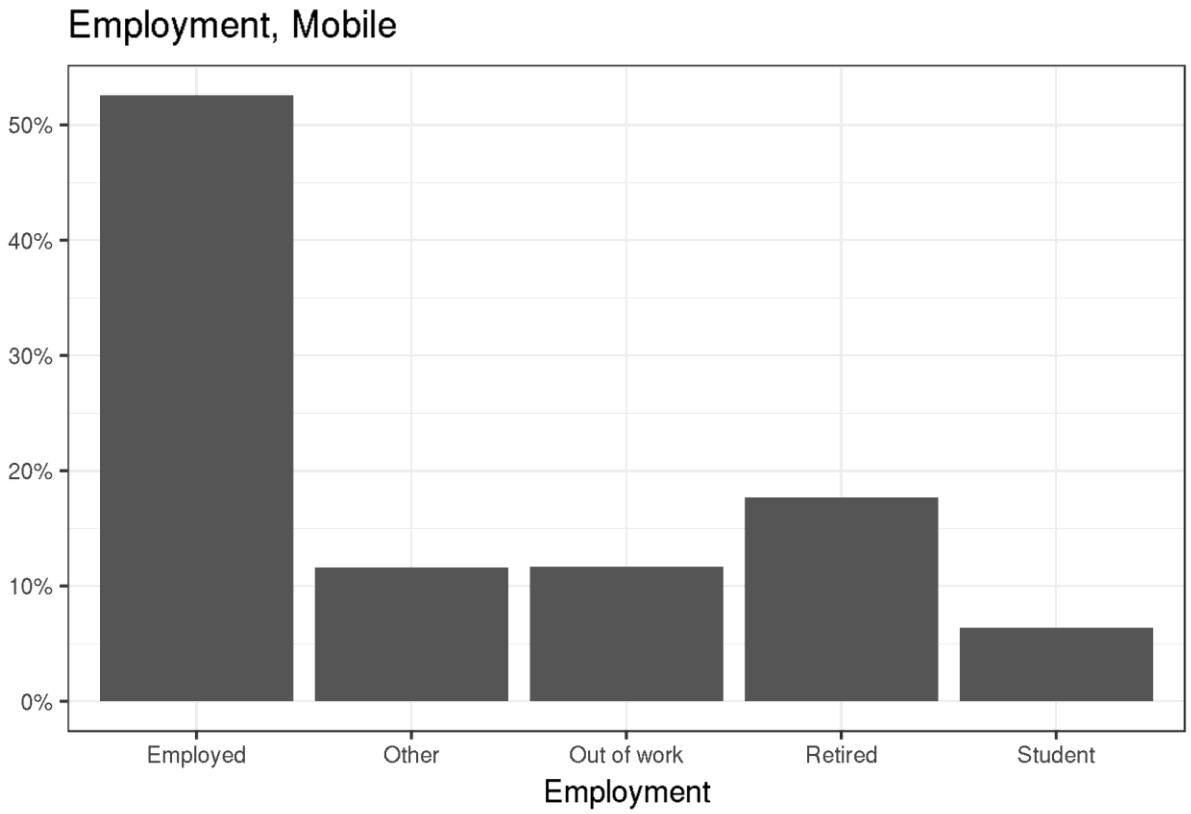
Figure 29: Weights for UX metric's contribution to the rank score (the weights are the estimated regression coefficients from fitting PLS regression with the rank scores and UX metrics).

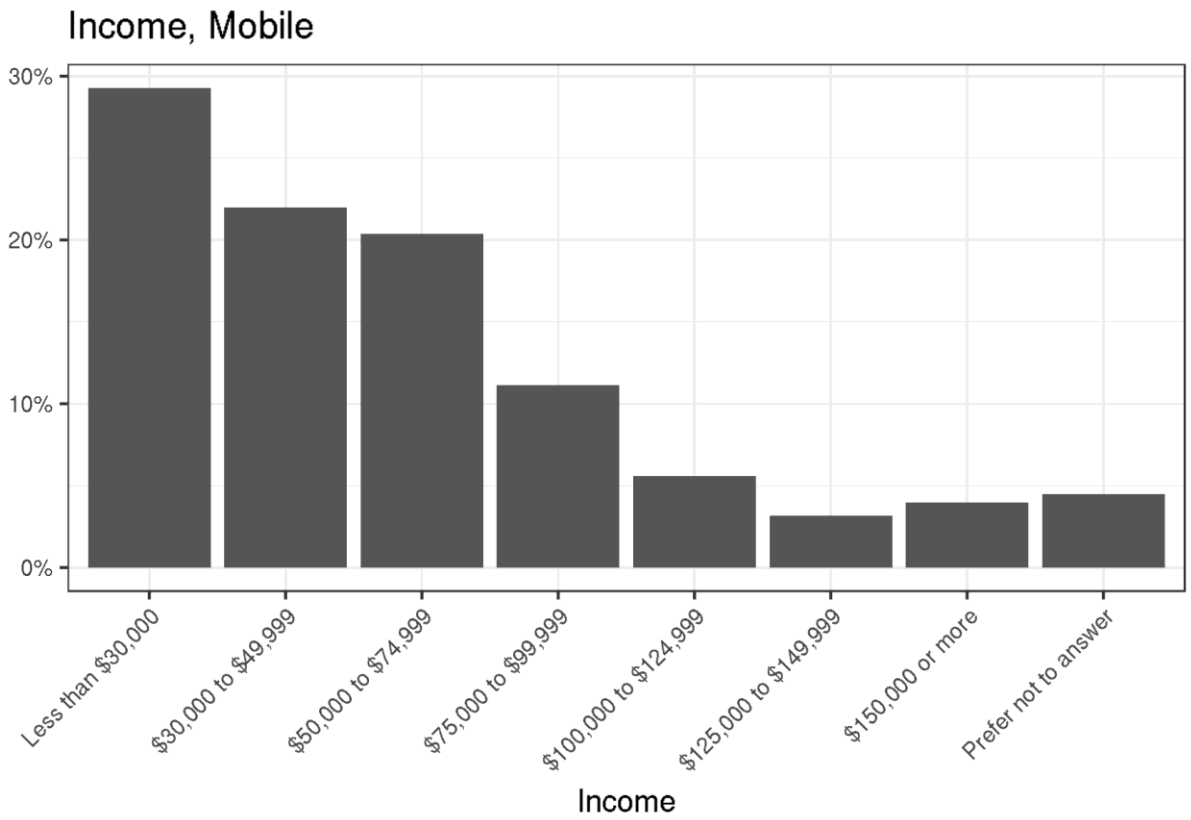
The figure above shows that ad length ("too.long"; whether the participant felt the ad was too long), unexpected appearance, and wait time ("skippability"; whether the participant felt they had to wait a long time before they could skip the ad) are three main negative contributors for both desktop and mobile, followed closely by annoyance and distraction. The latter two were the largest contributors to the rank score in the Multi-Ad Study. Overall experience with the webpage, how personal the ad seemed, and how predictable the web page contributed little. Enjoyability and relevance contributed positively to the rank score, as we would expect. This result confirms that the rank score correlates well with the UX metrics and is a good single metric to rely on.

