



In-View Ad-Refresh Experiment

Measuring the effect of expanded in-view ad refresh allowances on ad-blocking user behaviour and publisher revenue

Date: August 2026
Performed by: eyeo GmbH

Executive Summary

eyeo conducted a controlled A/B experiment from 1 May to 1 June 2026 to evaluate whether relaxing restrictions on in-view ad refresh affects the browsing experience of ad-blocking users. Approximately 9.9 million users were assigned to each of five groups: a control group receiving the current AA-compliant refresh experience, and four experimental groups each testing a different relaxation of the current Standard:

- **Group A — After Unfilled:** refresh triggered after an unsuccessful ad bid
- **Group B — Any Size:** the refreshed ad may be a different size from the original
- **Group C — All In-View:** all in-view ad units refresh, not just one
- **Group D — All Features:** all of the above combined

After more than 320MM aggregate page views, there was no measurable increase in negative user experience in any of the experimental groups. The rate of users uninstalling their ad blocker or opting out of seeing Acceptable Ads altogether were flat or marginally lower compared to the control group, none at a level that would indicate a genuine effect. Page views per session were also marginally lower across all groups, but within a range consistent with normal variation. A meaningful drop in page views per session could suggest users are leaving pages sooner in response to a disruptive ad experience; no such pattern was observed.

Groups C and D delivered meaningful revenue gains for publishers (+4.5% and +16.5% respectively); Groups A and B did not clear the threshold for a meaningful revenue difference. Group D, combining all features, was the strongest performing configuration.

Group D offers the best balance to AAC Stakeholders, demonstrating the highest meaningful increase to publisher ad revenue by providing more opportunities for successful bids while having no detectable negative impact on user experience. Ongoing monitoring of user experience over an extended period of time is warranted.

Table of Contents

[Executive Summary](#)

[Table of Contents](#)

[1. Introduction](#)

[2. Background](#)

[3. Research Method](#)

[4. A/B Experiment Design](#)

[5. Data Reliability](#)

[6. Results](#)

[7. Summary of Findings](#)

[8. Comparison with 2022 Survey Findings](#)

[9. Limitations](#)

[10. Conclusion and Recommendations](#)

[11. Next Steps](#)

1. Introduction

In 2023, the AAC approved a change to the Acceptable Ads Standard to permit specific implementations of in-view ad refresh, based on a commissioned user survey of 7,427 ad-blocking users across the US, Germany, and France. As with previous user survey studies, only a restricted set of ad-refresh variations could be tested owing to the complexities of study design, and those that were tested were based on mock-up environments that lack external validity. Given these limitations, and the constraints currently placed on publishers to implement ad-refresh in adherence to the AA Standard, publishers have asked whether these constraints can be relaxed without negatively affecting users.

This report presents the results of an A/B test evaluating four expanded ad refresh feature sets against the current AA Standard, conducted on ABP non-premium users across Blockthrough's publisher network (May-June 2026). An A/B test enables causal measurement of how ad implementations affect real user behaviour at scale, including uninstall rates, AA opt-outs, and other engagement signals, in ways that controlled surveys cannot replicate. Critically, it also enables simultaneous measurement of publisher ad revenue, providing the AAC with evidence of balance for both users *and* publishers. Publishers are a key stakeholder for the AAC, and evaluating whether a proposed change

delivers measurable value for publishers, alongside its acceptability for users, strengthens the basis for any AA Standard decision.

The results of the A/B test is a preliminary communication to the AAC under the Research Framework. Findings from the 2022 survey are drawn on where conditions are directly comparable. More broadly, this test demonstrates how behavioural data can help the AAC evaluate whether conditions balance user experience and publisher benefit, and may inform a principle that ad experiences are only added to the AA Standard where they show measurable value for publishers as well as acceptability for users.

2. Background

2.1 Current AA Standard provisions for in-view ad refresh

The Acceptable Ads Standard currently permits in-view ad refresh under the following conditions:

- One ad unit per page may refresh
- The refresh request must return the same ad size as the original
- Placement is restricted to the side rail only
- Refresh may not occur before 30 seconds of continuous in-view time on scrollable pages, or 60 seconds on non-scrollable pages

2.2 Untested variations

The 2022 survey could only test a finite number of conditions. The following implementation variations fall outside the scope of what was evaluated:

- Refreshing more than one in-view ad unit on a page simultaneously
- Allowing the refresh request to return a different ad size from the original
- Triggering refresh based on non-time-based events (e.g., a failed bid, a user action, or page engagement signals)
- Combinations of the above

2.3 Rationale for further research into untested implementations

At first glance, some of the above variations appear to be in tension with the direction of findings from the 2022 survey, particularly multi-unit refresh and size-flexible refresh. However, there are principled reasons to evaluate these behaviourally rather than treating the 2022 findings as conclusive.

