

An experimental methodology to measure consumer perceptions of ads in apps

November 2021

Table of Contents

Table of Contents

Abstract

Introduction

Experimental Overview

Experimental Methodology

 Overview: Experimental Design

 Main Adaptations to Study Casual Gaming Experiences

 Experiment Staging

 Study Participants

 Material

Results and Discussion

 Overview

 Ranking Chart

 Results

Conclusions

References

Appendix

 The survey instrument

 Ad Experiences We Tested

 Estimating the Rank Score Using the Bradley-Terry Algorithm

 Bradley-Terry Model

 PLS Regression - Rank Score on UX Metrics

 Demographic Distribution

 Full Results

Abstract

Ads can introduce users to new businesses and products that they might enjoy, enabling advertisers to grow their businesses and app developers to build sustainable app businesses. Many developers support their businesses with ads, helping them keep their apps free. Ad experiences that are interruptive or annoying can undermine the consumer's perception of the ad-supported app model leading users to uninstall apps or find other means to avoid ads entirely. Consumer dissatisfaction with ad experiences has the potential to disrupt the ad-supported app ecosystem and can create long term harm for publishers. Measuring consumer perceptions of ads can help the app publishing industry ecosystem to ensure ads don't drive users away.

The Coalition for Better Ads' predecessor project tackling consumer perceptions of desktop web-based ads that led to our Desktop Web Standard was built on the Multi-Ad Study (see *An Experimental Methodology to Rank N Ad Experiences by Consumers' Perceptions*, Ad Experience Research Group, 2016). In an article-reading environment, it was demonstrated users have a clear preference for ad experiences and measuring their preferences is possible. That work was extended into building a standard for short form videos (*An experimental methodology to measure consumer perceptions of ads in short-form video*). This paper presents the results from a new study to evaluate ad perceptions in an app environment. We utilized ads that were in common placements and formats in the casual gaming landscape. Names and definitions of these ad experiences will be introduced later in the paper.

This study utilized the same approach as previous CBA studies with adaptations to evaluate the perceptions of 31 in-app ad experiences in a casual game environment where participants played three levels. Each level contained a different ad experience. The overall game experience and each ad were ranked on various dimensions. The last task was for participants to rank the three ad experiences based on their preference within the game. Due to studying a large number of ad experiences we used the same iterative sampling approach as we have in our previous ad experience research to reduce the number of participants needed to evaluate the ads. The Bradley-Terry Algorithm (Turner and Firth, 2012) was utilized to create a unified stack rank of ad experiences by preference within the game.

We discovered clear differences between the ad experiences. The ads found to be most interfering by our respondents were those that interrupted the game play or were longer in length. The least interfering ads were those that did not cause a break in game play.

This research suggests that app developers can utilize the preferred ad experiences to monetize their app but also highlights the negative consumer perception of those that disrupt the main task. This study also provides the foundation for developing a standard that identifies the ad experiences that are least preferred by app users.

Introduction

Developers use ads in their apps to generate revenue. This helps keep free apps free, which allows more users to access valuable content that at least some wouldn't be able to access without ads. Ads themselves may sometimes introduce users to new businesses and products that they might enjoy. In the context of apps, ads can interfere with users' ability to use the app in several ways: interrupting content, distracting from content, or generally annoying users. Understanding how users perceive ads can help developers improve their apps' user experience and help ensure that ads shown in their apps don't drive users away.

Previous research by the Coalition for Better Ads has set the standard for mobile web, desktop web, and short form video. Our methodology for this study is based on previous CBA work, the Multi-Ad Study and short form video study (see *An Experimental Methodology to Rank N Ad Experiences by Consumers' Perceptions*, Ad Experience Research Group, 2016 and *An experimental methodology to measure consumer perceptions of ads in short-form video*, 2020). These studies rank ads experiences common on mobile and desktop web environments and in short form videos. In this paper we rank a set of ad experiences based on game players' perception of ad preference during game play. Though there are many app categories, we selected casual gaming because the casual gaming segment is one of the most popular segments for app ads. The ad experiences that were selected represent the most common ad experiences that show up in mobile apps.

The outline of the paper is as follows: a summary of the methods including adaptations from previous research, the results of testing 31 ad experiences with 45,289 participants, and a recommendation on where to set the standard to improve the overall quality of the ecosystem.

Experimental Overview

Described in this paper is a new experiment to evaluate ad experiences in a casual game environment. The ad experiences represent common formats (interstitial, banner, rewarded), creatives (image and video), and placements (app open, before level, after level).

Experimental Methodology

The methodology is based on two key principles: *adopting the user's perspective*, and measuring ads in the *context of common user activities online*. *An Experimental Methodology to Rank N Ad Experiences by Consumers' Perceptions* describes the methodology created to evaluate users' perceptions of ad experiences where the primary consumer experience was reading an article online. An ad experience is 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. *An experimental methodology to measure consumer perceptions of ads in short-form video* extended the research to include the consumer experience of watching short form videos.

A staged process that optimized which ad experiences to show to each participant was used 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, we used the Bradley-

Terry (B-T) Algorithm (Turner and Firth, 2012) to combine them into a unified ad preference ranking.

This methodology was similar to previous CBA multi-ad studies with slight adaptations to tailor the experience to a casual gaming environment. This section describes the methodology and how we adapted it from previous CBA studies.

Overview: Experimental Design

Similar to the previous experiments, a within-subjects design was employed. Each participant played three levels of a game, each of which contained an ad, and then answered questions about their experience. Similar to the video study, we removed the (ad-less) experience to keep the overall length of the study manageable. After playing each level, participants answered questions about their general game play experience, perception of the ads, and likelihood to engage in behaviors to avoid similar ad experiences in the future. At the end of the study they ranked the three ads based on the degree to which each ad they preferred.

Main Adaptations to Study Casual Gaming Experiences

The study design closely followed the established CBA methodology for studying consumer preferences for digital advertising experiences. We adapted the survey instrument to ensure that the ad attributes evaluated were relevant to the game play environment. We also added ad avoidance behavior to the instrument so that participants responded to the questions for each ad experience they encountered.

Stage 0: Modifying the Survey Framework

We ran a series of surveys with US participants to identify the factors that influence the perception of ads in apps. The initial phase was qualitative, in which 1,614 participants answered open ended questions about their perception of advertisements in smart phone apps. We selected 15 negative attributes and 15 positive attributes from the set of themes and had 1,114 respondents rate the extent to which adjective they considered characteristic of advertisements in apps. These data were used in a factor analysis to determine which adjectives had similar ratings. The factor analysis showed there were five underlying dimensions. The dimensions were how *useful*, *enjoyable*, *annoying*, *unsettling*, or *trustworthy* an ad was.

In creating the survey we used the 6 dimensions from the factor analysis. From the past CBA stack ranking studies and additional qualitative studies on the comprehension of the factors we conducted in 2020 we determined that the following ad-specific dimensions should be added:

- Distracting
- Relevant
- Misleading

We utilized three questions to understand how the ad impacted the participants' game play experience:

- Difficulty of playing the level
- How fun the level was
- Satisfaction with game play

We added questions on what behaviors participants would engage in to avoid ads like the one they experienced during the level. These questions were based on the ones used in previous studies. We added these questions to gather data on all formats to inform the recommendations for setting the standard.

The final survey contained the three game play experience questions, eight ad specific questions, and three ad avoidance questions. All of the questions used a labeled 5-point vertical scale.

- How easy or difficult was it to figure out how to play the game? (from *Very easy* to *Very difficult*)
- How fun was this game to play? (from *Extremely fun* to *Not fun at all*)
- How satisfied or dissatisfied are you with the overall experience playing this game? (from *Very satisfied* to *Very dissatisfied*)
- How annoying was the ad? (from *Extremely* to *Not at all*)
- How distracting was the ad? (from *Extremely* to *Not at all*)
- How appropriate was the timing of the ad's appearance during the game? (from *Extremely* to *Not at all*)
- How useful was the ad? (from *Extremely* to *Not at all*)
- To what extent does each of the following words or phrases describe the ad?
 - "Relevant to me" (from *Not at all* to *A great deal*)
 - "Trustworthy" (from *Not at all* to *A great deal*)
 - "Visually Pleasing" (from *Not at all* to *A great deal*)
 - "Misleading" (from *Not at all* to *A great deal*)
- Ad avoidance questions
 - How likely would you be to play a game with ads like this in the next 30 days? (from *Extremely* to *Not at all*)
 - How likely would you be to share a game with ads like this with someone else? (from *Extremely* to *Not at all*)
 - Based on your experience with this level, how worthwhile would you consider the effort to install an ad blocker? (from *Extremely* to *Not at all*)

Finally, we used the same construct as the video experiment where the ranking exercise was about *ad preference* (e.g., "Which one did you most / least prefer?"). We evaluated *ad preference* like the *webpage preferences* (from the Multi-Ad study).

Experiment Staging

Similar to previous CBA Multi-Ad and short form video experiments, we studied the experiences in three stages per platform, which maximized our 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 played three levels, each containing one ad experience from the full set we wanted to study, viewing a total of three ad experiences. We chose which three to ensure that, across the full set, the number of pairwise comparisons was balanced. For example, Anchored banners during the level appeared with in-game image interstitials just as often as they did with end-of-level video interstitials. We determined the number of participants such that each pair of ads was evaluated by roughly 16 participants. We used 17,324 participants in this stage. Given each participant's ranking of the three ads they saw, we used the Bradley-Terry algorithm to aggregate the rankings across all participants to create an initial stack-rank based on preference (see [Appendix](#)).

Stage 2: Refining the stack-rank

The stack-rank generated in Stage 1 allowed us to distinguish preference 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, we therefore restricted the sets of ads participants saw to only include those that were within ten ranks of each other. For example, a participant could see a set of ads that were ranked 2, 5, and 7 but not ones that were ranked 2, 5, and 16. We used 11,888 participants in this stage; each pair of ads was evaluated by roughly 16 participants. We then utilized the data from Stage 2 to create an updated stack-rank.

Stage 3: Further refining the stack-rank

In Stage 3, we further refined the stack-rank from Stage 2 by comparing experiences whose ranks were within six places of each other. For example, a participant could see a set of ads that were ranked 2, 5, and 7 but not ones that were ranked 2, 5, and 12. We used 16,077 participants in this stage.

Using the Bradley-Terry (B-T) Algorithm

We used the Bradley-Terry (B-T) algorithm (Turner and Firth, 2012) 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 preference; the greater the ability, the smaller the preference. We used the estimates of preference to rank the ads, and the confidence intervals allowed us to determine whether the differences between ads were statistically significant.

Study Participants

We used Kantar (formerly known as Answer Research), a market research firm, to recruit 45,289 mobile users. We surveyed participants across 7 geographical regions: Brazil, Germany, Great Britain, India, Indonesia, Japan, and the United States. We chose Kantar because of their ability to reach a large representative sample of global mobile users.

The company emailed individuals who had signed up to participate in studies and who matched our target criteria (i.e., device type and broad demographic traits matching that of each market's 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

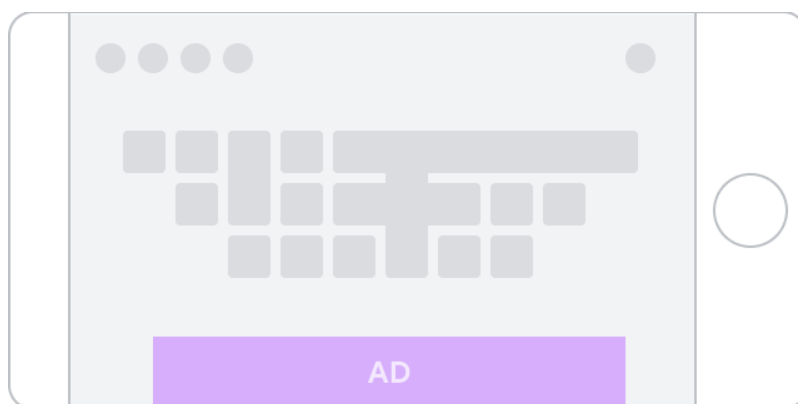
Overview

The experiment varied the ad experience while holding the game environment, ad creatives, and survey questions consistent. Each participant played three different levels of a game. Each level had a different ad experience and different creative.

Ad Creatives

We created video and image ad creatives for three fake casual games. We pretested the creatives to confirm that they were suitable. Survey pretests confirmed participants perceived the creatives for each of the three fake casual games as similarly boring or interesting. Since our goal was to test ad formats and experiences that affect game play, pretesting was necessary to minimize the effect of preference for ad creatives.

In total, we used 93 ad creatives for 3 casual game apps. Each app had three different versions of banner ads: one animated, one static, and one with excessive flashing (repeatedly flashing more than once per second). There were two image interstitials: portrait and landscape, with the same content in different configurations. There were two video creatives: 15 second videos and 30 second videos, both in landscape. The 15 second videos were cropped versions of the 30 second videos. We also included creatives that were made to imitate misleading ad experiences. These included an interstitial that was a fake playable and a video interstitial with a fake close button.



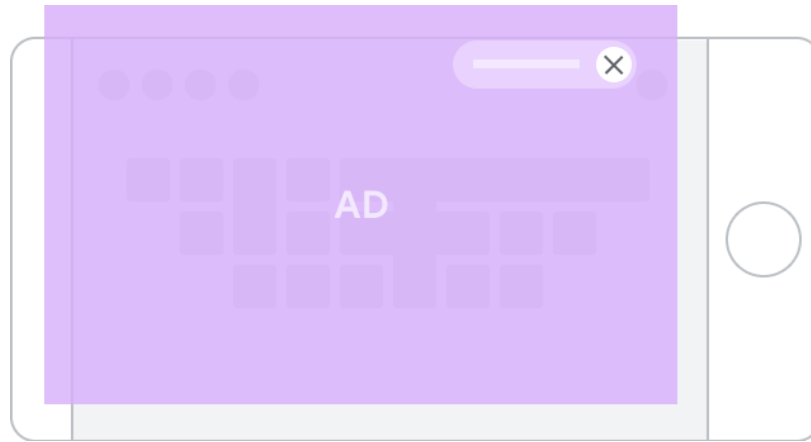


Figure 1. Examples of two ad creatives: a banner experience and an interstitial overlay.

Game Levels

To help ensure user engagement and minimize the potential for noise we evaluated the game levels against two criteria. The first criteria was how “fun” the game level was, the level had to meet a threshold to reduce the number of participants dropping out of the study. Second the levels had to be approximately the same level of “challenging” while mimicking a real game where levels get more challenging as you progress. Pretesting initially showed that level three was too challenging compared to the other levels, so we adjusted to make it only slightly harder while still showing a progression in how challenging the game was. Pretesting showed that over 80% of participants reported the game to be at least slightly fun and 90% rated it less than very difficult.