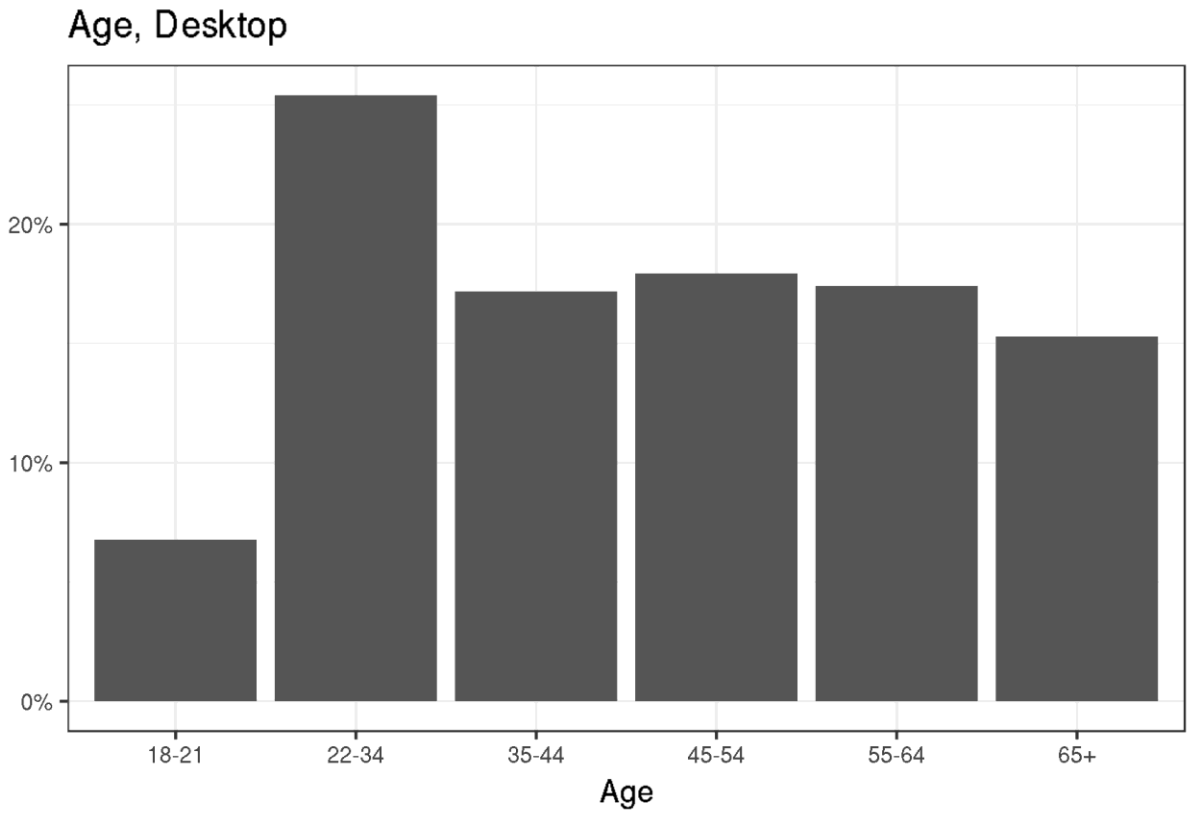
Demographic Distribution

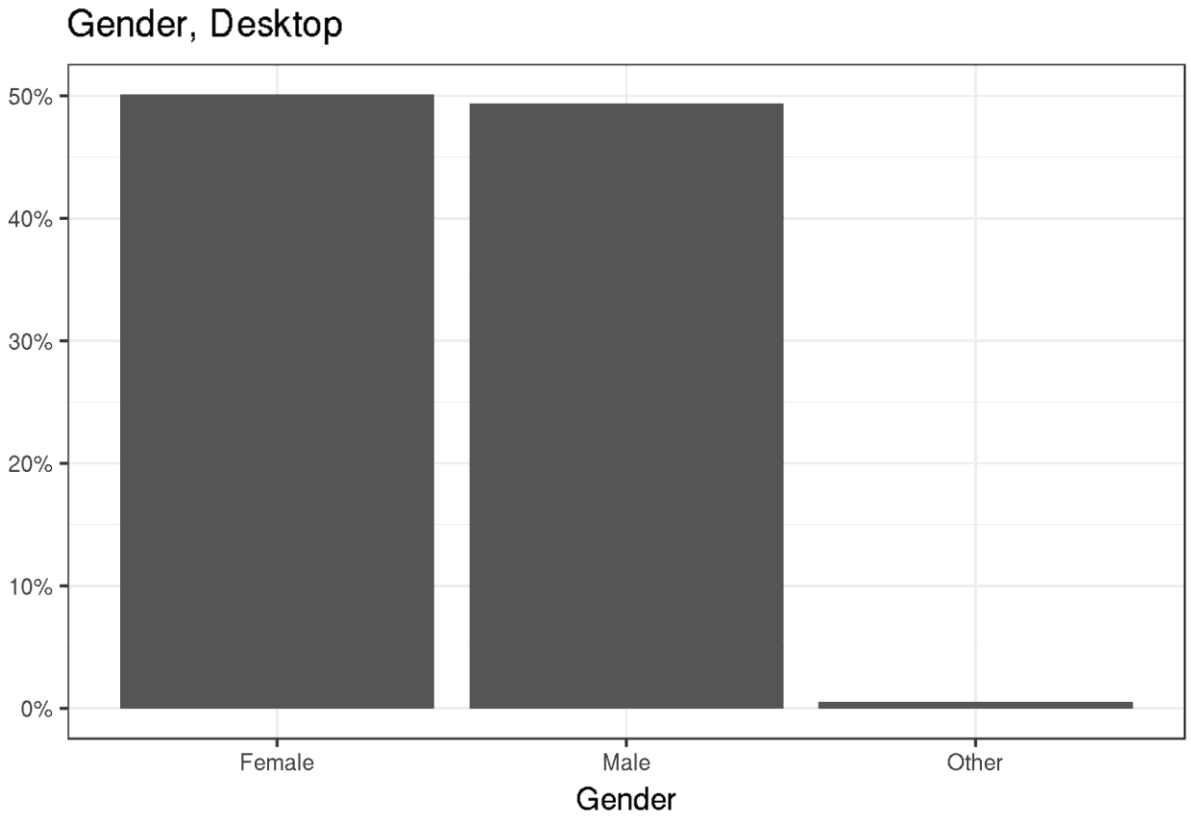

Age, Mobile
20%
10%
0%
18-21
22-34
35-44
45-54
55-64
65+
Age

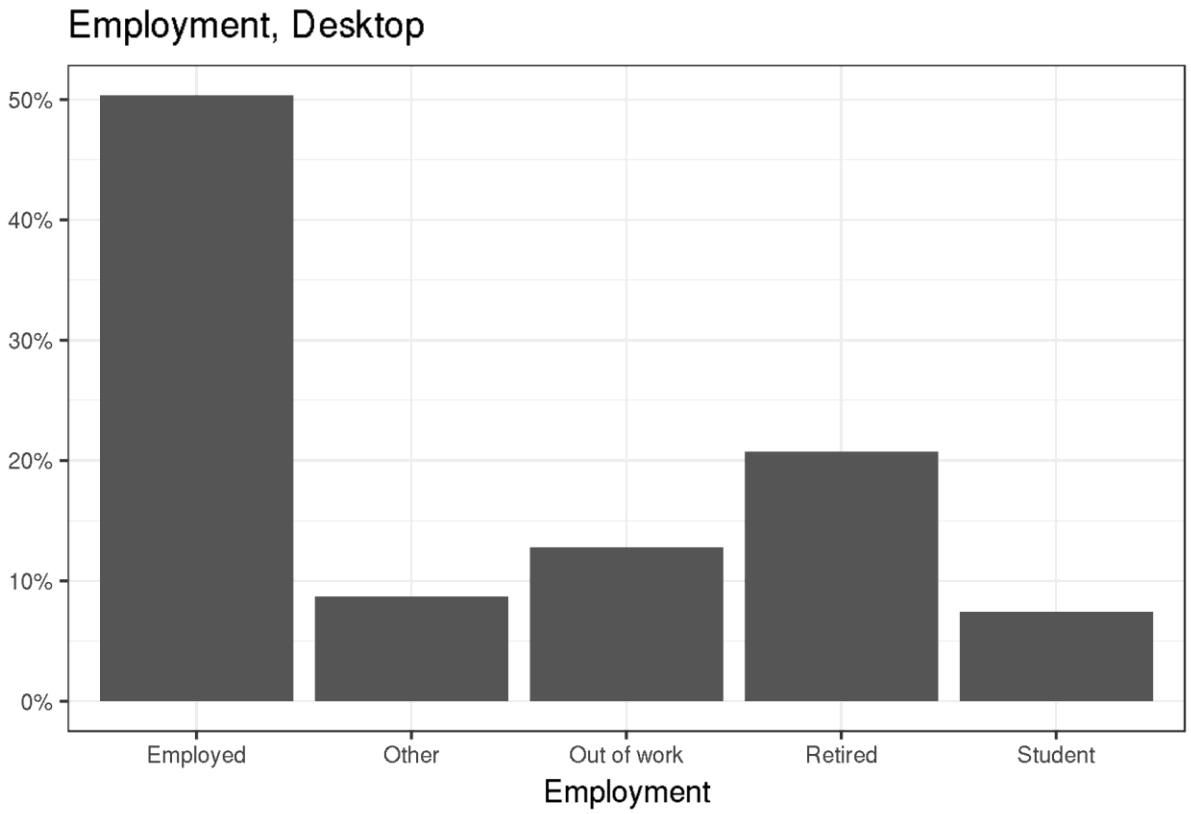

Gender, Mobile
40%
20%
0%
Female
Male
Other
Gender

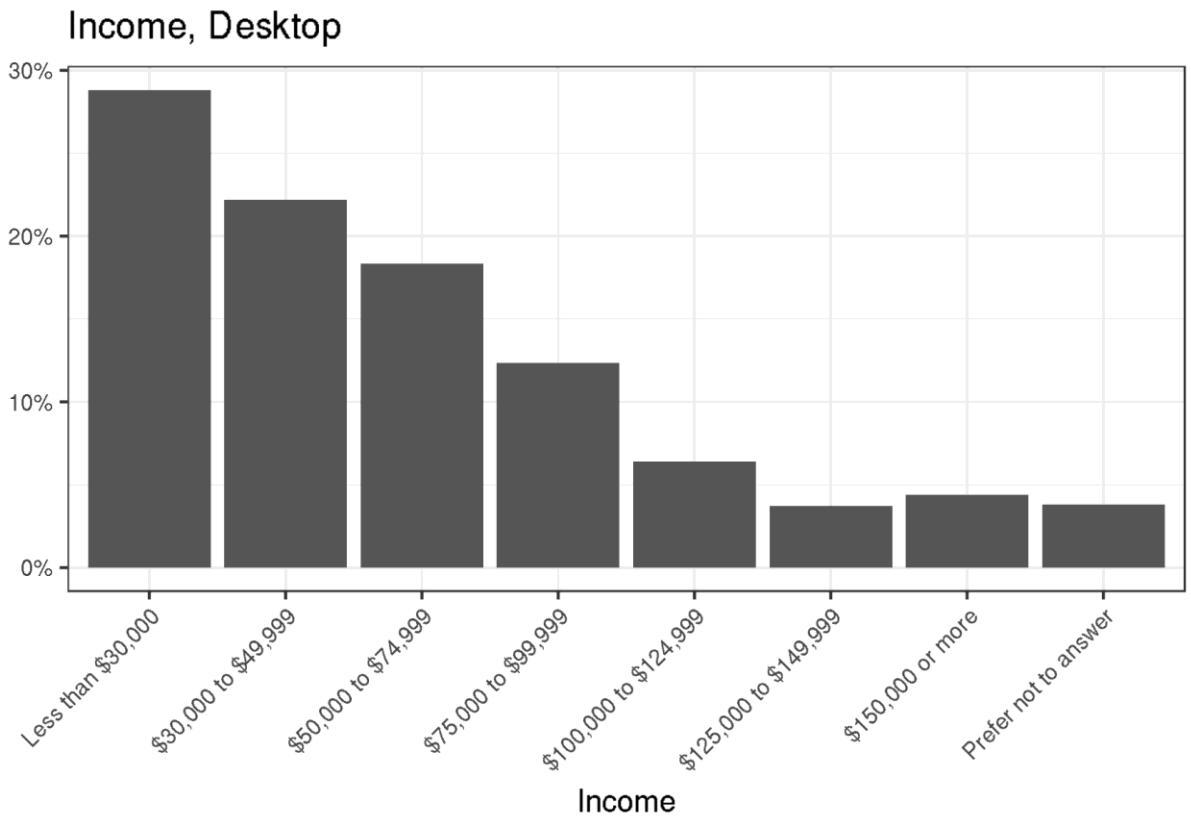

Employment, Mobile
50%
40%
30%
20%
10%
0%
Employed
Other
Out of work
Retired
Student
Employment