- The 2022 study's multi-ad findings were not a clean test of multi-unit refresh at a single placement. Conditions involving multiple ads always paired a side-rail unit with a top-banner placement, confounding unit count with placement type. It is not established that refreshing multiple side-rail units simultaneously would produce the same effect.
- The constraint to allow only a single-unit ad-refresh reflected a precautionary interpretation rather than a clear case for exclusion. Concerns about multiple ads arose in a particular page context (non-scrollable) involving a combination of placements that may not reflect the broad range of scenarios under consideration. The 2022 report noted that testing ad refresh in combination with other ads "will help extend these findings". An A/B test is well-placed to evaluate whether that caution was warranted in practice.
- The 2022 survey tested whether users *perceived* different ad formats as annoying in a mock environment, not whether they *acted* on that annoyance in real browsing. Self-reported disruptiveness and behavioural churn are distinct outcomes: users may tolerate implementations that they would rate negatively in the abstract, particularly when engaged with real content. Behavioural data from a live test provides an ecologically valid measure of whether any effect is material in practice.

Finally, the AA Standard exists to serve multiple stakeholders (users, publishers, and advertisers) and any proposed change should also demonstrate value to publishers without imposing a cost on users. **A behavioural experiment allows for both dimensions to be measured simultaneously: if the data show a revenue uplift without user harm, the case for amending the AA Standard is strong; if user harm is detected, the commercial case is irrelevant.**

3. Research Method

In an A/B test, users are randomly assigned to experimental groups that each vary one or more features of the ad experience. Each group in this test isolates a specific modification to what the AA Standard already permits: extending the number of refreshing units, allowing size variation on refresh, or enabling a non-time-based trigger. The behavioural effect of each change can be measured causally against a control group receiving the current AA Standard experience.

According to the Research Framework, A/B testing is appropriate for this kind of research question as it is clearly defined and the aim is to evaluate a specific implementation in a real-world environment.

4. A/B Experiment Design

4.1 Test conditions

The experiment evaluated five groups, each isolating one or more refresh behaviours relative to the current AA Standard.

Group	Code	Behaviour tested	Requires AAC approval
Control	refresh-control	Today's AA-compliant refresh: one in-view unit, same size, limited placement	—
A — After Unfilled ¹	refresh-unfilled	Refresh a slot after it returns unfilled. There would be delay in the loading of the ad unit (e.g., after a 30 second interval).	No*
B — Any Size	refresh-multisize	Refresh request may return a different ad size. All ad units	Yes

¹ Group A was included in the experiment because it was a technical detail for ad-refresh that had been overlooked when specifying its implementation. Rather than assume it had a minor impact on user experience, we wanted to test it in isolation (to compare to the control) and include it in Group D as part of the 'All Features' implementation. Including it as a comparison group helped to validate that the experience itself was not detrimental while characterising its revenue impact.

		remained within the AA Standard compliant sizes.	
C — All In-View	refresh-all-inview	There was no limit to the number of ad units that could refresh at a time, if triggered. All ad units could refresh. ² In this group, ad refresh units were not restricted to the side rail.	Yes
D — All Features	refresh-all	A + B + C combined. All three of the restrictions that were lifted for each of the Groups A, B, and C were allowed in Group D. In practice, this means that any combination (or none) of these conditions could be shown depending on the auction.	Yes

Consistent with the current AA Standard, ad-refresh could occur no earlier than 30 seconds for scrollable pages and 60 seconds for non-scrollable pages. All other Acceptable Ads requirements relating to size, coverage, placement, animation, and ad density continued to apply.

4.2 Test Overview

The test ran from 1 May to 1 June 2026 (31 days) across ABP non-premium users on Blockthrough publishers running BT Focus. Each group received approximately 20% of eligible traffic (~9.9M users per group). Users were allocated randomly to test groups. No user characteristics, subgroups or quotas were applied, with the exception that Premium users were not part of the experiment.