Comprehension, Overall Experience, and Ad-specific Questions

After each level, participants answered overall experience and ad-specific questions. The overall experience questions covered the level of fun and difficulty of the level. The ad-specific questions covered annoyance, distraction, usefulness, relevance, trustworthiness, the timing of the ad’s appearance, visual appeal, and misleading. To ensure participants paid attention while answering the questions, we included two attention-check questions 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?*”). We filtered out responses 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. The participants were shown instructions asking them to rank ads by which most and least interfered with their ability to play the game. Then the screenshots of the three ads were shown to refresh their memory and participants were asked to rate their choices based on preference. We filtered out responses from participants who chose the same ad experience as the most and least preferred.

It was intended that participants were asked about which most interfered and which least interfered. The follow up questions asking participants to rate their preference instead of interference was an error in the survey instrument. While this was an error, we believe that the data remains valid. The instructions displaying interference are likely to cause priming, could have possibly created confusion, and an awkward transition from interference to preference. Previous research demonstrated a high correlation between annoyance and distraction and the measures used in the stack ranking, we found a similar pattern that will be discussed in the results section. Because preference was found to be an effective measure in previous research both in our pilot studies and for the web standards we believe it to be effective in this study despite the potential confusion caused by mixed instructions.

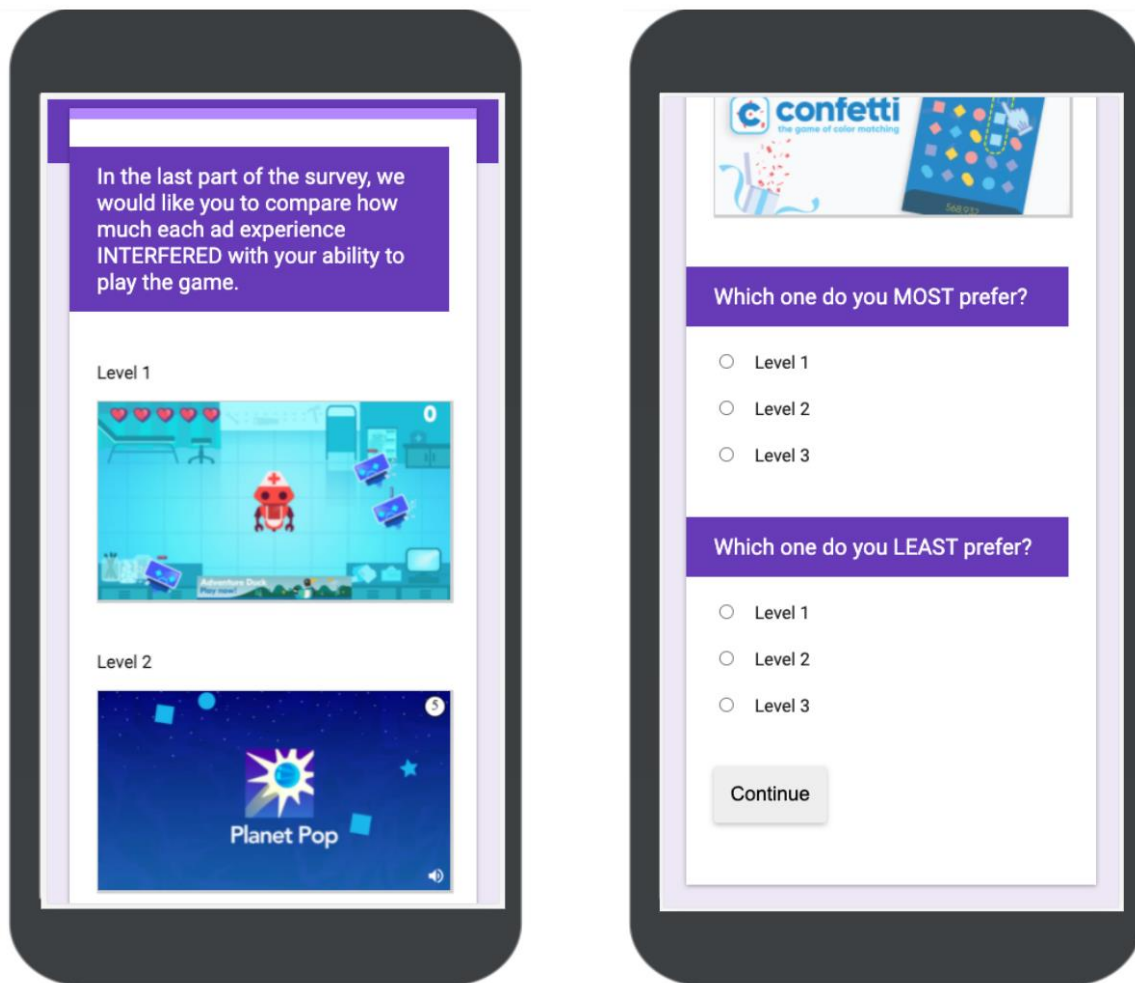


Figure 2. The top and bottom of the final ranking survey in the instream video experiment.

Localization

To adapt the environment to local markets, all text in the creatives, game levels, and survey were adapted to the language of the local market. These translations were verified with native speakers to ensure accuracy.

Results and Discussion

Overview

The analysis presented is aggregated across all markets. We present the overall ranking results, followed by the important user experience (UX) metric scores. The ranking shows how much the ads were preferred during the users' game play experience, while the UX metrics help us understand why ads are ranked in a certain way. These metrics are also a good measure of the broader user experience because they evaluate different ad attributes.

Ranking Chart

Each ad experience has a rank score estimate with a 95% confidence interval. For ads in which the error bars overlap, the rank scores are similar. This similarity of score means that perceived preference among the ads do not significantly differ.

Results

Image interstitial ads after the level has been played are rated as the best ad experiences. These are significantly better than interstitials that are before or during game play as well as video interstitials. The main differences between video and image interstitial ads are the amount of time before dismissal and whether there is motion and sound during the experience.

Anchored banner ads are clustered as the second best rating. One of the differentiating characteristics of banner ads is that they don't require dismissal, allowing the game experience to remain present on the screen.

Video interstitials rank as the worst ad experiences. Length, skip delay times, and placement of the ad in the environment all contribute to the preference score.

Mobile Ad Experience Rankings

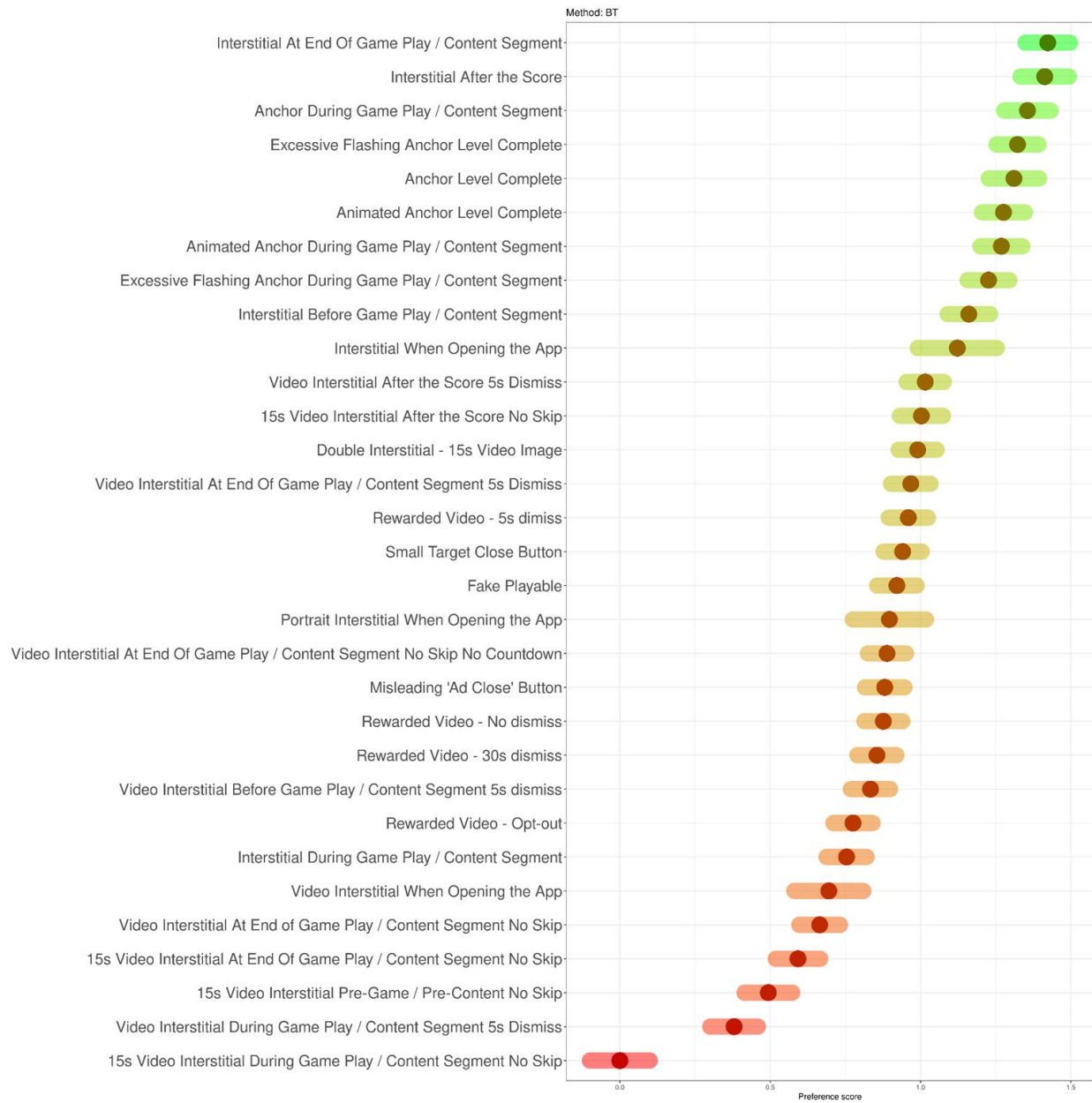


Figure 3. Rank scores of the 31 mobile ad experiences, based on preference

Timing matters: Interrupting game play makes the ad experience worse

This pattern is observed across ad experiences we tested (i.e. pre game, during the level, post completion and post score). When looking at the ad experiences, those that showed up during game play or before game play ranked consistently near the bottom. Ad experiences like the anchor during game play or once game play was completed ranked closer to the top. This pattern makes sense as ads that show during or before game play are interrupting the primary task the user is trying to accomplish.

Taking control away makes an ad experience rank worse

The first way that control was removed was by altering the skip time. The skip time is the duration of the ad users had to watch before they could proceed to the game. Ad experiences that were unskippable or had longer watch times before skip was available generally ranked worse than those that could be skipped immediately or with a short duration. In line with removing control, turning the rewarded experience from opt-in to opt-out reduced the preference score compared to the opt-in experiences.

Ad experiences with bad intent don't fool everyone

Three trick-to-click experiences were included in the test. These were the fake close button, the small close button, and a fake playable. These experiences ranked low on preference but not near the bottom. Not all participants were tricked by the experience with 44% trying to interact with the fake playable, 23% trying to click on the fake close button, and 17.6% missing the close button on the small target.

Annoyance and distraction are the most predictive of rating difference across ad experience

Of the 8 metrics we collected, a Partial Least Squares Regression analysis demonstrated how annoying and how distracting an ad experience was most predicted the rank score. The next most predictive measures were the ads timing, how misleading the ad was, how visually pleasing the ad was, and how trustworthy the ad was. See Appendix for more details.

Annoyance, Distraction, Misleading, Appropriate timing, and Visually Pleasing

The figures below order the ad experiences by their position in the overall rank (from the most preferred to the least preferred). The stacked bars represent the distribution of participant ratings on a 1-5 scale for each ad metric.

Annoyance and **distraction** directionally aligns with the overall preference ranking with both metrics most strongly correlating. The ads during game play receive slightly higher annoyance and distraction ratings than other ads placements including the equivalent in other placements. Additionally the small target close button ranks as more annoying and distracting likely due to it being harder to tap.