Income, Mobile
30%
20%
10%
0%
Less than $30,000
$30,000 to $49,999
$50,000 to $74,999
$75,000 to $99,999
$100,000 to $124,999
$125,000 to $149,999
$150,000 or more
Prefer not to answer
Income

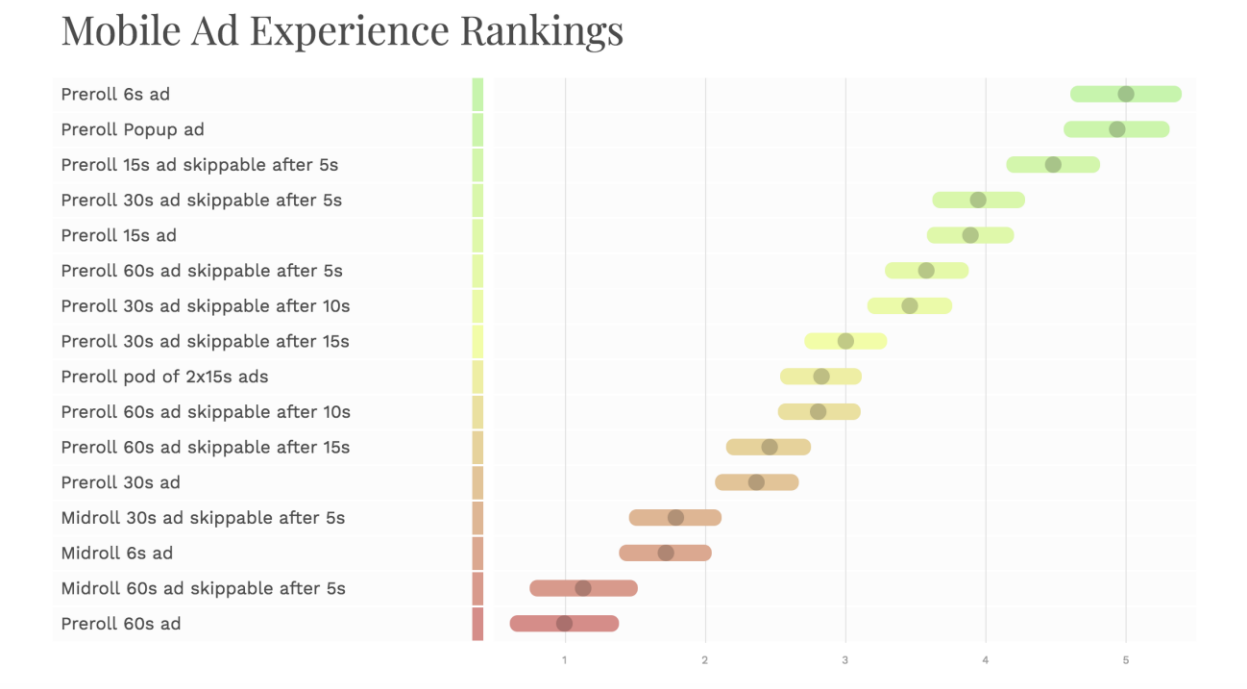

Age, Desktop
20%
10%
0%
18-21
22-34
35-44
45-54
55-64
65+
Age

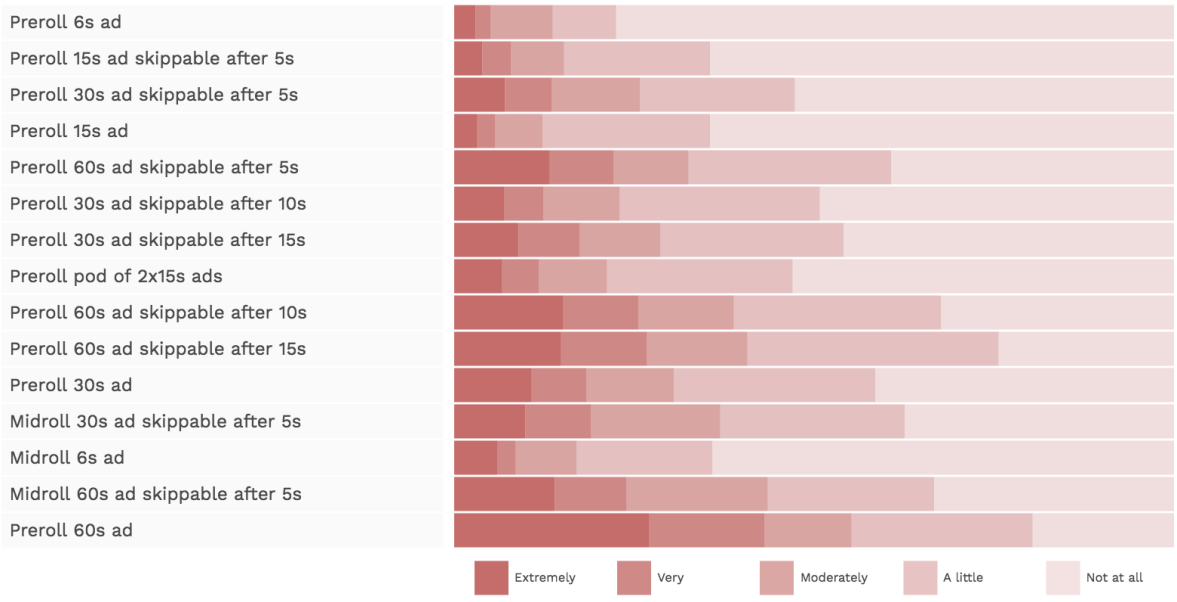

Gender, Desktop
50%
40%
30%
20%
10%
0%
Female
Male
Other
Gender


Employment, Desktop
50%
40%
30%
20%
10%
0%
Employed
Other
Out of work
Retired
Student
Employment

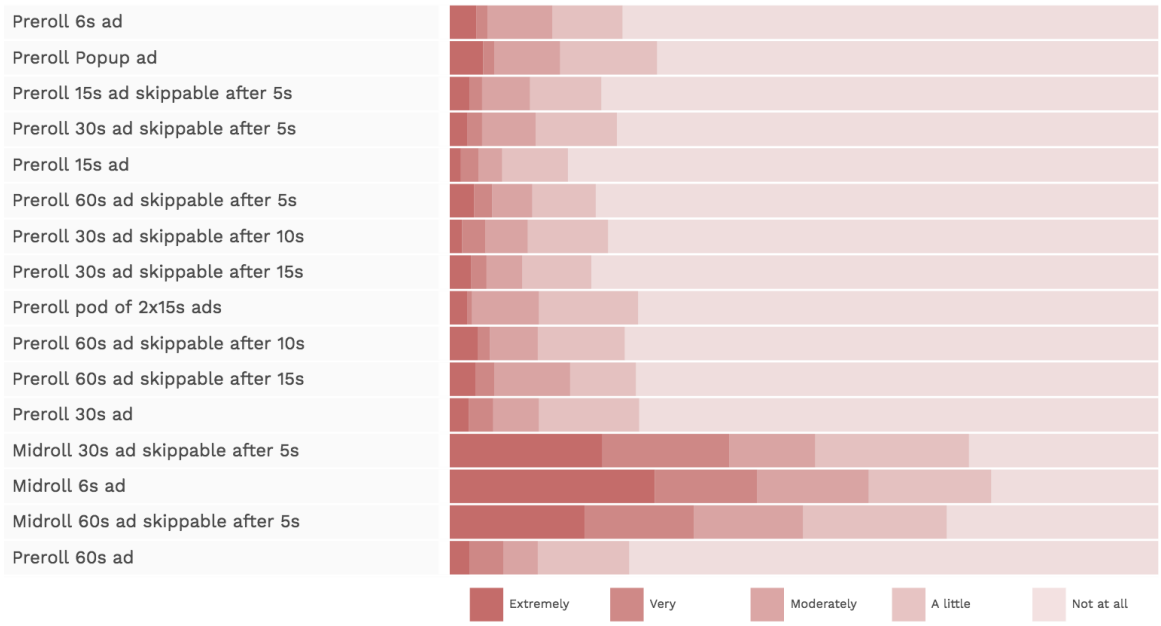

Income, Desktop
30%
20%
10%
0%
Less than $30,000
$30,000 to $49,999
$50,000 to $74,999
$75,000 to $99,999
$100,000 to $124,999
$125,000 to $149,999
$150,000 or more
Prefer not to answer
Income