Results are assessed against a significance threshold of $p < 0.05$, meaning a difference between groups is only treated as real if there is less than a 5% probability it occurred by

² The typical behaviour is that a page's ad units get auctioned at once when the page first loads, so they usually end up with the same cadence and refresh simultaneously. However, there is no guarantee that all will get triggered at once – or that they actually get refreshed at all, depending on whether or not compliant bids were received for the units.

chance alone. For each key metric, a minimum meaningful effect size was established before the test began: the smallest change that would be considered practically significant rather than statistical noise. This step is necessary because at the scale of this test (~9.9M users per group), even extremely small differences will register as statistically significant. Without a pre-specified threshold, the analysis would treat trivial fluctuations as meaningful findings.

The thresholds were set as follows:

- **Uninstall rate: 1.6% relative change.** At the observed control uninstall rate of 0.734%, this corresponds to approximately 1,200 additional uninstalls per group: the minimum increase considered large enough to indicate a genuine effect on user retention rather than random variation.
- **AA opt-out rate: 0.008 percentage points.** AA opt-outs are rare events; across ~9.9M users this threshold corresponds to approximately 800 additional opt-outs: the minimum increase considered a meaningful signal of user dissatisfaction with the Acceptable Ads experience.
- **Page views/Page views per Blockthrough session:** On average of 2.0 pageviews per session for every Blockthrough site visit, a meaningful change would see this drop by 4% relative.
- **Revenue: 4% relative change.** A smaller revenue difference would not be reliably distinguishable from normal day-to-day fluctuation in publisher earnings.

A result that is statistically significant but falls within its threshold is treated as noise, not a genuine effect. Group B (Any Size) illustrates this: its revenue lift of +2.1% was statistically significant but fell below the 4% threshold and is therefore not treated as a meaningful finding.

Two measurement systems feed into this report, covering different user populations. Total user counts are drawn from eyeo's central data platform (covering 100% of users, with a small approximation error of ~0.5%). Revenue and opt-out events are logged directly from the extension but cover a 10% sample of users. Some figures therefore differ in their population base between sections; both systems tend to understate rather than inflate any treatment difference.

4.3 Outcome metrics

Metrics span strong and moderate behavioural cues plus revenue, per the Research Framework. Weak cues and additional strong cues are planned for the follow-on phase (Section 11).

Strong behavioural cues

If any of these exceed a pre-specified threshold relative to control, there is ***strong evidence*** of a negative effect on user experience.

Metric	Data source
Extension uninstall count and rate (per group, per day)	Extensions / websites
AA opt-out count and rate (per group, per day)	Extensions
AA-related uninstall and opt-out reasons (e.g., "I'm still seeing ads"; "I don't like the Acceptable Ads programme")	Extensions / websites
Ad-related issue report submissions	Extensions

Moderate behavioural cues

If exceeded, there is ***moderate evidence*** of a negative effect.

Metric	Data source
Pageviews per Blockthrough session	Blockthrough

Page views per session measures how many pages a user visits within a single browsing session. A lower value in a treatment group relative to control would suggest users are navigating away from pages or ending their sessions sooner, signalling disruption to the browsing experience. A similar or unchanged value indicates the ad experience did not affect how deeply users engaged with content.

Revenue metric

Metric	Data source
USD normalised revenue per variant	Blockthrough

5. Data Reliability

5.1 Traffic balance

Each group received approximately 20% of eligible traffic (~9.9M users per group), with groups balanced to within ~0.1% of one another via eyeo's central data platform. No meaningful imbalance between groups was detected, giving confidence that any differences in outcomes reflect the ad experience tested, not pre-existing differences between groups.

5.2 Treatment fidelity

Extensive pre-experiment testing ensured that the randomisation worked correctly across data sources (extensions data and Blockthrough data) and that users in each cohort saw the correct ad-refresh behaviour. Auction data increased in accordance with expectations for each respective study group, suggesting users in these groups did experience the ad-refresh behaviour.

5.3 Analysis approach: Intention-to-treat

Users are analysed in the group they were assigned to, regardless of whether they actually encountered the refreshed ad during the test window. This is the methodologically correct approach as it is least subject to bias. However, it means the true effect on users who did encounter the new experiences could likely be somewhat larger than the figures reported here.

5.4 Revenue data note

Revenue figures in this report use assigned-user ARPU, defined as total revenue divided by all users assigned to a group, with users who generated no revenue counted as zero. This is the correct basis for comparison.