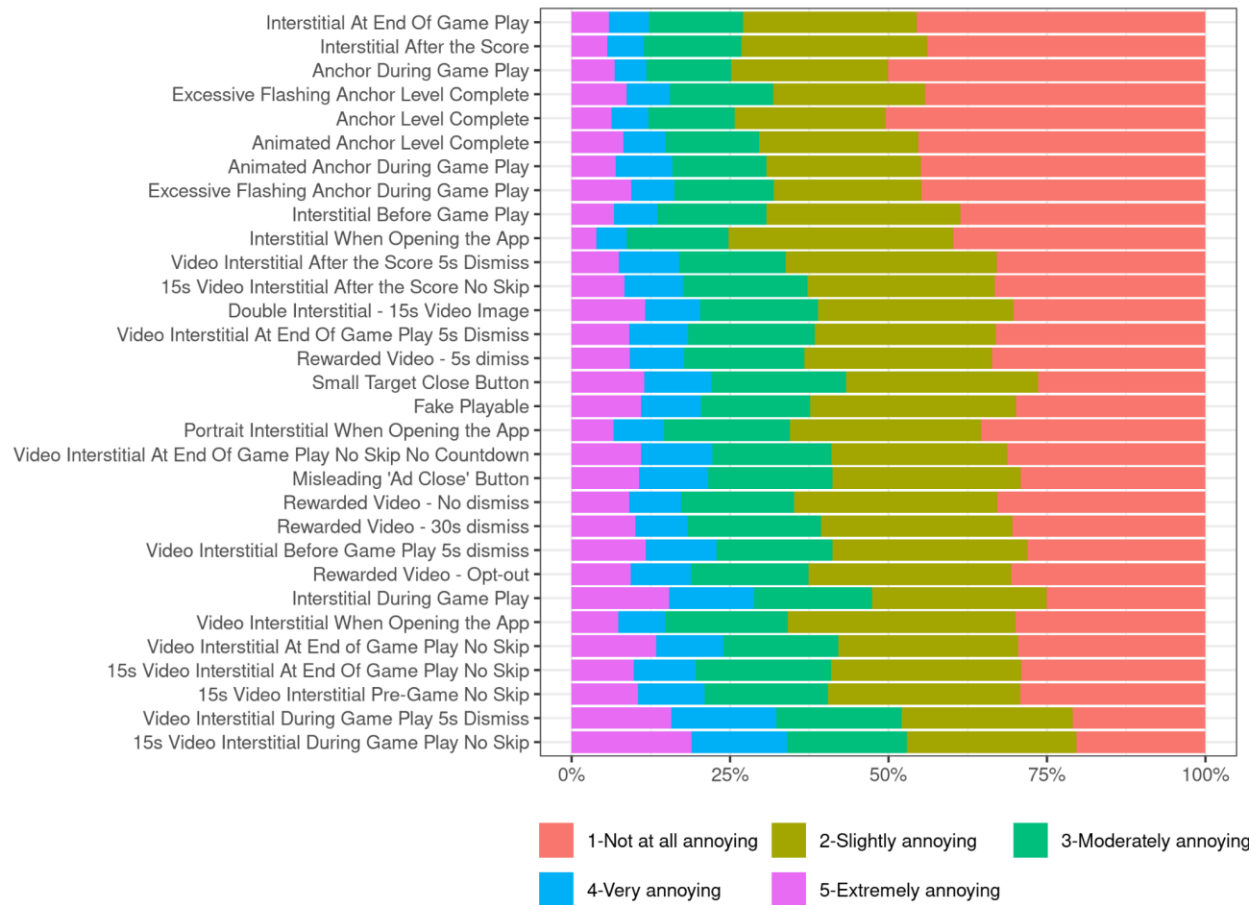


Figure 4. Annoyance

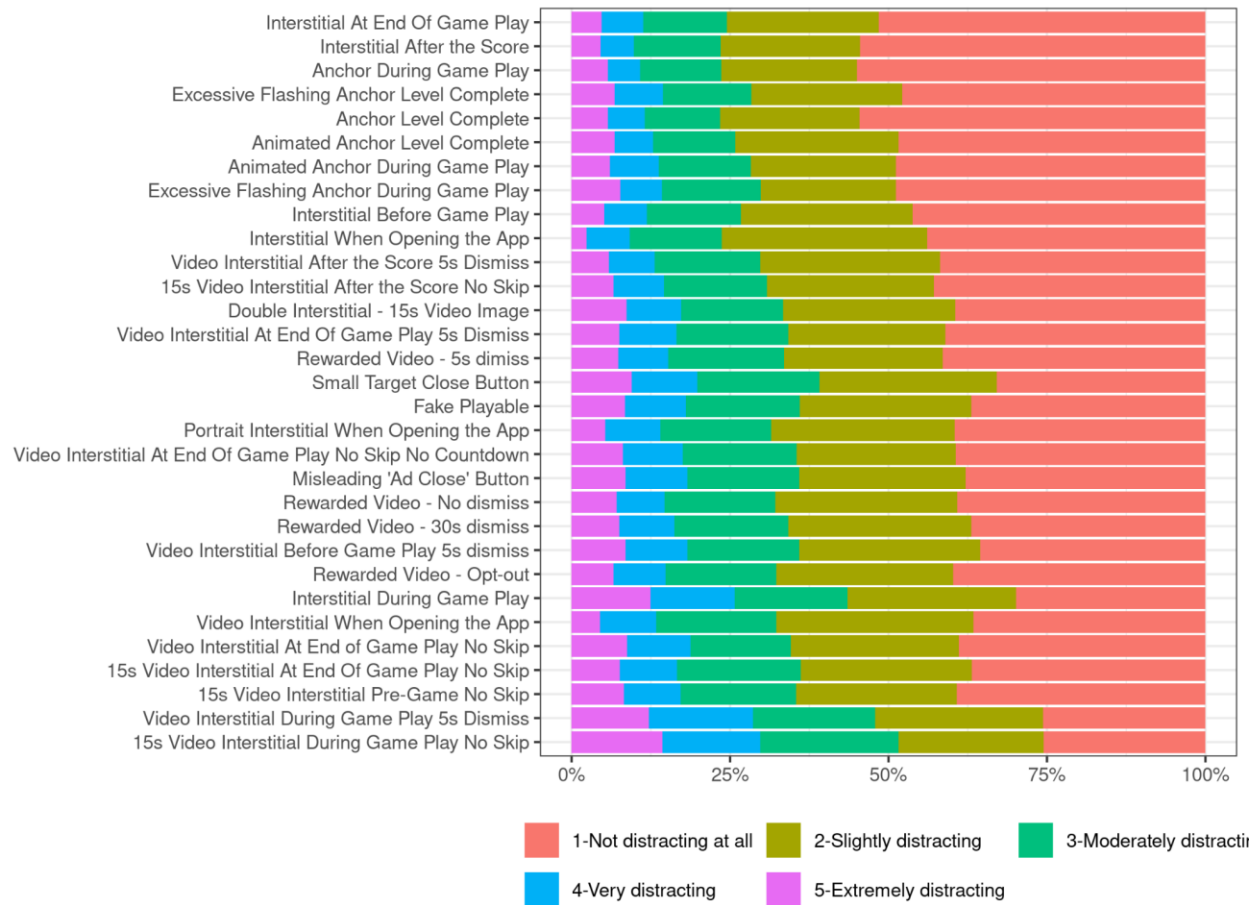


Figure 5. Distraction

The “**Appropriate Timing**” metric was noticeably worse for the three ad experiences that interrupted game play (statistically significant). Most other ads are perceived more favorably and have similar ratings. This pattern is intuitive because interrupting the game with an ad is a poorly ranked experience.

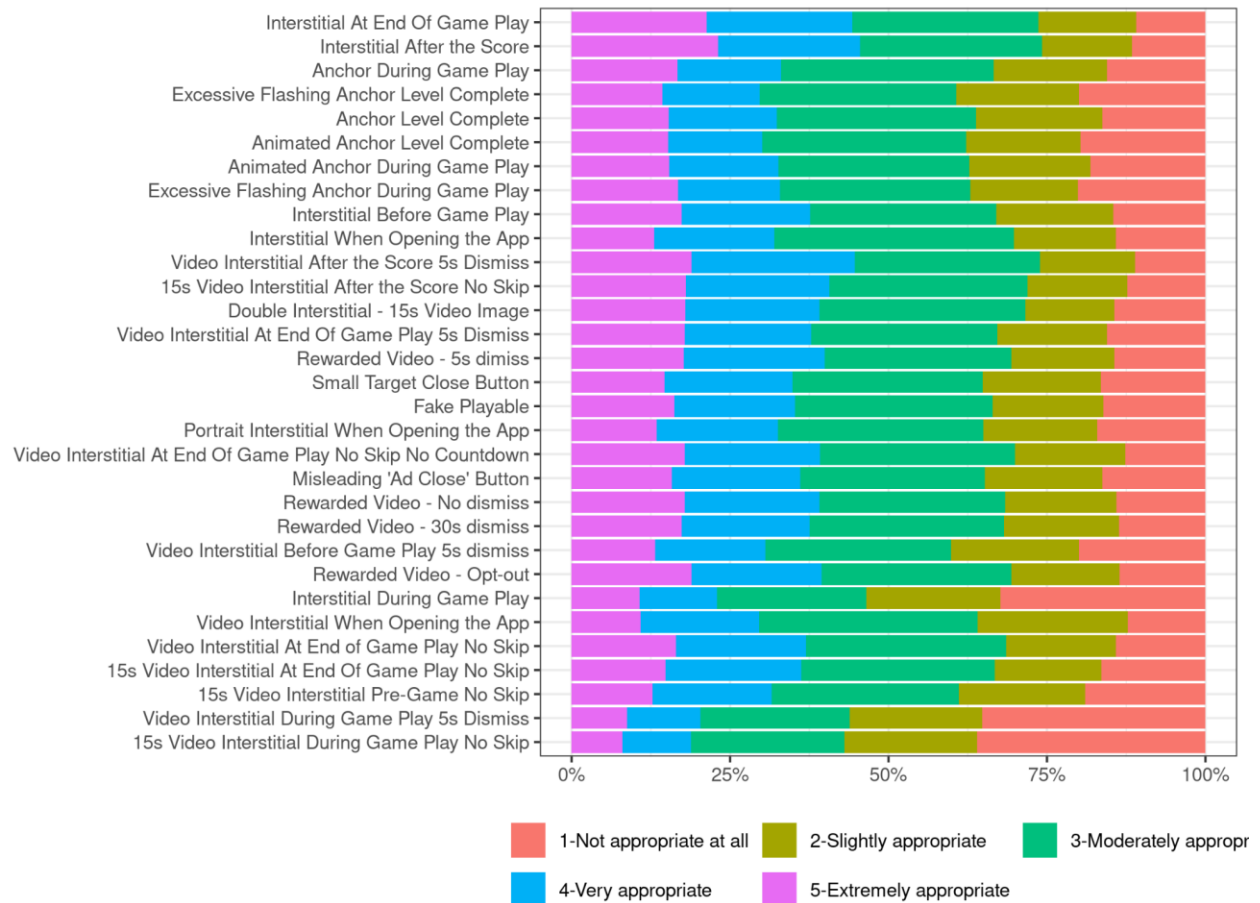


Figure 6. Appropriate timing

The “**Misleading**” metric. While it is fairly consistent across ad experiences, the fake playable experience, interstitials when opening the app, and interstitials during game play are all considered slightly more misleading (statistically significant). The fake playable is an obviously misleading experience so the ranking is intuitive. Interstitials when opening the app or during game play could be misleading due to the expectation of playing the game but encountering an ad experience instead.

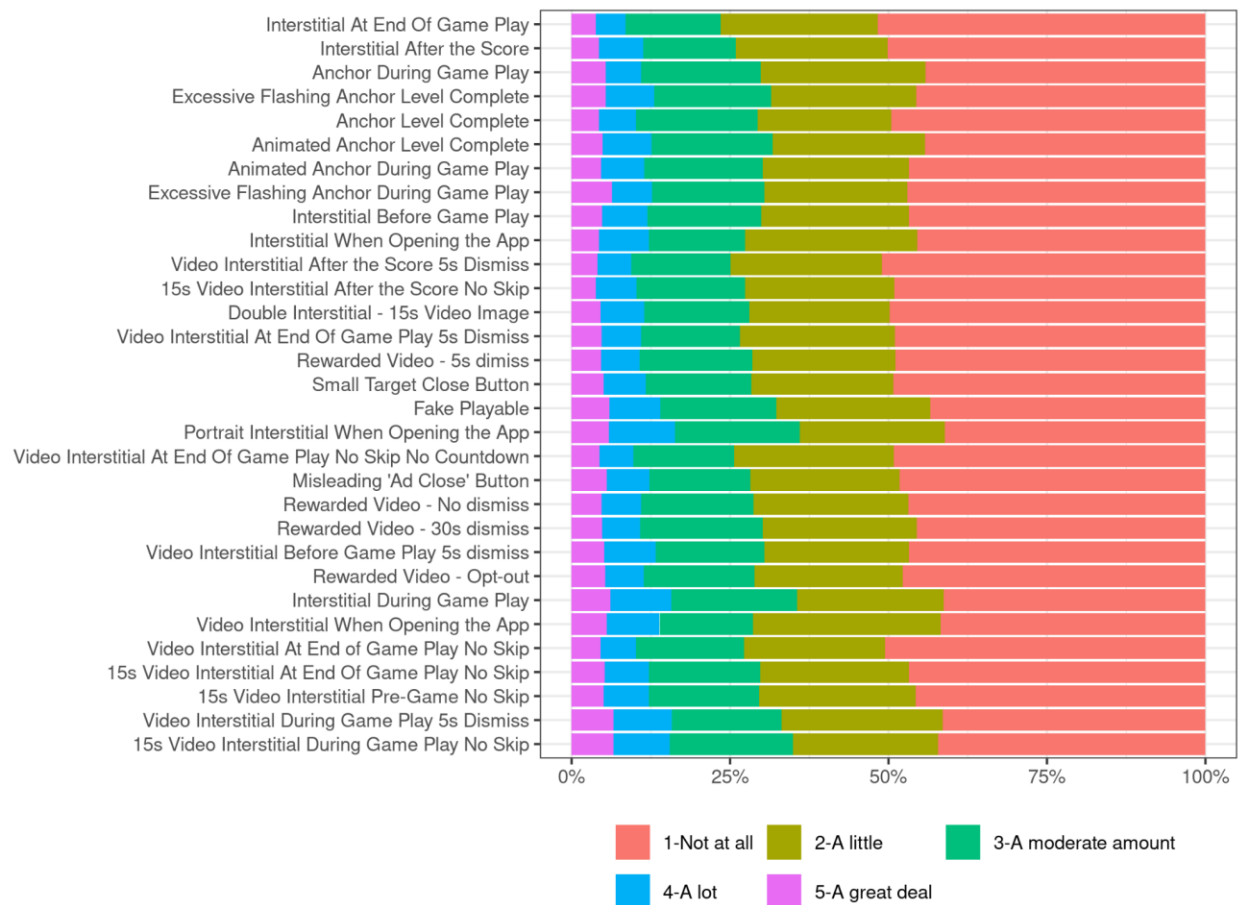


Figure 7. Misleading

Visually pleasing is the final metric that is related to the overall preference ranking. While it is fairly consistent across ad experiences, the anchor and fake playable experiences have a lower ranking (statistically significant). This suggests that utilizing the full screen with engaging creatives allows for the increase in visual appeal and higher preference.



Figure 8. Visually Pleasing

The Other Metrics: Relevant, Useful, and Trustworthy

Most ads generally were rated similarly to each other for the relevant, useful, and trustworthy metrics (the graphs visualizing these metrics can be found in the appendix). Since these ads are not targeted toward the user it is expected that relevance and usefulness wouldn't be any different. Since we used creatives for standard games, seeing no movement on trustworthiness is expected.

Conclusions

This paper describes how the CBA adapted the Multi-Ad Study methodology to evaluate perceptions of 31 ad experiences in a casual game environment. Using this approach, we efficiently ranked ad experiences by preference during the game playing experience using an algorithm to optimize which experiences we showed each participant and the Bradley-Terry model to estimate the overall ranking.

Results suggest certain ad experiences are preferred more than others by users while playing a game. Generally, the least preferred ads were ones that interrupted users during game play or

made users wait before they could resume game play. Additional analysis exploring the relationship between individual ad dimensions and the stack-rank confirmed that perceived annoyance, distraction, and the timing of the ad's appearance most explained variance in the stack-rank.

The approach described in this paper can be effectively used to study user perceptions of additional or future in-game ad experiences. As with the previous studies, it can discriminate consumer perceptions of annoyance and interruption between ad experiences across many ads, in the context of a natural activity. Moreover, new ad experiences can be incorporated into the stack-rank in response to industry innovation, and can be adapted to study other environments, while maintaining the same core design.

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/.

Ad Experience Research Group (2016). *Determining a Better Ads Standard Based on User Experience Data*. From betterads.org/research/standardpaper/.

Ad Experience Research Group (2020). *An experimental methodology to measure consumer perceptions of ads in short-form video*. 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.

How often do you play games on your mobile phone?

- Multiple times a day
- Once a day
- Multiple times a week
- Once a week
- Multiple times a month
- Once a month or less

- Never. I don't play games on my phone

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

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 \$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

Overall Experience Questions

How easy or difficult was it to figure out how to play the game?

- Very easy
- Somewhat easy
- Neither easy nor difficult
- Somewhat difficult
- Very difficult

How fun was this game to play?