Full Results

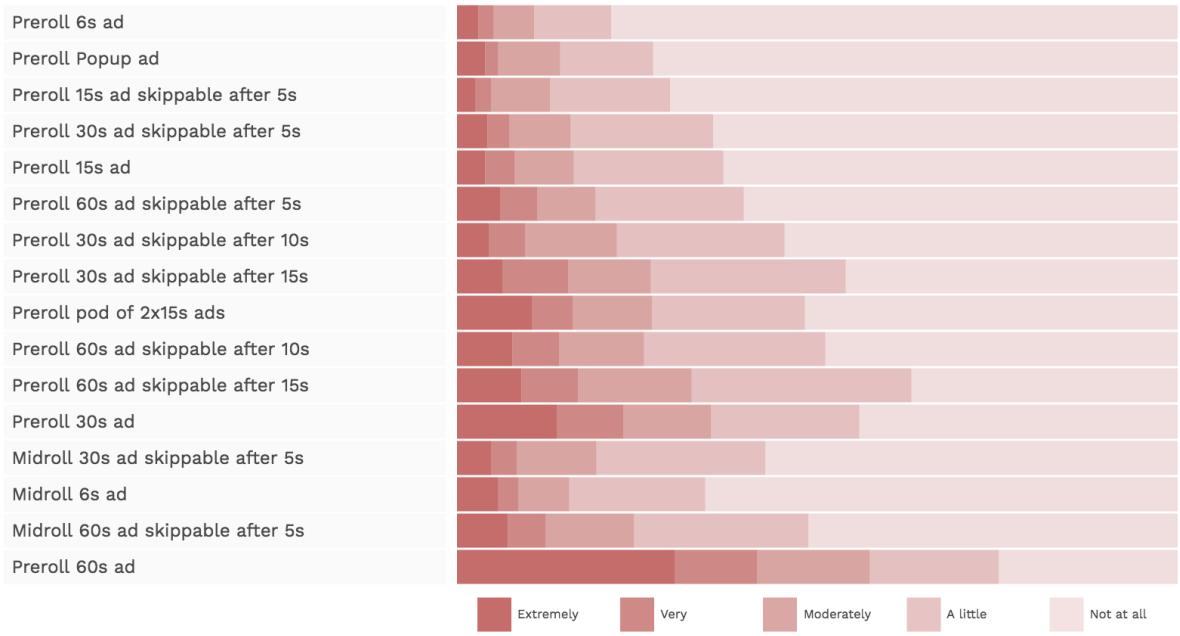




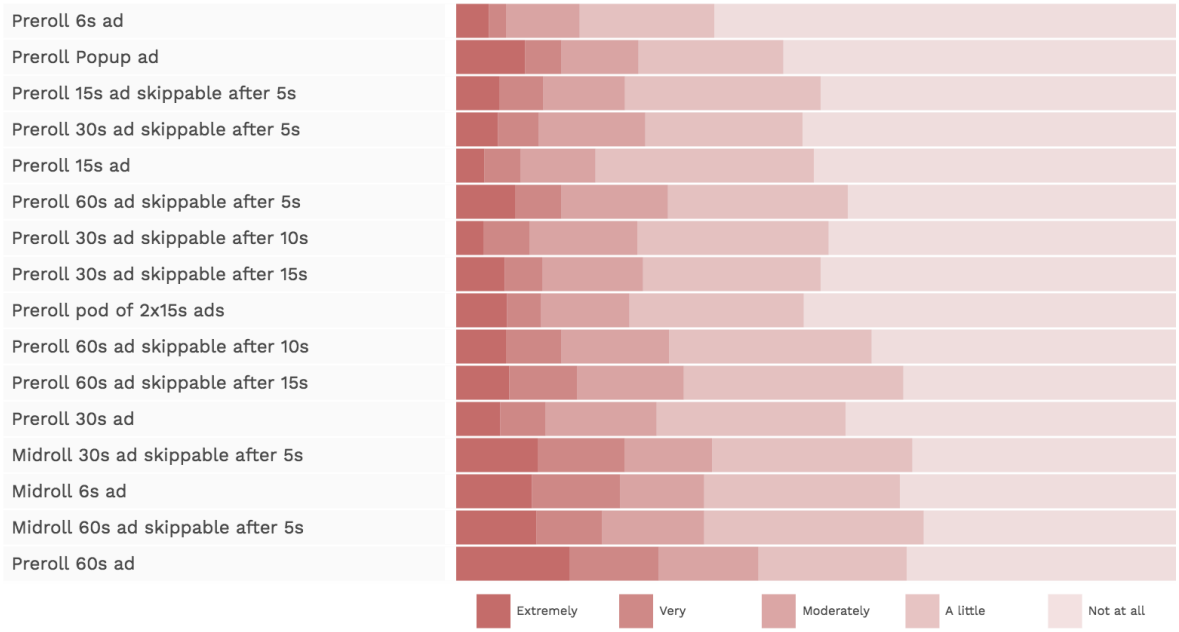
MOBILE - Unexpected Appearance



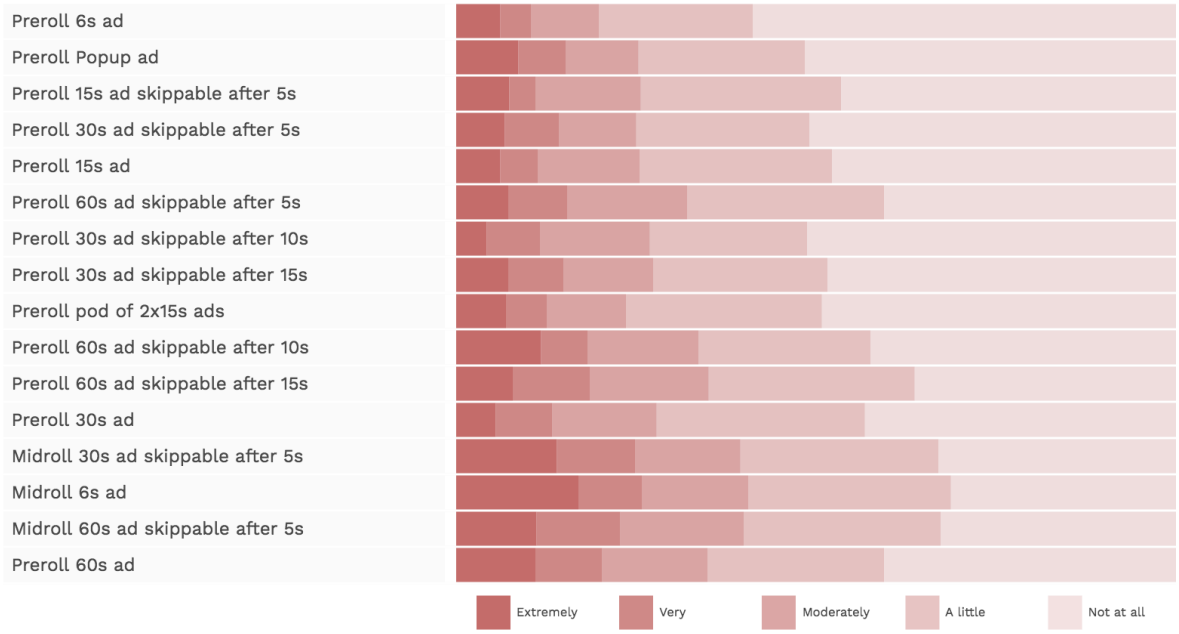
MOBILE - Wait Time



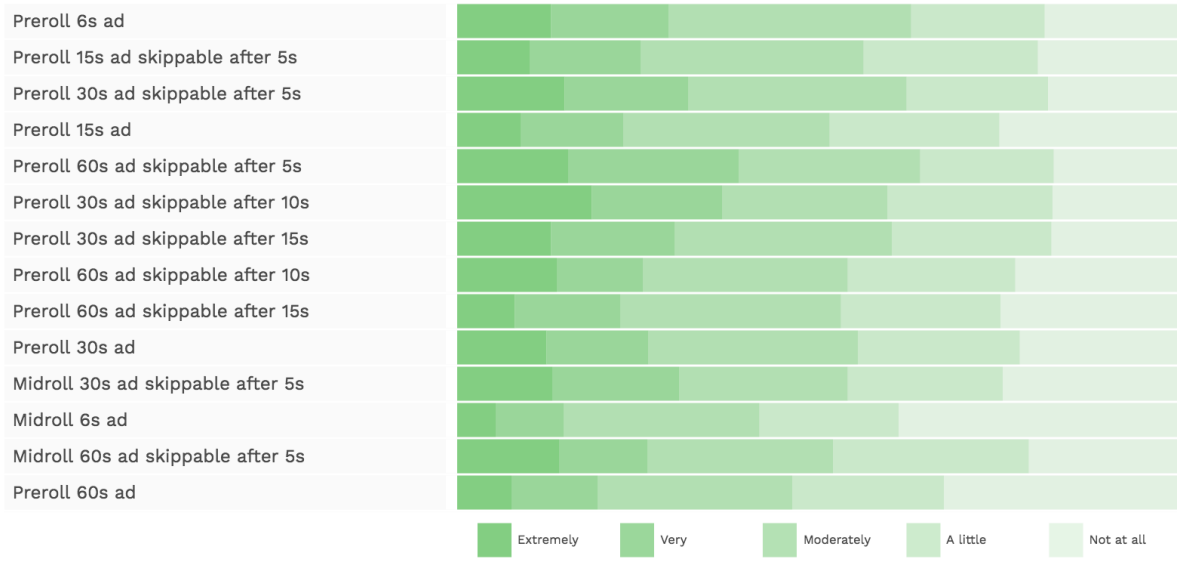
MOBILE - Annoying



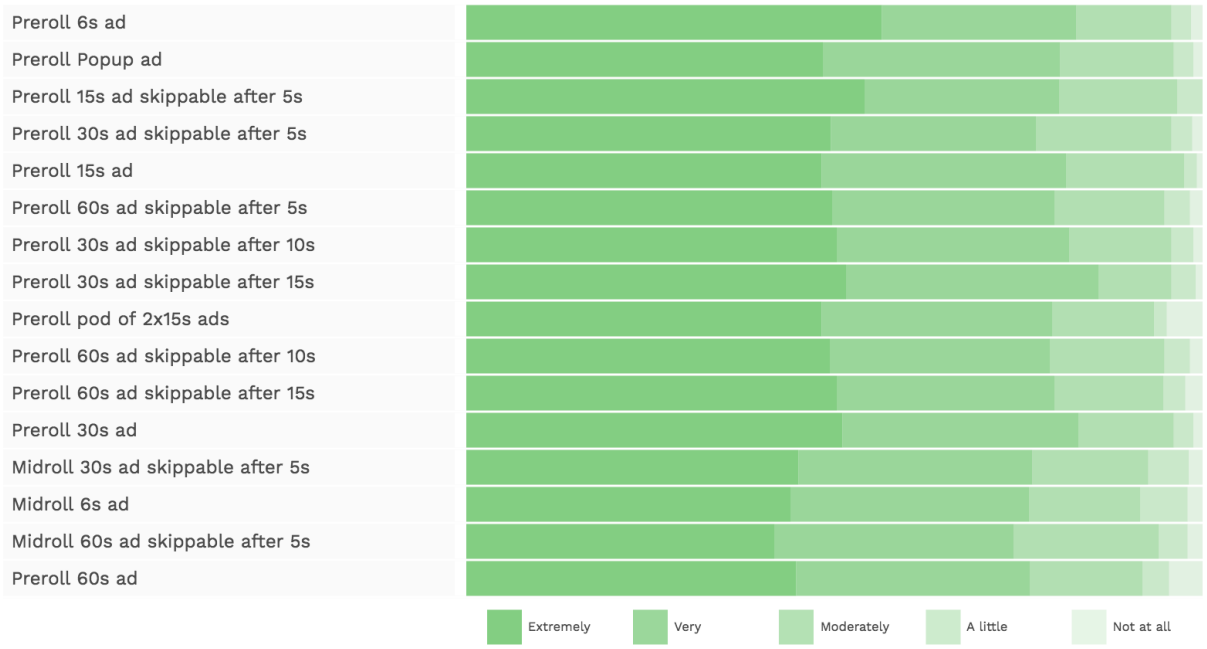
MOBILE - Distracting



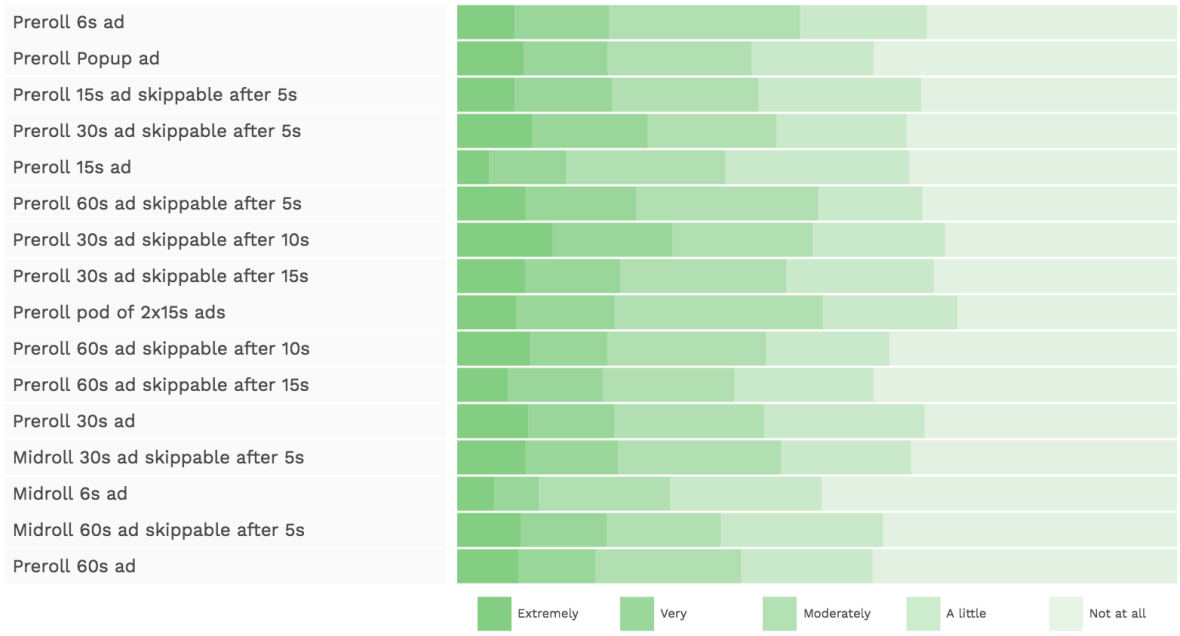
MOBILE - Enjoyable (Ad Creative)



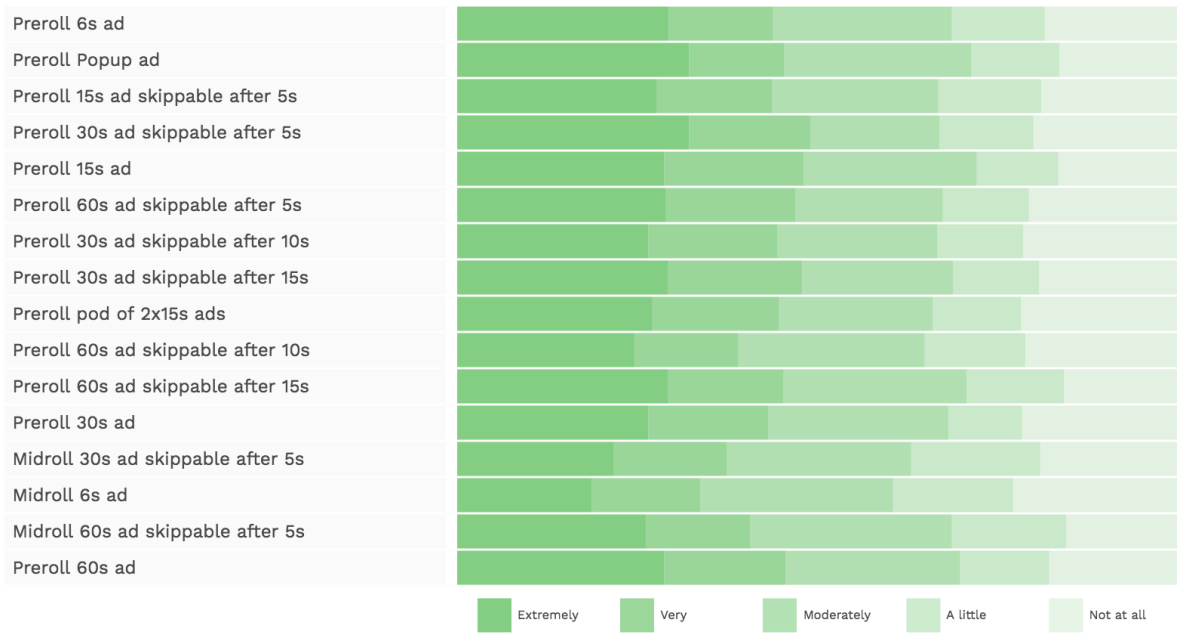
MOBILE - Overall Satisfaction Watching Main Video