- Extremely fun

- Very fun
- Moderately fun
- Slightly fun
- Not fun at all

How satisfied or dissatisfied are you with the OVERALL EXPERIENCE playing this game?

- Very satisfied
- Somewhat satisfied
- Neither satisfied or dissatisfied
- Somewhat dissatisfied
- Very dissatisfied

Ad Specific Questions

We showed a screenshot of the ad participants saw to refresh their memory of the ad.

In the next part of the survey, we would like to ask you about an ad you may have notified while playing the game.

[Participant sees screenshot of the ad experience]

How annoying was the ad?

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

How distracting was the ad?

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

How appropriate was the timing of the ad's appearance during the game?

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

How useful was the ad?

- Extremely useful

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

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

“Relevant to me”

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

“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

“Misleading”

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

We would like you to continue thinking about the ad you may have noticed while playing the game.

How likely would you be to play a game with ads like this in the next 30 days?

- Not at all likely
- Slightly likely
- Moderately likely

- Very likely
- Extremely likely

How likely would you be to share a game with ads like this with someone else?

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

Based on your experience with this level, how worthwhile would you consider the effort to install an ad blocker?

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

Ad Ranking Exercise

In the last part of the study, we would like you to compare how much each ad experience INTERFERED with your ability to play the game.

Level 1 Ad

[Participant sees screenshot of the ad experience]

Level 2 Ad

[Participant sees screenshot of the ad experience]

Level 3 Ad

[Participant sees screenshot of the ad experience]

Which one do you MOST prefer?

- Level 1
- Level 2
- Level 3

Which one do you LEAST prefer?

- Level 1
- Level 2
- Level 3

Follow Up Questions

Do you use any of the following ways to avoid ads in mobile apps?

- Close the App

- Use a different app for the same thing
- Put the phone down
- Close the ad to stop seeing it
- Pay to remove ads
- Put phone in airplane mode
- Ignore the ads
- Uninstall the App
- None of the above

Do you block ads on your mobile phone?

- Yes
- No
- Unsure

Do you currently use an ad blocker on your desktop computer?

- Yes
- No
- Unsure

If you could install ad blocking software on your mobile phone, how likely would you be to install it in the next 3-6 months?

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

Ad Experiences We Tested

In-Game Experiences

Name	Description	Link
Animated Anchor During Game Play / Content Segment	An animated banner ad that is present at the bottom of the screen during game play / during a content segment	http://snapshot-dot-poetic-glass-136423.appspot.com/experience?exp=GAM E_APP_ANCHOR_BOTTOM_ANIMATED&advertiser_id=9&article_id=0
Animated Anchor Level Complete	An animated banner ad that is present at the bottom of the level complete screen	http://snapshot-dot-poetic-glass-136423.appspot.com/experience?exp=GAM E_APP_ANCHOR_BOTTOM_END_ANIMATED&advertiser_id=9&article_id=0
Anchor During Game Play / Content Segment	A static banner ad that is present at the bottom of the screen during game play / during a content segment	http://snapshot-dot-poetic-glass-136423.appspot.com/experience?exp=GAM E_APP_ANCHOR_BOTTOM&advertiser_id=9&article_id=0

Anchor Level Complete	A static banner ad that is present at the bottom of the level complete screen	http://snapshot-dot-poetic-glass-136423.appspot.com/experience?exp=GAM E_APP_ANCHOR_BOTTOM_END&advertiser_id=9&article_id=0
Portrait Interstitial When Opening the App	Full screen interstitial image ad displayed in portrait orientation that appears after the app icon is selected and before the splash screen	http://snapshot-dot-poetic-glass-136423.appspot.com/experience?exp=GAM E_APP_APP_OPEN_INTERSTITIAL_PORTRAIT&advertiser_id=9&article_id=0
Interstitial Before Game Play / Content Segment	Full screen interstitial image ad that appears after the splash screen and before game play / before a content segment	http://snapshot-dot-poetic-glass-136423.appspot.com/experience?exp=GAM E_APP_INTERSTITIAL_PREGAME&advertiser_id=9&article_id=0
Interstitial During Game Play / Content Segment	Full screen interstitial image ad that appears during game play at the beginning of the level / during the beginning of a content segment	http://snapshot-dot-poetic-glass-136423.appspot.com/experience?exp=GAM E_APP_INTERSTITIAL_INGAME&advertiser_id=9&article_id=0
Interstitial At End Of Game Play / Content Segment	Full screen interstitial image ad that appears after game play before the level complete screen / at the end of a content segment	http://snapshot-dot-poetic-glass-136423.appspot.com/experience?exp=GAM E_APP_INTERSTITIAL_ENDGAME&advertiser_id=9&article_id=0
Interstitial After the Score	Full screen interstitial image ad that appears after the level complete screen	http://snapshot-dot-poetic-glass-136423.appspot.com/experience?exp=GAM E_APP_INTERSTITIAL_AFTERSCORE&advertiser_id=9&article_id=0
Video Interstitial Before Game Play / Content Segment 5s dismiss	Full screen interstitial 30s video ad that appears after the splash screen and before game play / a content segment. Skippable after 5s with countdown	http://snapshot-dot-poetic-glass-136423.appspot.com/experience?exp=GAM E_APP_VIDEO_PREGAME_CD5&advertiser_id=9&article_id=0
Video Interstitial During Game Play / Content Segment 5s Dismiss	Full screen interstitial 30s video ad that appears during game play at the beginning of the level / during the beginning of a content segment. Skippable after 5s with countdown	http://snapshot-dot-poetic-glass-136423.appspot.com/experience?exp=GAM E_APP_VIDEO_INGAME_CD5&advertiser_id=9&article_id=0
Video Interstitial At End Of Game Play / Content Segment 5s Dismiss	Full screen interstitial 30s video ad that appears after game play before the level complete screen / at the end of a content segment. Skippable after 5s with countdown	http://snapshot-dot-poetic-glass-136423.appspot.com/experience?exp=GAM E_APP_VIDEO_ENDGAME_CD5&advertiser_id=9&article_id=0

Video Interstitial After the Score 5s Dismiss	Full screen interstitial 30s video ad that appears after the level complete screen. Skippable after 5s with countdown	http://snapshot-dot-poetic-glass-136423.appspot.com/experience?exp=GAM E_APP_VIDEO_AFTERSCORE_CD5&advertiser_id=9&article_id=0
15s Video Interstitial Pre-Game / Pre-Content No Skip	Full screen interstitial 15s video ad that appears after the splash screen and before game play / before a content segment. Unskippable with countdown	http://snapshot-dot-poetic-glass-136423.appspot.com/experience?exp=GAM E_APP_15S_VIDEO_PREGAME_NOSKIP&advertiser_id=9&article_id=0
15s Video Interstitial During Game Play / Content Segment No Skip	Full screen interstitial 15s video ad that appears during game play at the beginning of the level / during the beginning of a content segment. Unskippable with countdown	http://snapshot-dot-poetic-glass-136423.appspot.com/experience?exp=GAM E_APP_15S_VIDEO_INGAME_NOSKIP&advertiser_id=9&article_id=0
15s Video Interstitial At End Of Game Play / Content Segment No Skip	Full screen interstitial 15s video ad that appears after game play before the level complete screen / at the end of a content segment. Unskippable with countdown	http://snapshot-dot-poetic-glass-136423.appspot.com/experience?exp=GAM E_APP_15S_VIDEO_ENDGAME_NOSKIP&advertiser_id=9&article_id=0
15s Video Interstitial After the Score No Skip	Full screen interstitial 15s video ad that appears after the level complete screen. Unskippable with countdown	http://snapshot-dot-poetic-glass-136423.appspot.com/experience?exp=GAM E_APP_15S_VIDEO_AFTERSCORE_NOSKIP&advertiser_id=9&article_id=0
Video Interstitial At End Of Game Play / Content Segment No Skip No Countdown	Full screen interstitial 30s video ad that appears after game play before the level complete screen / at the end of a content segment. Unskippable with no countdown	http://snapshot-dot-poetic-glass-136423.appspot.com/experience?exp=GAM E_APP_VIDEO_AFTERSCORE_NOSKIP_NOCD&advertiser_id=9&article_id=0
Video Interstitial At End of Game Play / Content Segment No Skip	Full screen interstitial 30s video ad that appears after game play before the level complete screen / at the end of a content segment. Unskippable with countdown	http://snapshot-dot-poetic-glass-136423.appspot.com/experience?exp=GAM E_APP_VIDEO_AFTERSCORE_NOSKIP&advertiser_id=9&article_id=0
Double Interstitial - 15s Video, Image	Full screen interstitial 15s video ad that appears after the level complete screen. Unskippable with countdown. Followed by an image interstitial.	http://snapshot-dot-poetic-glass-136423.appspot.com/experience?exp=GAM E_APP_DOUBLE_INTERSTITIAL_AFTERSCORE&advertiser_id=9&article_id=0

Interstitial When Opening the App	Full screen interstitial image ad that appears after the app icon is selected and before the splash screen	http://snapshot-dot-poetic-glass-136423.appspot.com/experience?exp=GAM E_APP_APP_OPEN_INTERSTITIAL&advertiser_id=9&article_id=0
Video Interstitial When Opening the App	Full screen interstitial 30s video ad that appears after the app icon is selected and before the splash screen. Skippable.	http://snapshot-dot-poetic-glass-136423.appspot.com/experience?exp=GAM E_APP_APP_OPEN_VIDEO&advertiser_id=9&article_id=0
Rewarded Video - 5s dismiss	Opt in rewarded ad experience. Full screen interstitial 30s video that appears after the level complete screen / after the end of a content segment. Can dismiss after 5s and receive reward	http://snapshot-dot-poetic-glass-136423.appspot.com/experience?exp=GAM E_APP_REWARDED_AFTERSCORE_CD5&advertiser_id=9&article_id=0
Rewarded Video - 30s dismiss	Opt in rewarded ad experience. Full screen interstitial 30s video that appears after the level complete screen / after the end of a content segment. Countdown displayed. If dismissed before 30s reward is forfeit	http://snapshot-dot-poetic-glass-136423.appspot.com/experience?exp=GAM E_APP_REWARDED_AFTERSCORE&advertiser_id=9&article_id=0
Rewarded Video - No dismiss	Opt in rewarded ad experience. Full screen interstitial 30s video that appears after the level complete screen / after the end of a content segment. No countdown displayed. Cannot dismiss for 30s	http://snapshot-dot-poetic-glass-136423.appspot.com/experience?exp=GAM E_APP_REWARDED_AFTERSCORE_NOS KIP_NOCD&advertiser_id=9&article_id=0
Rewarded Video - Opt-out	Rewarded ad experience with a 5s opt out. Full screen interstitial 30s video that appears after the level complete screen / after the end of a content segment. Can dismiss after 5s countdown and receive reward	http://snapshot-dot-poetic-glass-136423.appspot.com/experience?exp=GAM E_APP_REWARDED_OPTOUT_AFTERSCORE_CD5&advertiser_id=9&article_id=0
Fake Playable	Opt in rewarded ad experience. Full screen interstitial animated ad that appears after the level complete screen / after the end of a content segment. Ad looked like it was interactive but clicking on it went to a fake app download page. Can dismiss after 5s countdown and receive reward	http://snapshot-dot-poetic-glass-136423.appspot.com/experience?exp=GAM E_APP_REWARDED_FAKE_PLAYABLE_AFTERSCORE_CD5&advertiser_id=9&article_id=0
Excessive Flashing Anchor During Game Play	An animated banner ad with excessive flashing that is present at the bottom of the screen during game play / during a content segment	http://snapshot-dot-poetic-glass-136423.appspot.com/experience?exp=GAM E_APP_ANCHOR_BOTTOM_END_EXCESSIVE_FLASHING&advertiser_id=9&article_id=0

Excessive Flashing Anchor Level Complete	An animated banner ad with excessive flashing that is present at the bottom of the level complete screen	http://snapshot-dot-poetic-glass-136423.appspot.com/experience?exp=GAM_E_APP_ANCHOR_BOTTOM_EXCESSIVE_FLASHING&advertiser_id=9&article_id=0
Misleading 'Ad Close' Button	Full screen interstitial 30s video ad that appears after game play before the level complete screen / after a content segment. Immediately Skippable. A close button that doesn't close the ad is present on the screen.	http://snapshot-dot-poetic-glass-136423.appspot.com/experience?exp=GAM_E_APP_FAKE_CLOSE_VIDEO_ENDGAME&advertiser_id=9&article_id=0
Small Target Close Button	Full screen interstitial 30s video ad that appears after game play before the level complete screen / after a content segment. Immediately Skippable. The close button target is reduced in size to 60%.	http://snapshot-dot-poetic-glass-136423.appspot.com/experience?exp=GAM_E_APP_DIFFICULT_CLOSE_VIDEO_ENDGAME&advertiser_id=9&article_id=0

Estimating the Rank Score Using the Bradley-Terry Algorithm

The data we collected are all individual-level ranks of three ad combinations, out of about 31 ad experiences. The total number of ad experiences is not directly relevant to the analysis. We aggregate the participants' rankings of the three ads they see into a set of paired comparisons. We then apply the Bradley-Terry model to this data and estimate a latent ability parameter for each ad, where ability means how much the ad is preferred during participants' game playing experience.

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 (larger values of α mean higher 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 a 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.

PLS Regression - Rank Score on UX Metrics

In order to check that the rank score is a valid single metric for measuring ad preferences, we would like to model its relationship with the UX metrics we collect. 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. We would like to 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, annoying and distracting have a highly positive correlation (Figure 15). The multicollinearity between the UX metrics means that directly fitting a multiple regression with the UX metrics as predictors will be problematic. We therefore use Partial Least Squares (PLS) regression (Abdi 2010), 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 UX metrics (see Figure 16 below).

(Note: the rank score is measured at the ad level, while the UX metrics are measured at the participant level. We therefore calculate the percentage of the least favorite responses for each ad and use these as the explanatory variables in PLS.)