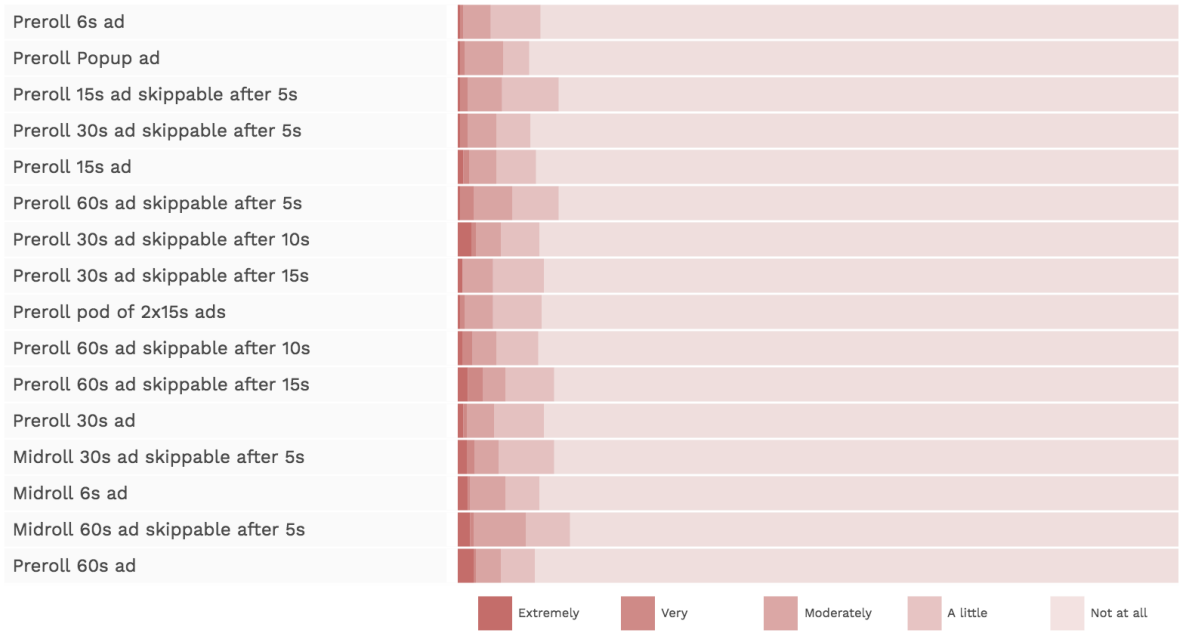
MOBILE - Ad's Relevance to User



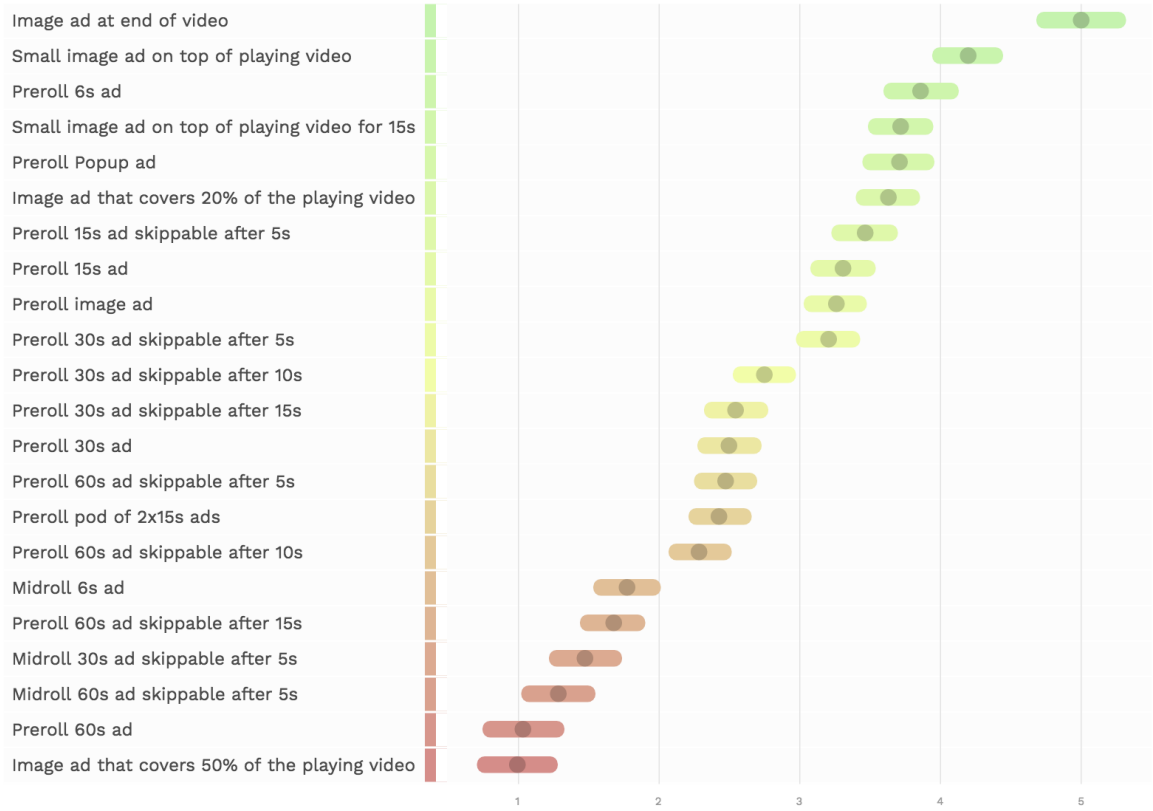
MOBILE - Overall Webpage Behavior Predictability



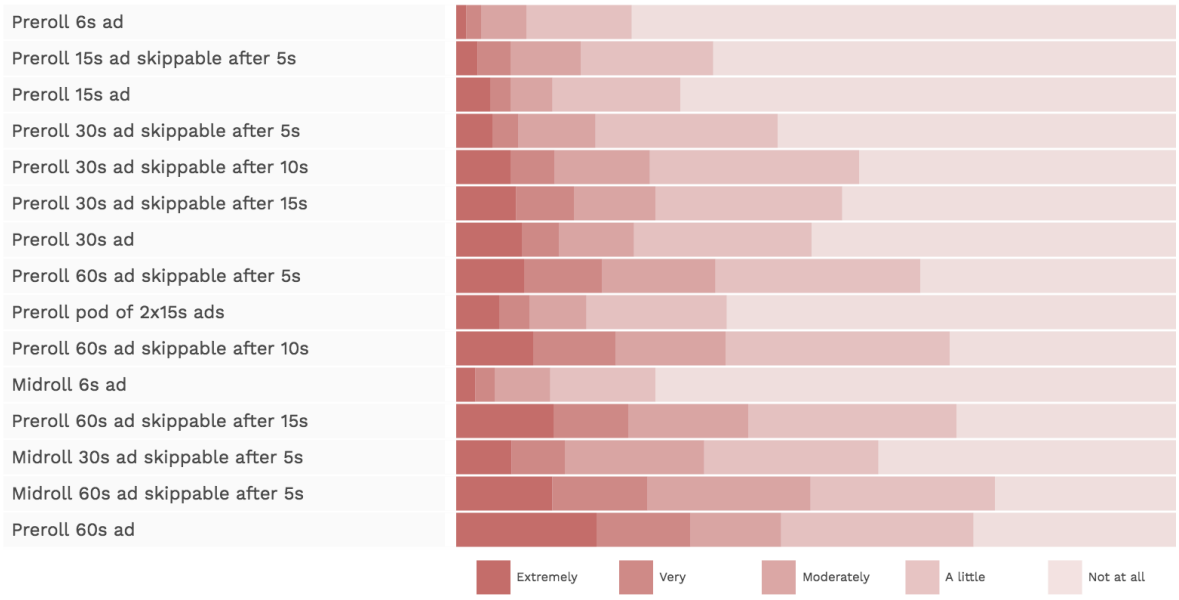
MOBILE - Too Personal (Ad Creative)



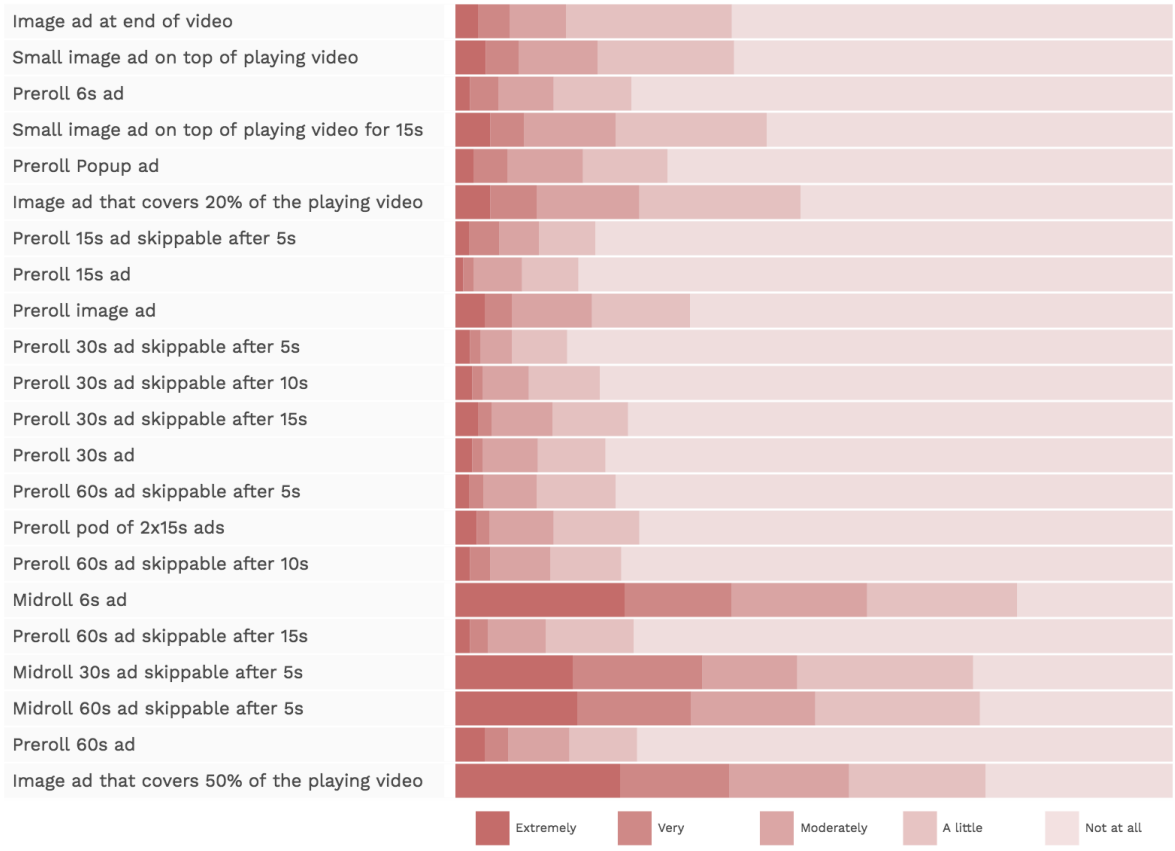
Desktop Ad Experience Rankings



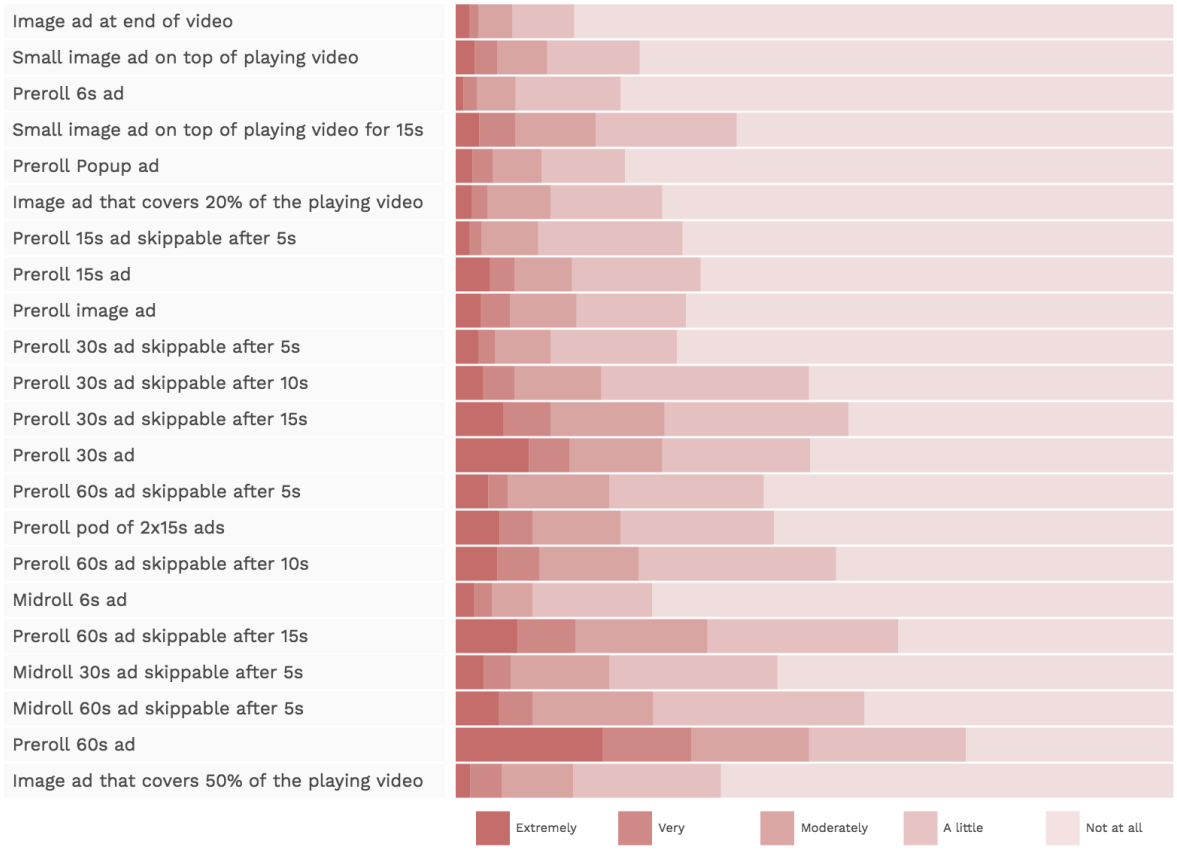
DESKTOP - Ad Length (Too Long)