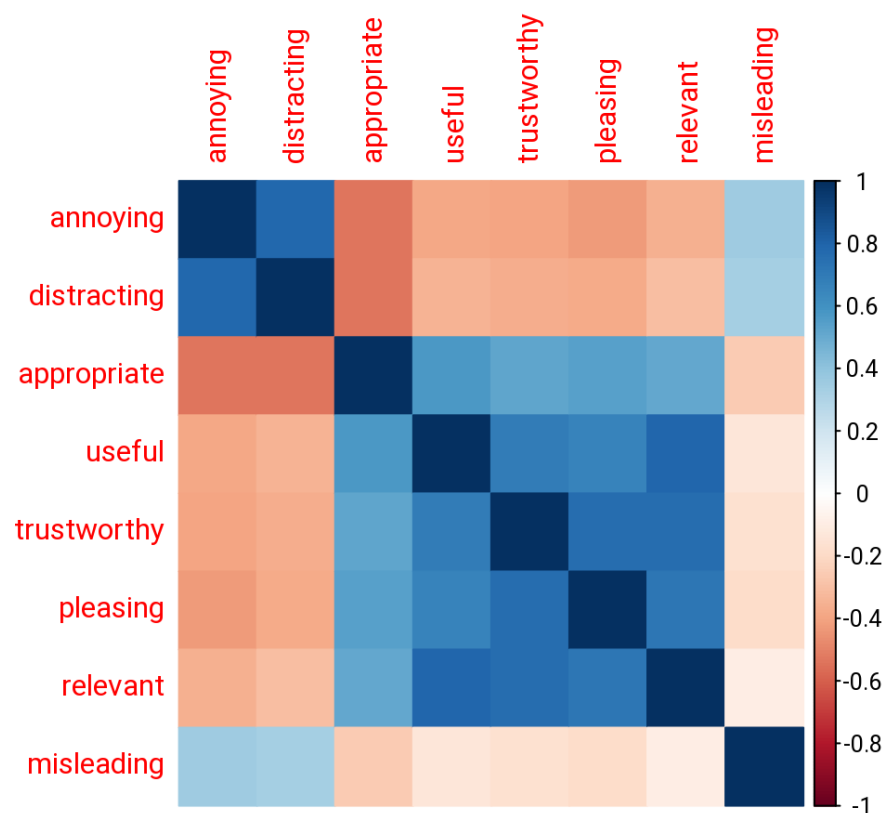


Figure 15: UX metrics - pairwise correlation matrix plot

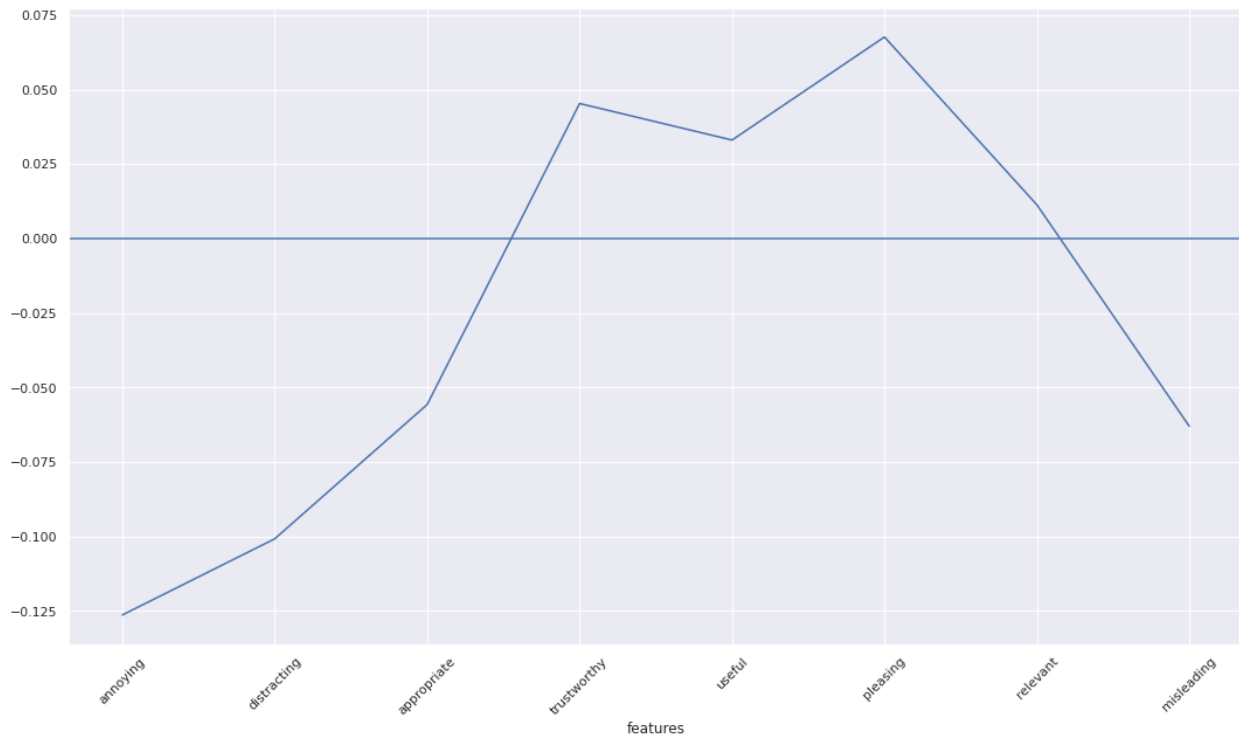
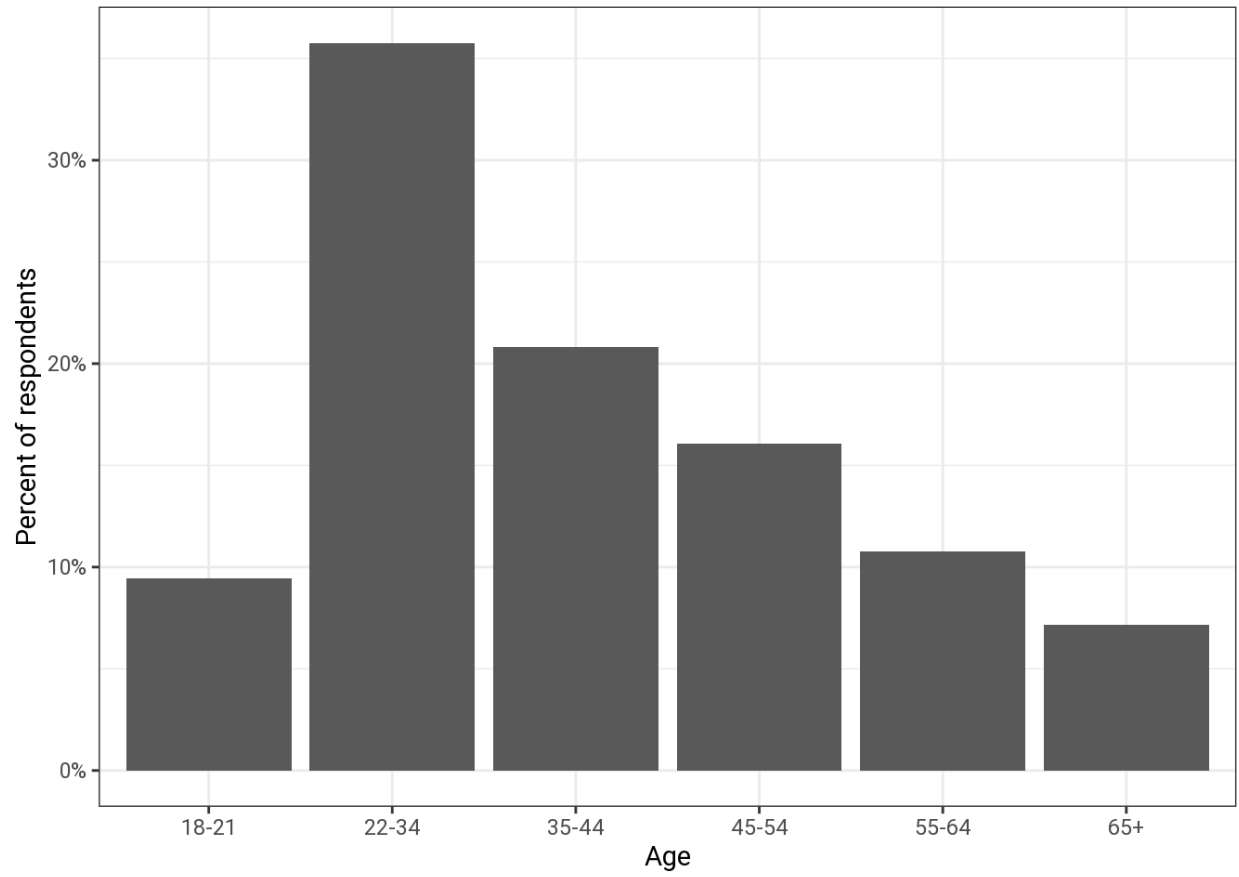
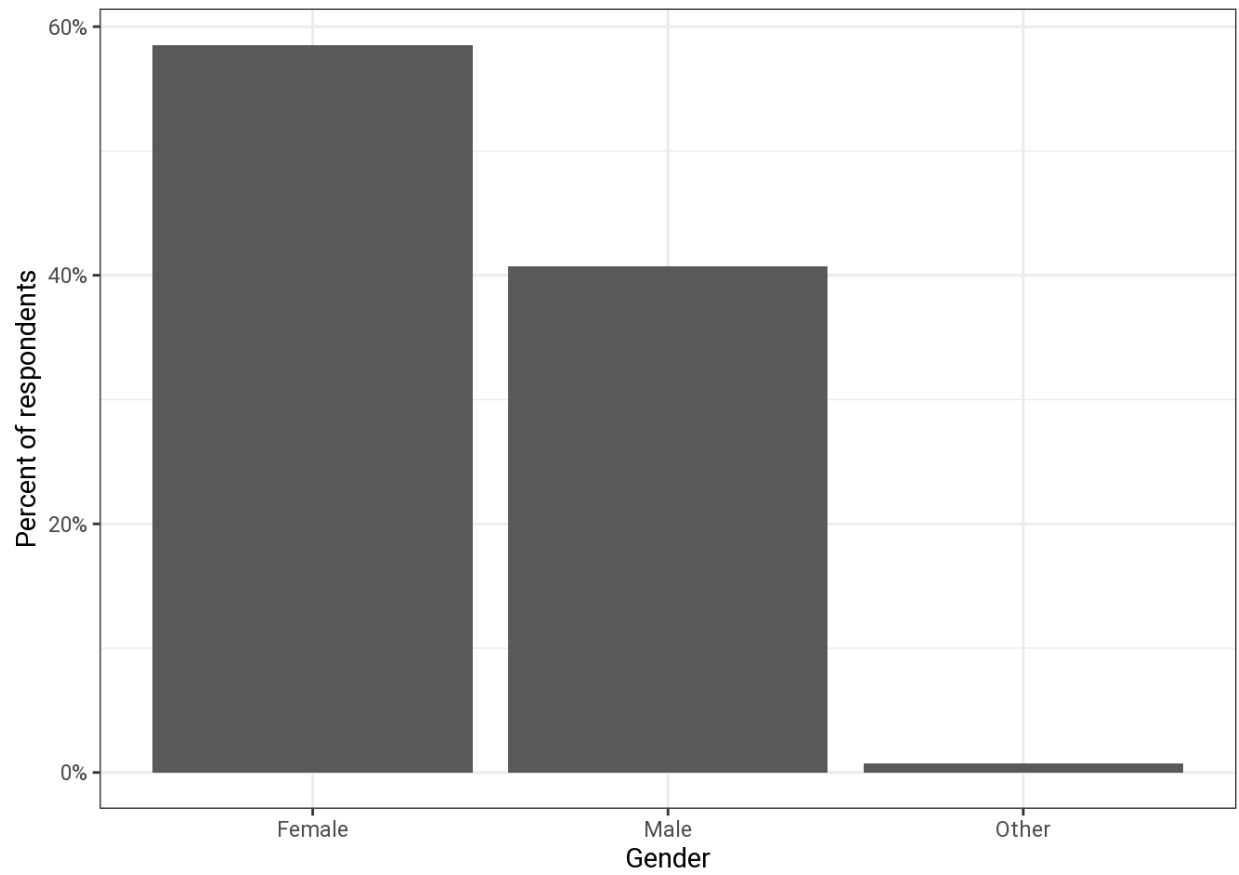


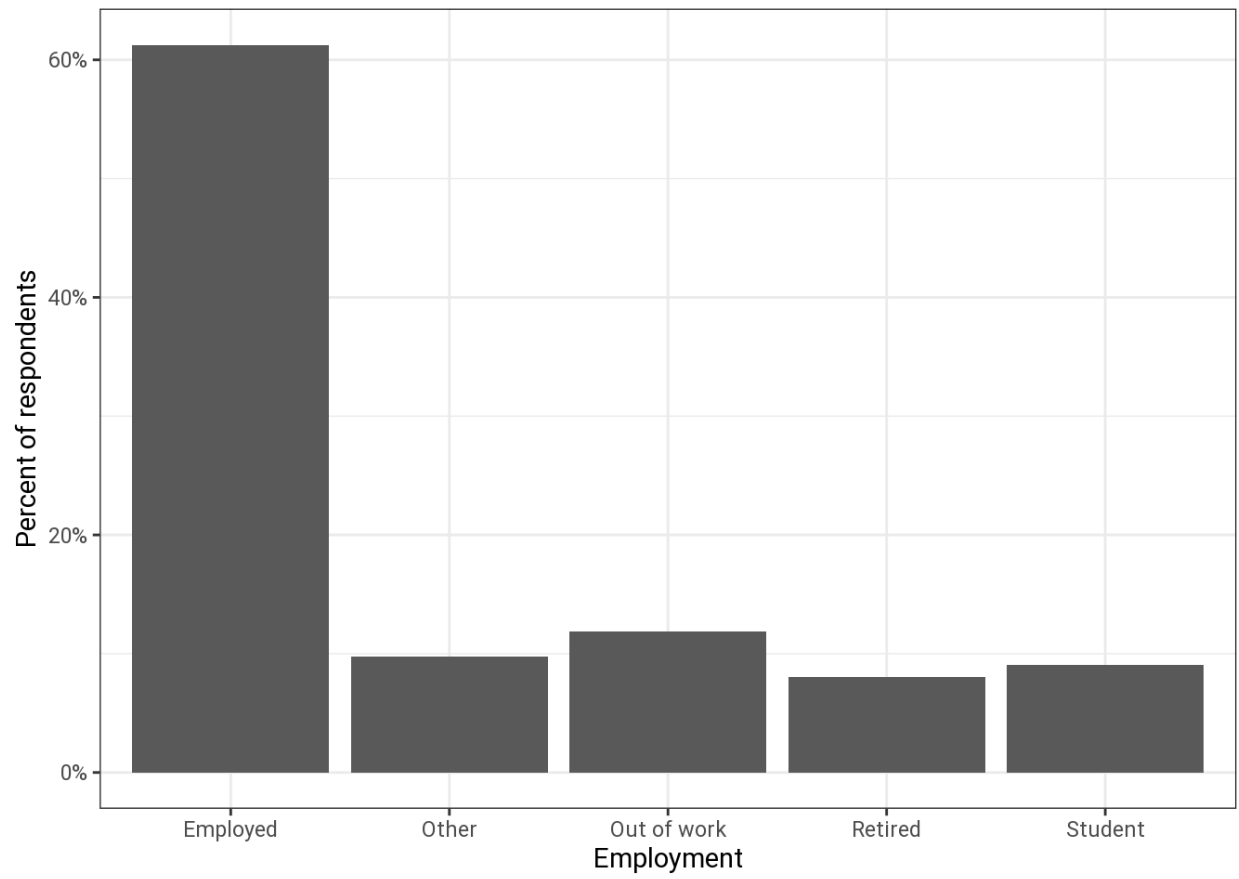
Figure 16: Weights for UX metric's contribution to the rank score (the weights are the estimated regression coefficients from fitting PLS regression with the rank scores and UX metrics).