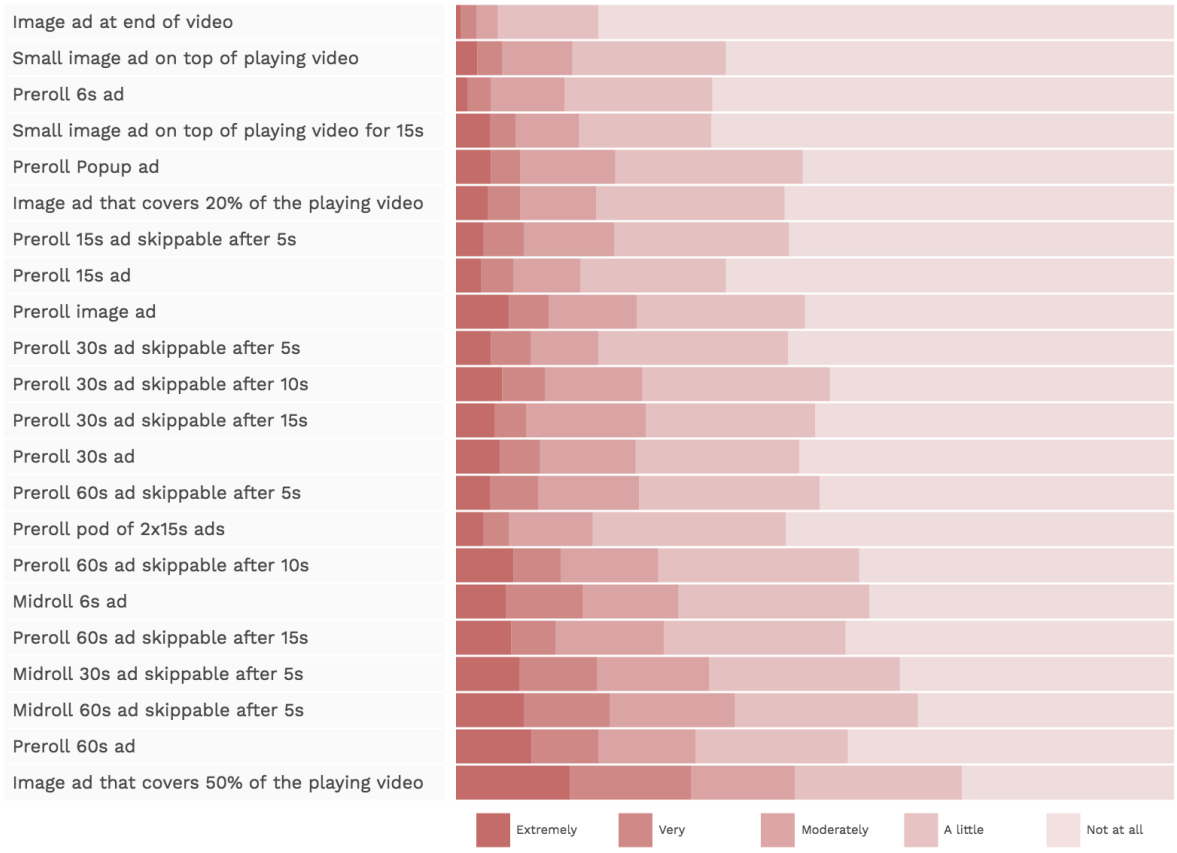
DESKTOP - Unexpected Appearance



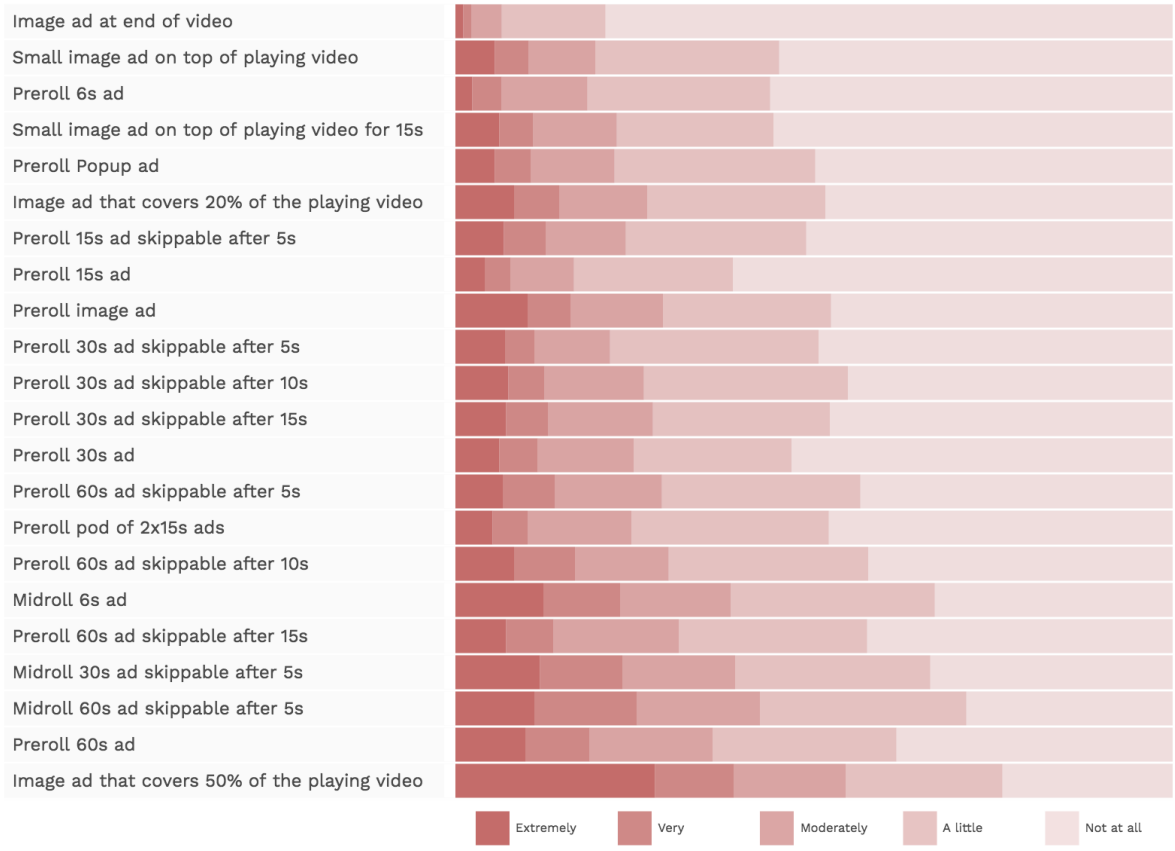
DESKTOP - Wait Time



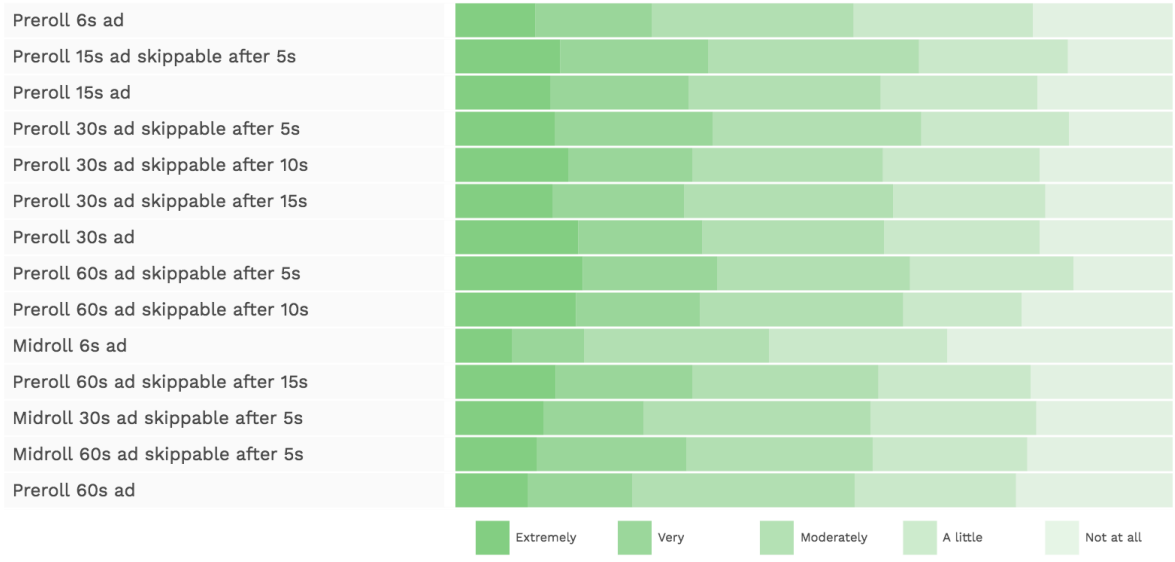
DESKTOP - Annoying



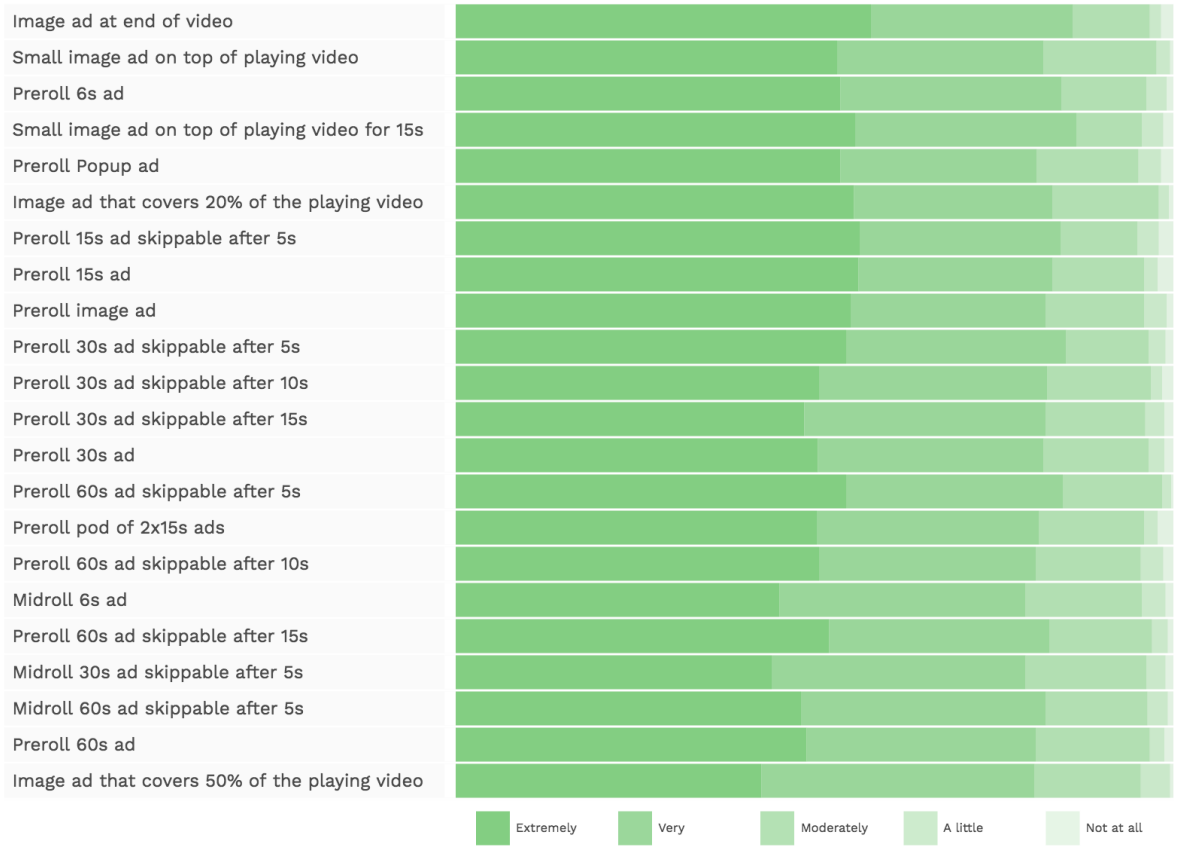
DESKTOP - Distracting