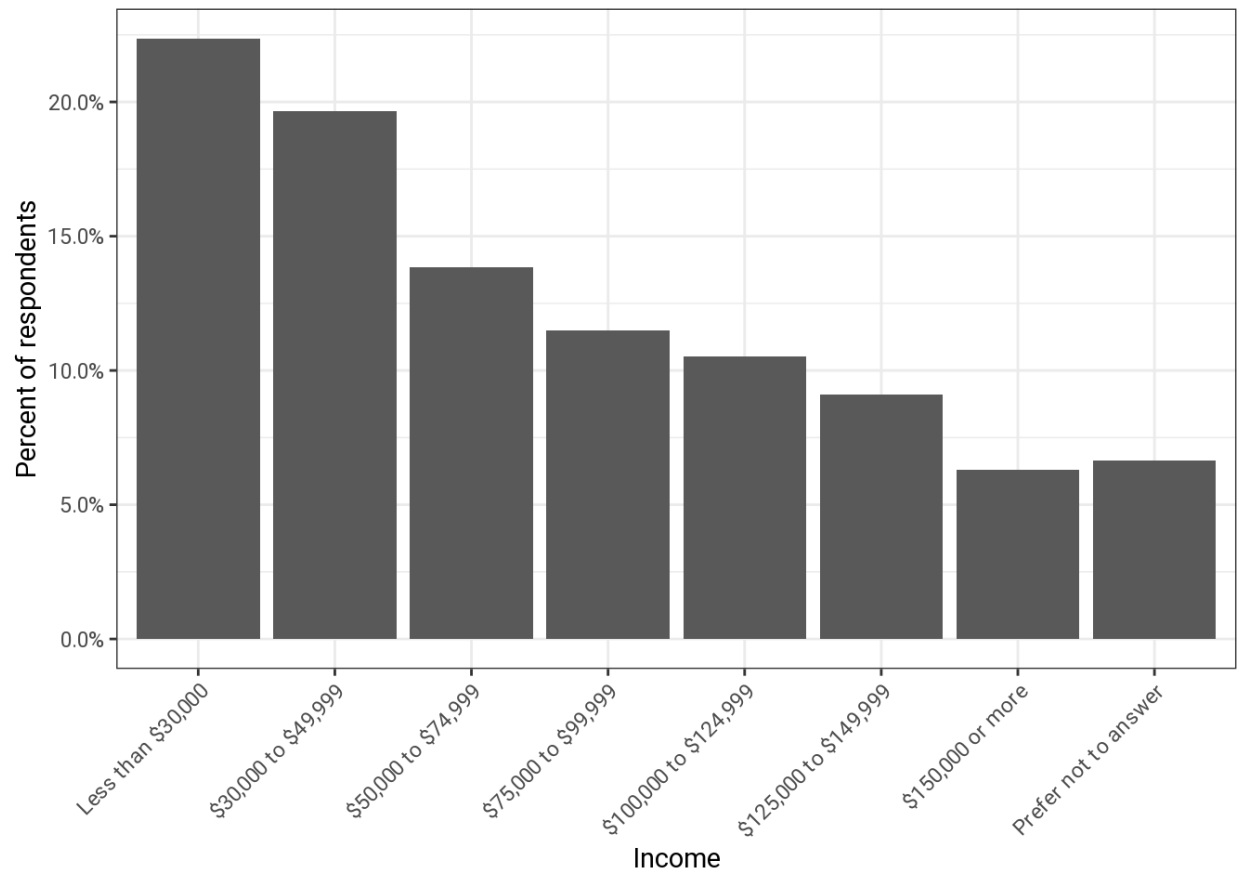
The figure above shows that “annoying”, “distracting”, “misleading” and “appropriate” are the four negative contributors. On the other hand, “trustworthy”, “useful” and “pleasing” contribute positively to the score, which is expected. What is interesting is that “relevant” does not have much impact on the rank score. This result confirms that the rank score correlates well with the UX metrics and is a good single metric to rely on.

Demographic Distribution



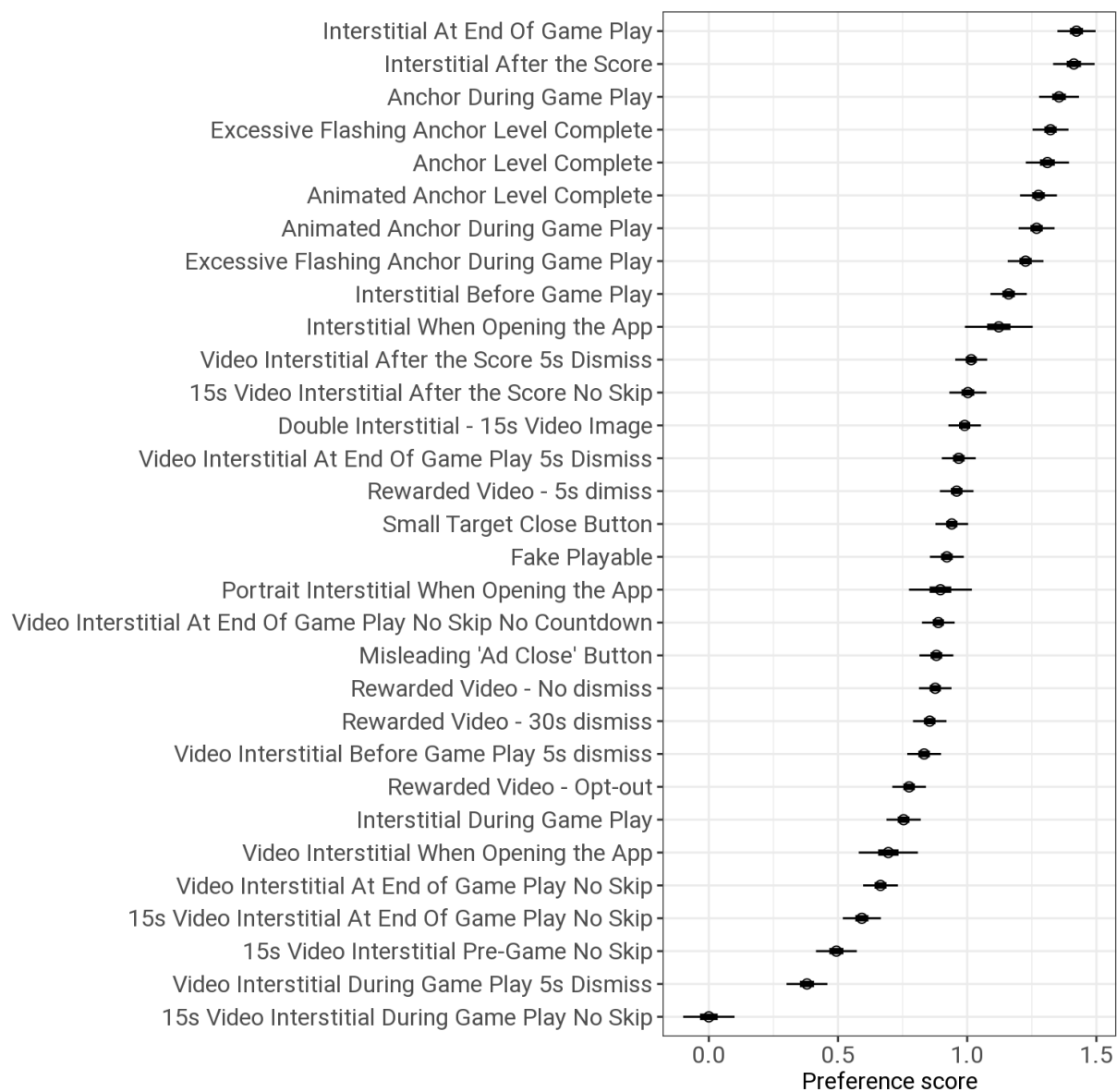




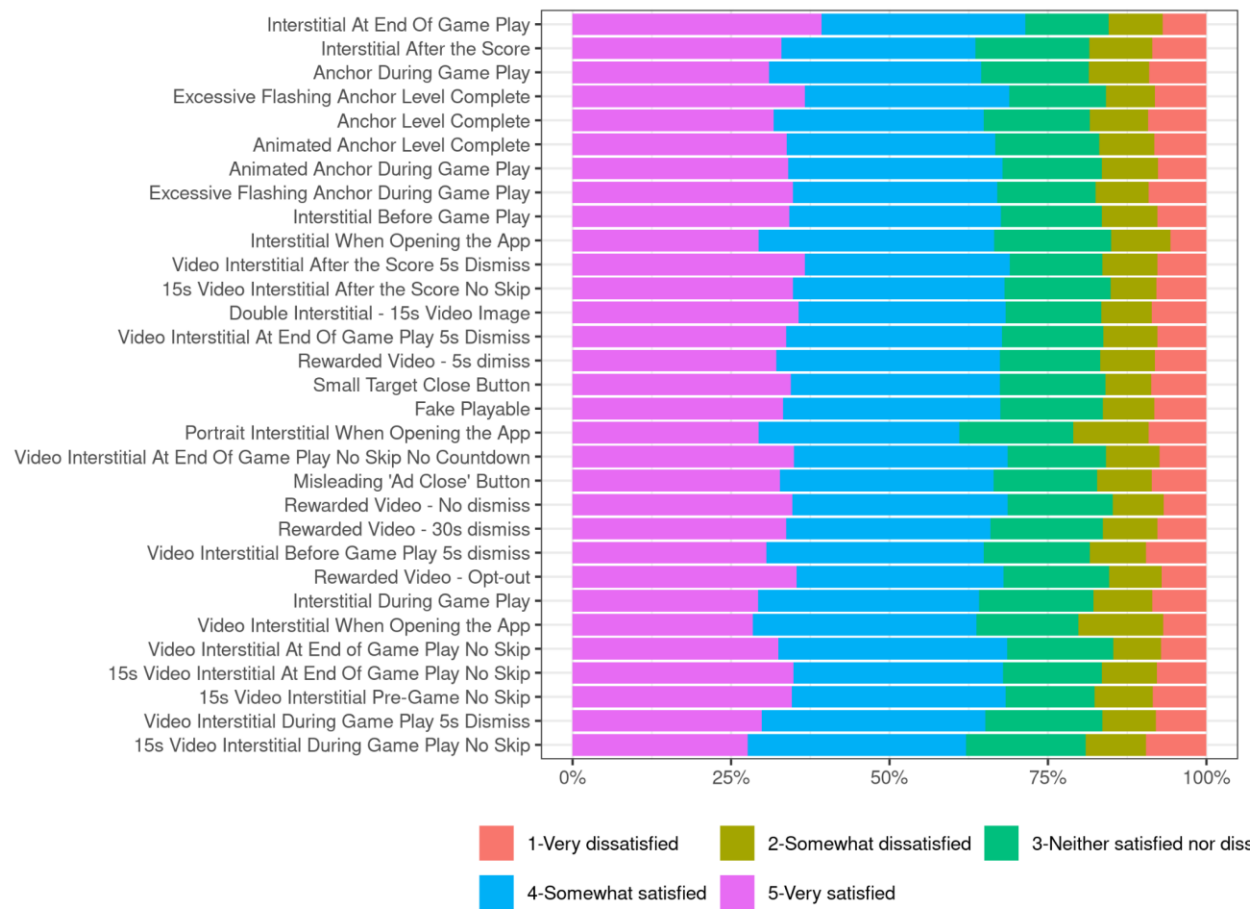


Full Results

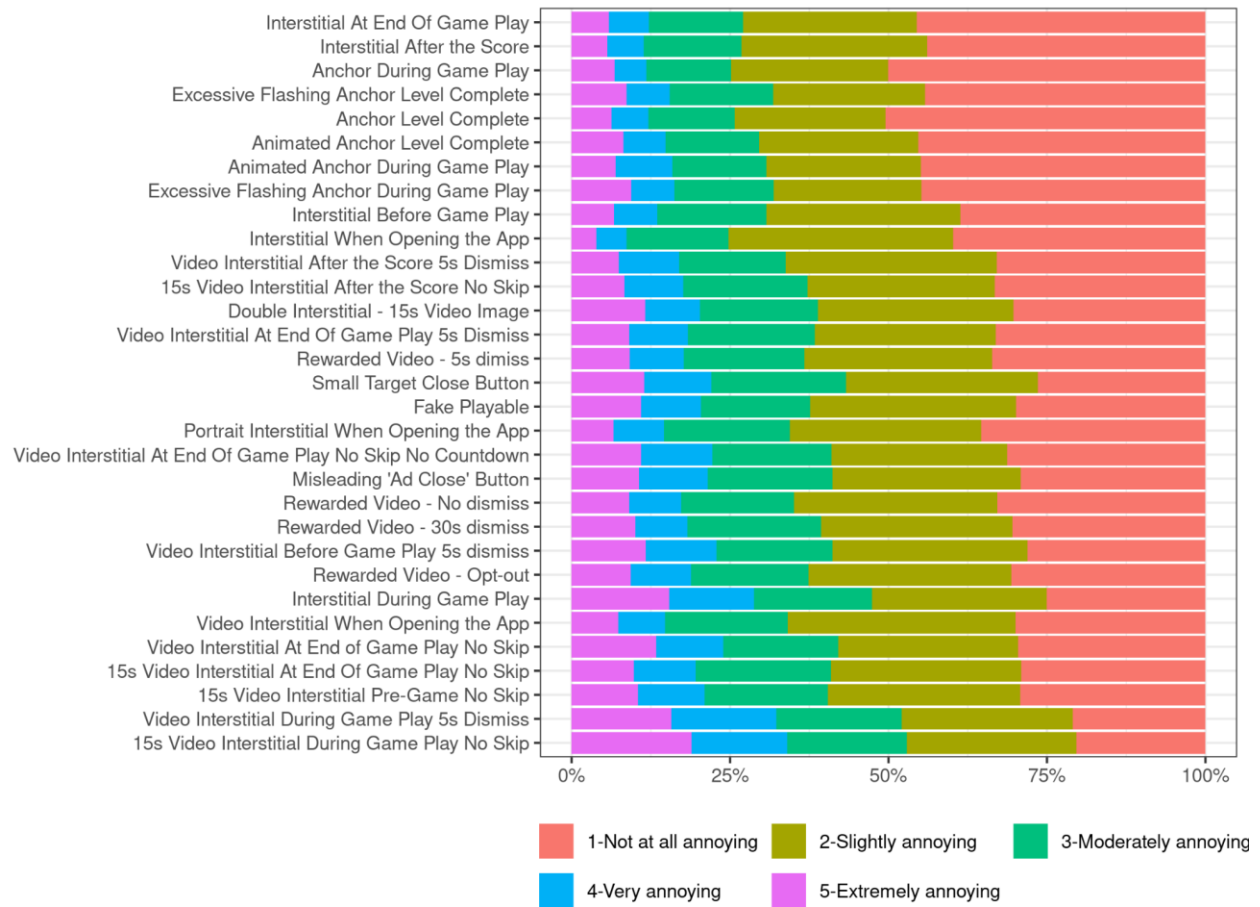
Method: BT



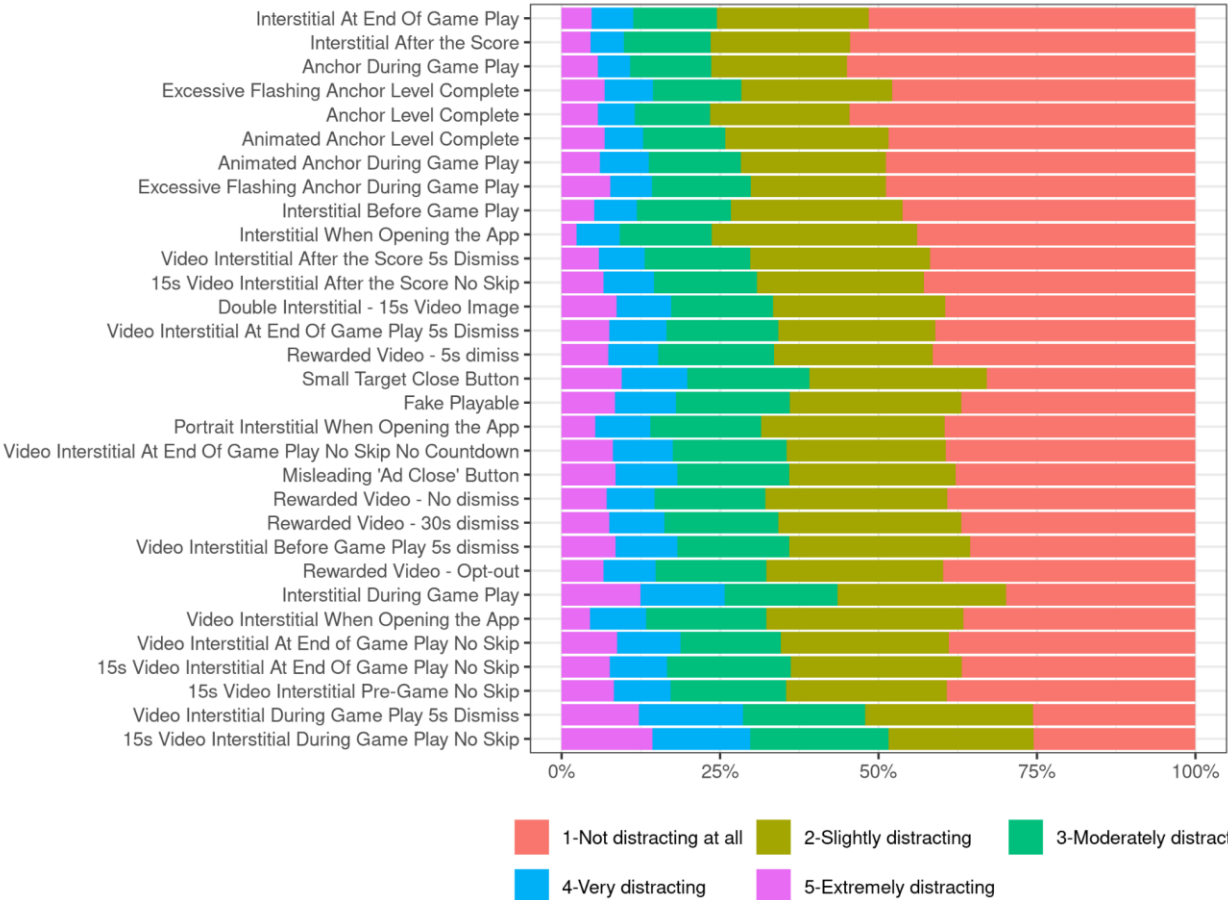
Satisfaction - Overall Experience



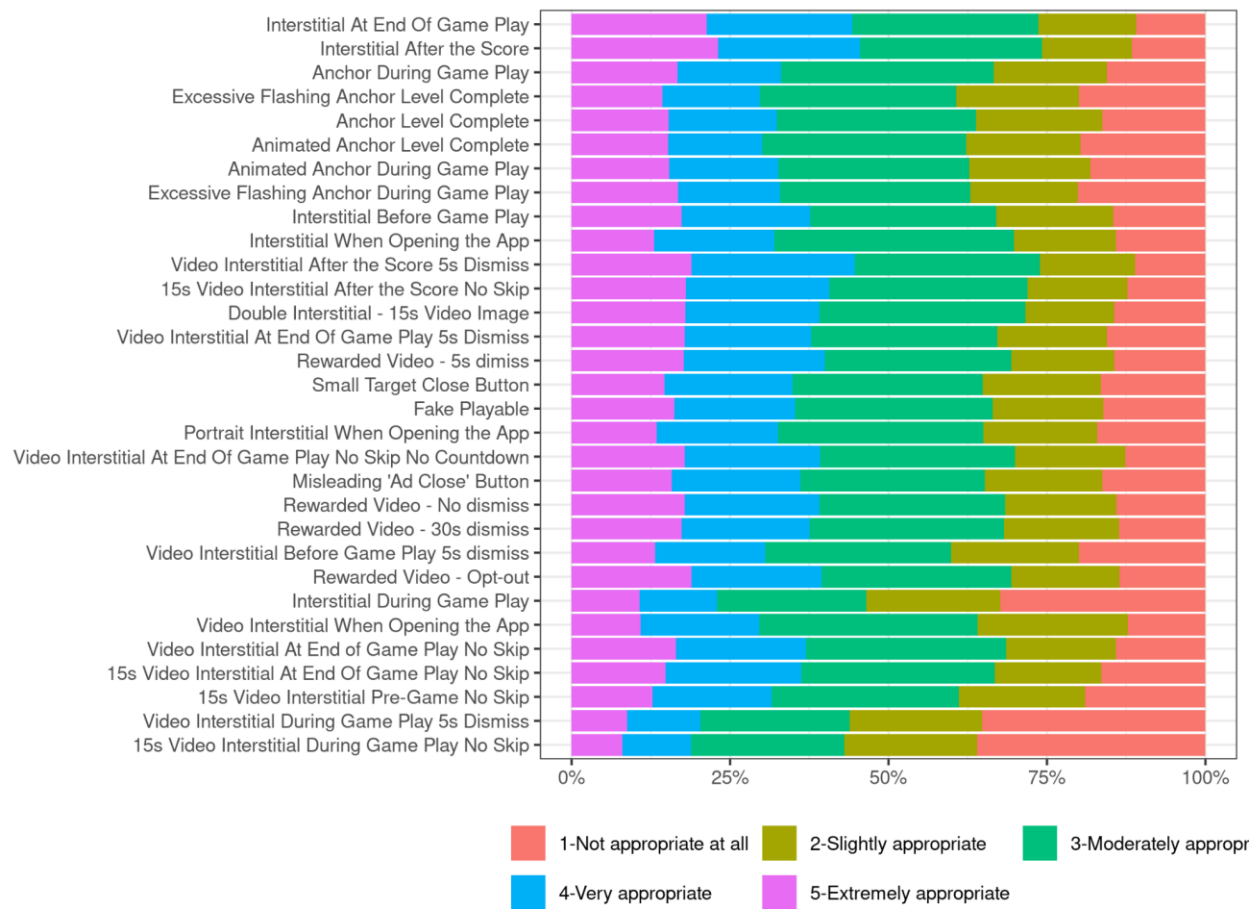
Annoyance



Distracting



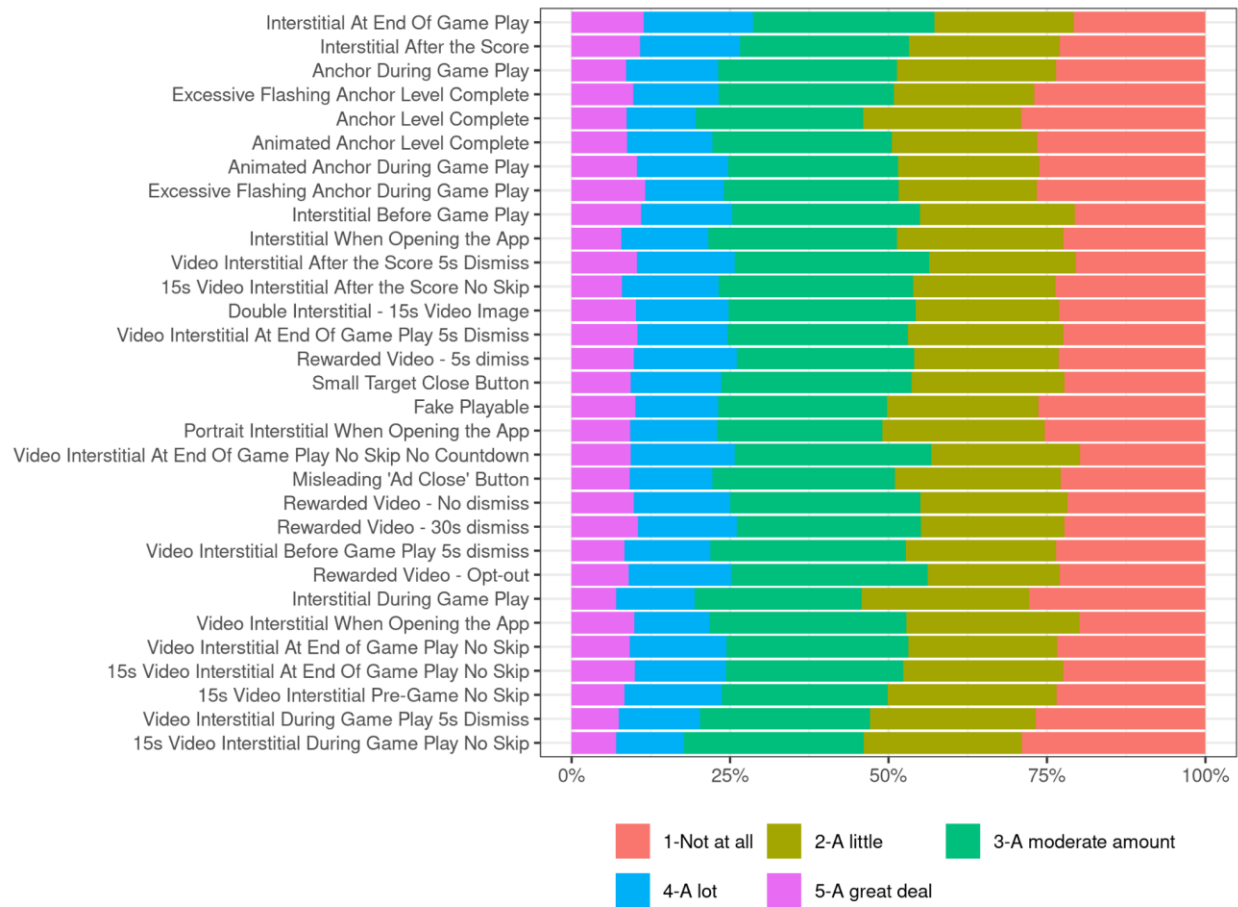