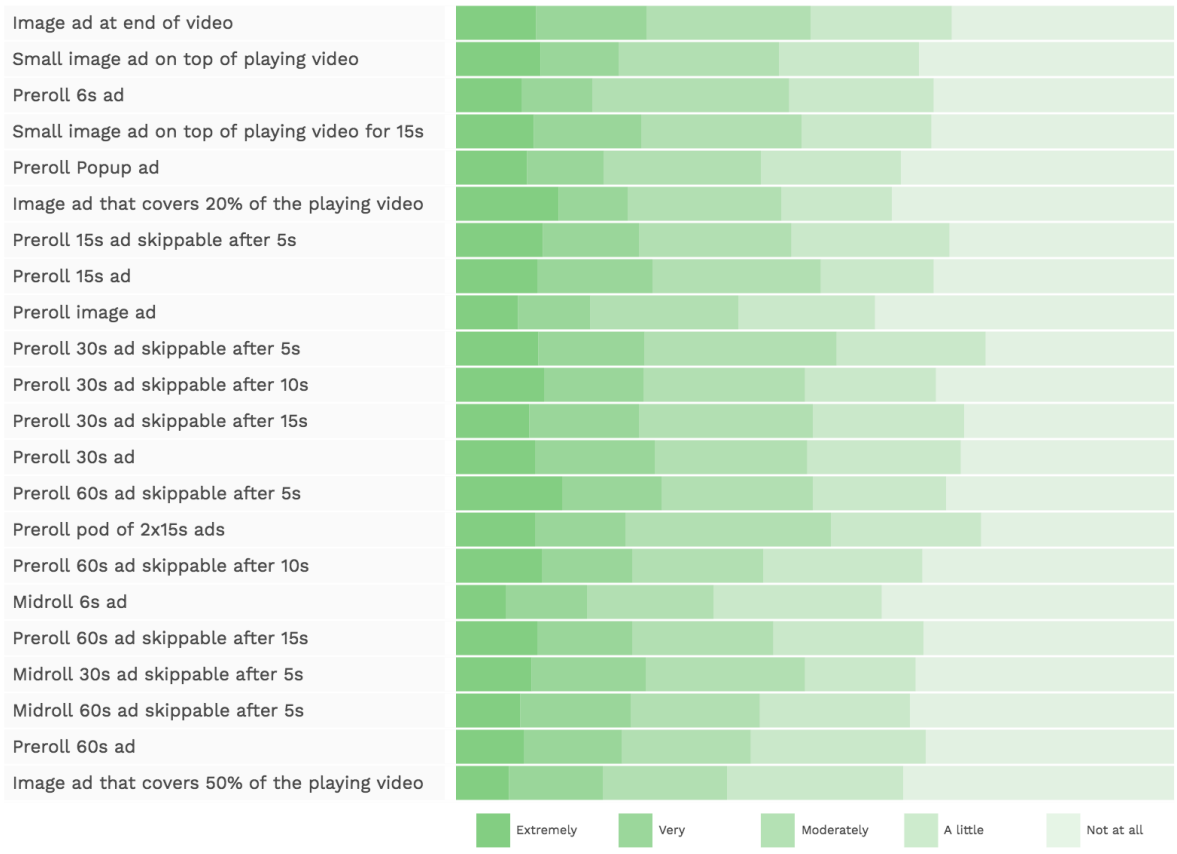
DESKTOP - Enjoyable (Ad Creative)



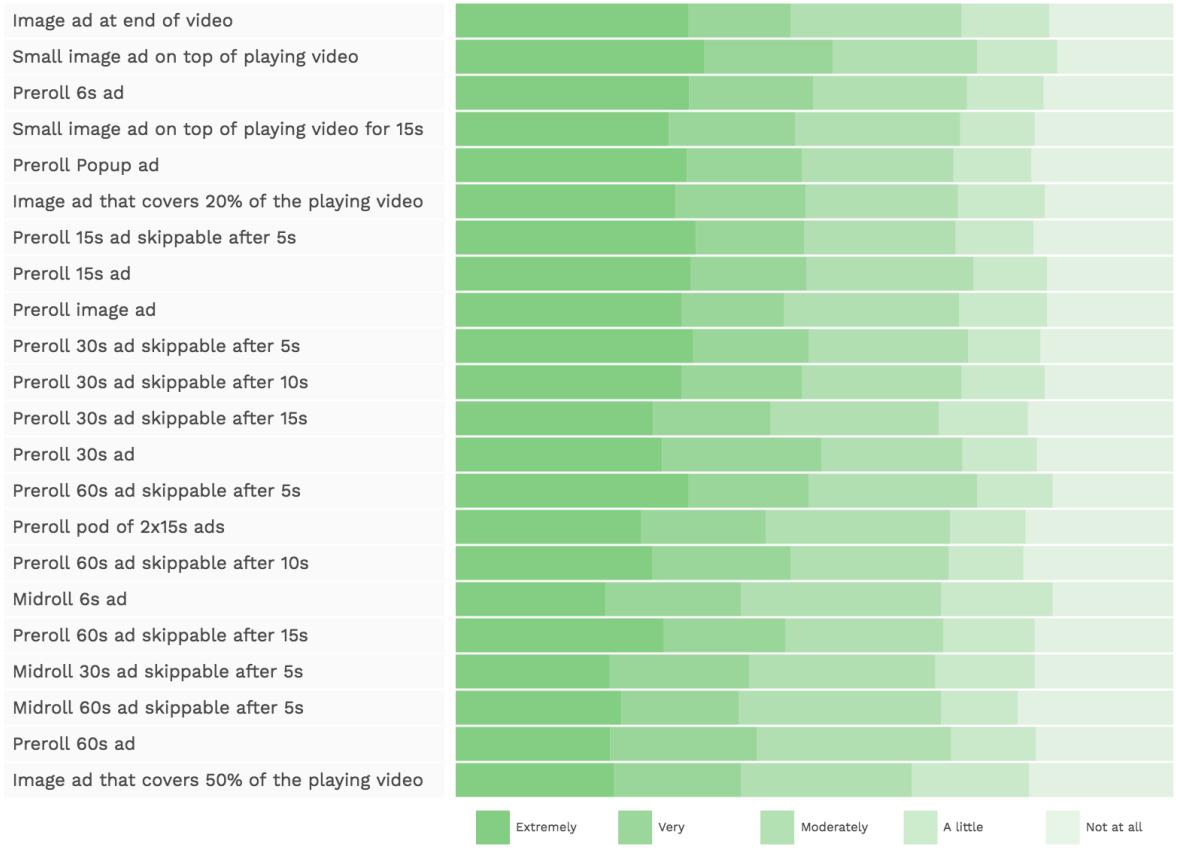
DESKTOP - Overall Satisfaction Watching Main Video



DESKTOP - Ad's Relevance to User



DESKTOP – Overall Webpage Behavior Predictability



DESKTOP - Too Personal (Ad Creative)