Appropriate Timing



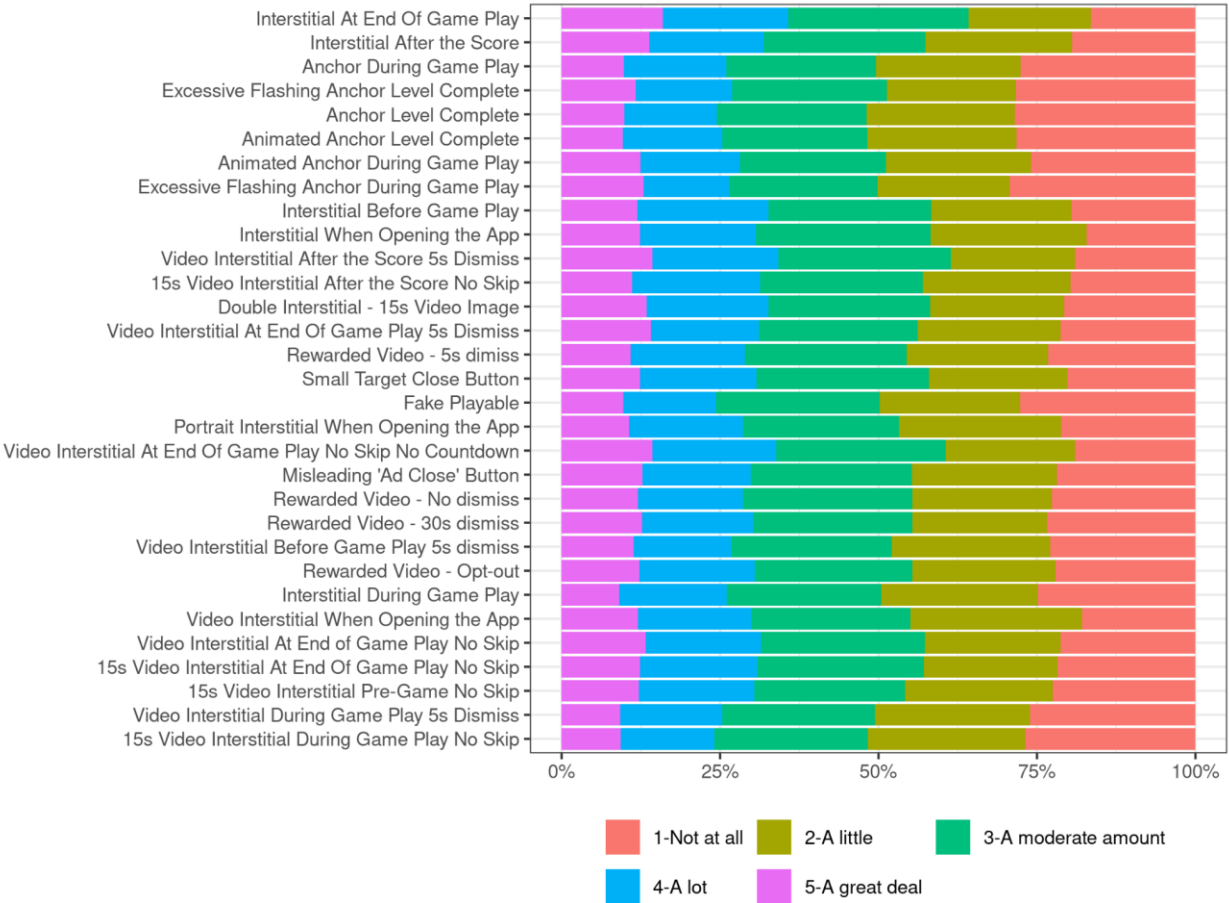
Useful



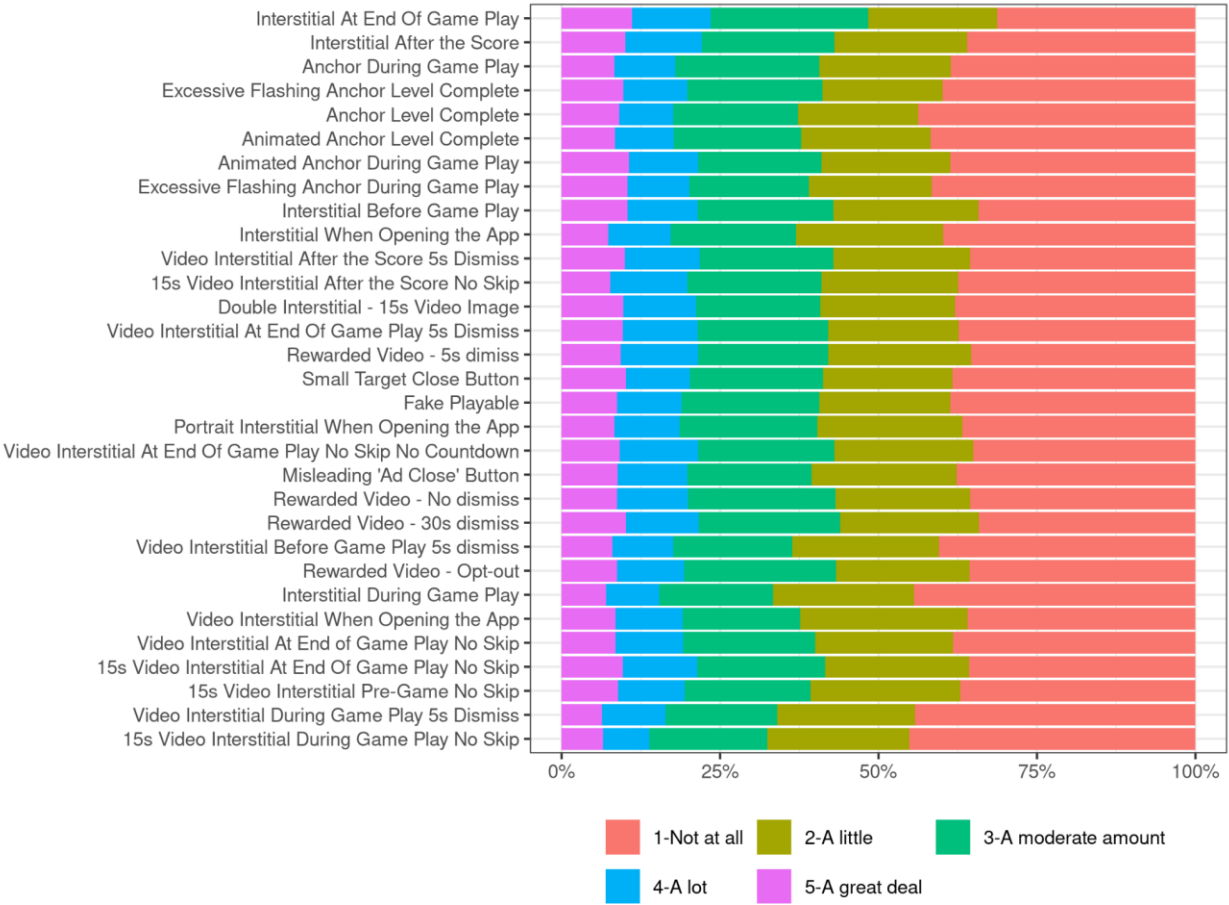
Trustworthy



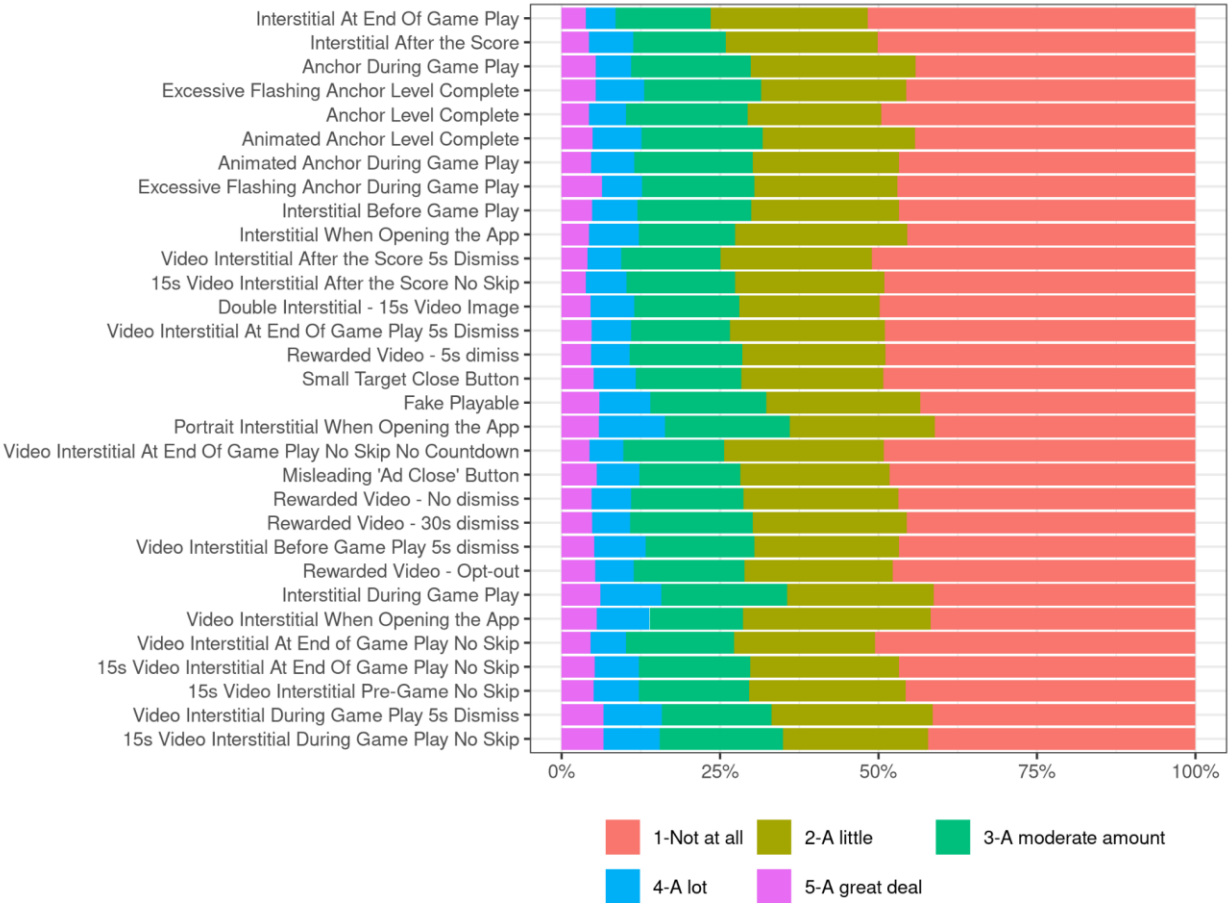
Visually Pleasing



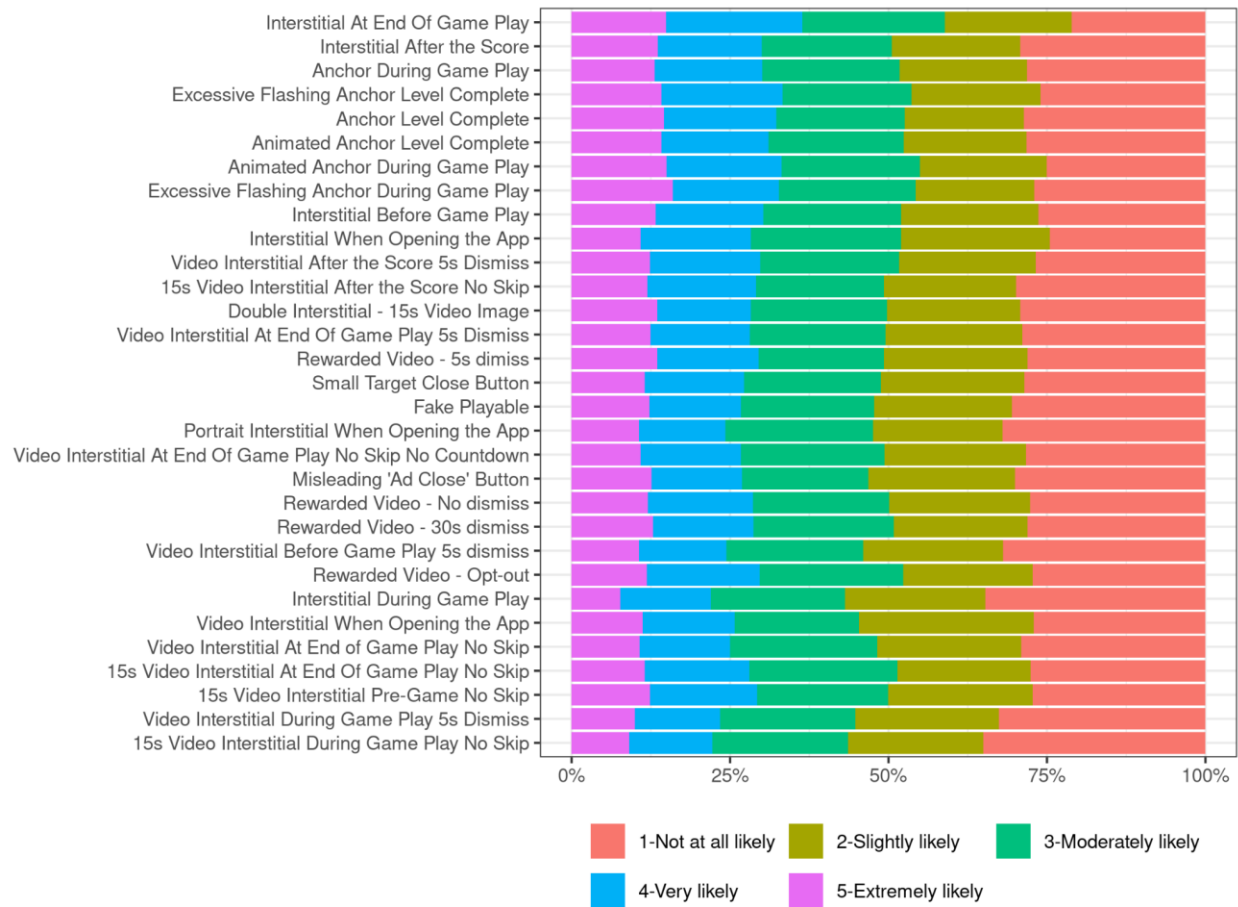
Relevant



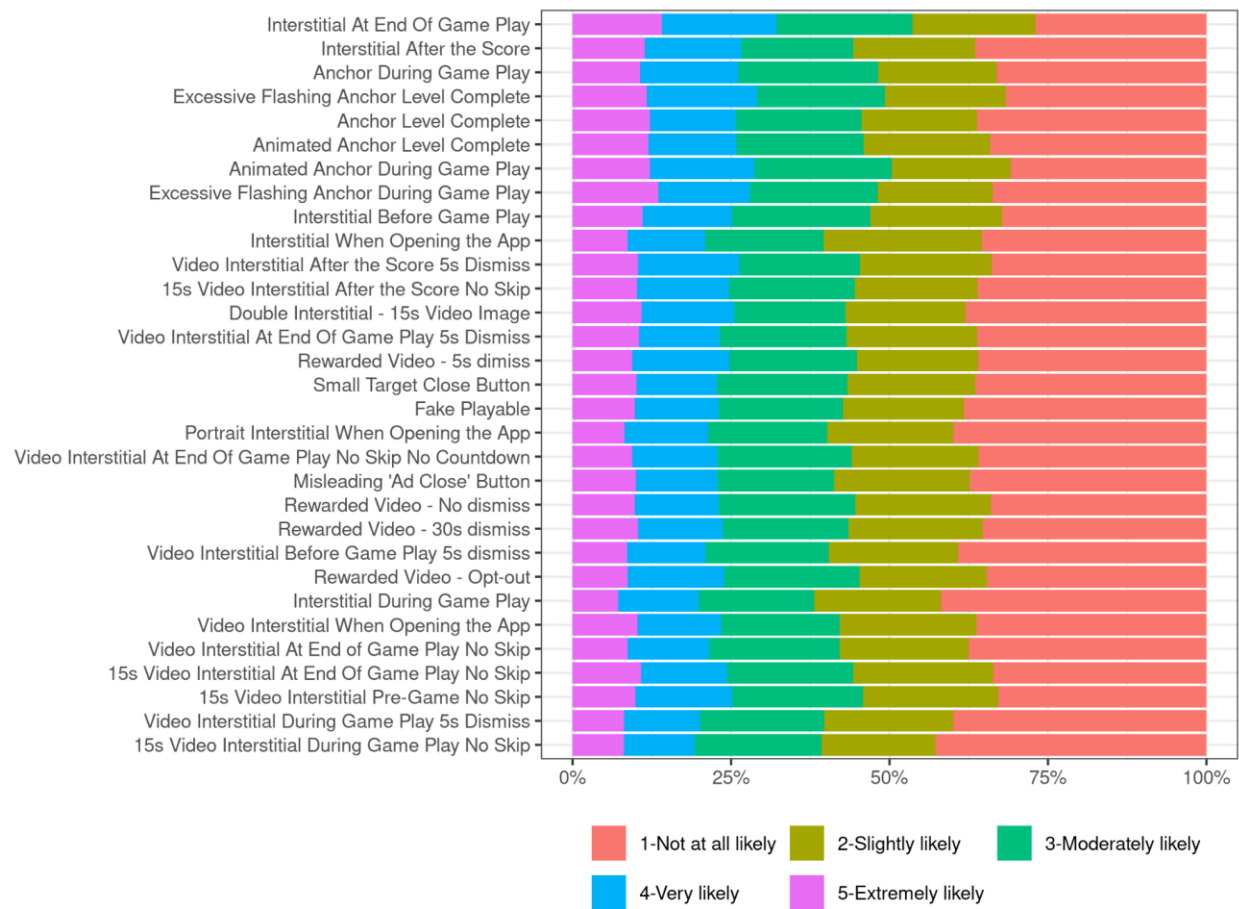
Misleading



Likelihood to Play



Likelihood to Share



Likelihood to Block