5.5 Issue reports

Issue reports are counted as distinct reporters — unique extension installations that submitted at least one ad-related report — rather than as raw report submissions. This reflects what the metric is meant to measure: how many users encounter a problem, not how many times the report control is pressed. Raw submission counts are vulnerable to

two distortions that do not reflect the breadth of user friction. First, a single user can submit the same complaint repeatedly; in this dataset one group's raw count was inflated by a single user submitting a near-identical complaint over a hundred times on one domain. Second, automated or non-representative submissions can cluster on atypical domains; a small set of reports with no comment or description appeared on otherwise-unseen domains. Counting distinct reporters neutralises both and gives a per-user measure consistent with the user-experience framing applied to the other behavioural cues.

6. Results

No ad-refresh experience shows a deterioration against the control group on any metric. Reading the strong behavioural cues together: if ad-refresh were degrading the user experience, we would expect more uninstalls, more opt-outs, or shorter session lengths on pages where the ad-refresh experience could be served. None of those metrics differed from the control group. The user experience is indistinguishable from the current AA Standard compliant ad-refresh experience across all groups.

General summary:

Group	Users (CDP)	Uninstall rate %	AA opt-out rate %	Issue reports per 1M users, distinct
Control	9904810	0.734	0.0102	41.8
After Unfilled	9909977	0.735	0.0104	41.48
Any Size	9906125	0.728	0.0098	42.1
All In-View	9910944	0.734	0.0111	42.09
All Features	9907138	0.727	0.0108	42.71

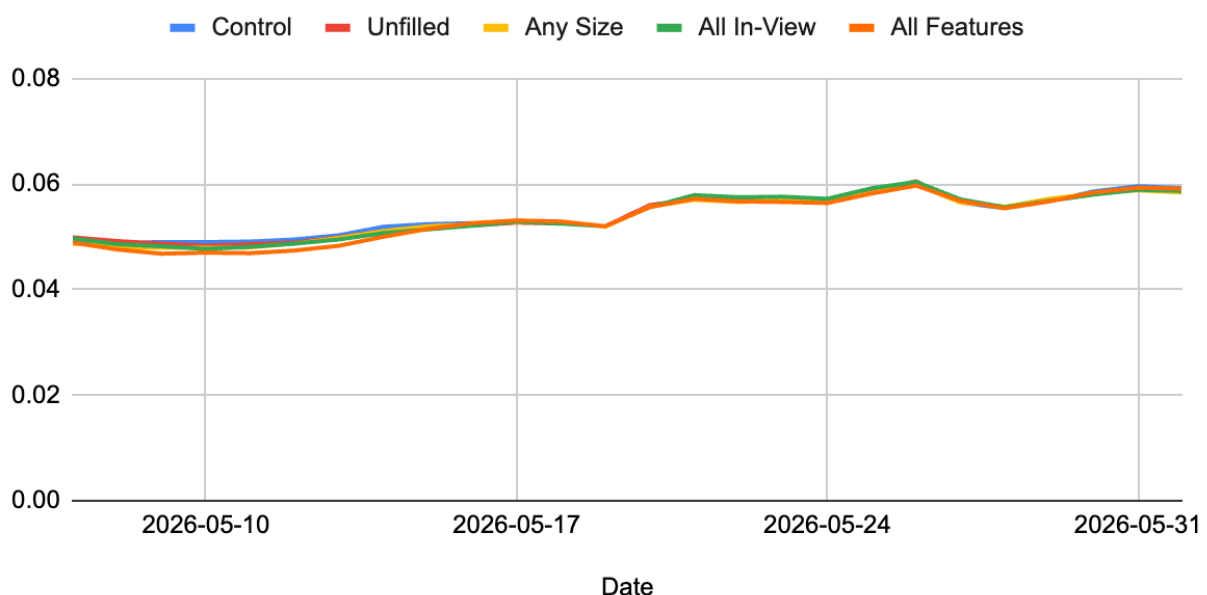
6.1 Strong behavioural cues

Uninstall rate

Uninstall rates were flat to slightly lower across all treatment groups. No group showed a statistically significant increase relative to the control (current AA Standard).

Group	Uninstall rate	vs. control	p-value
Control	0.734%	—	—
A — After Unfilled	0.735%	+0.06%	0.91
B — Any Size	0.728%	−0.82%	0.12
C — All In-View	0.734%	−0.09%	0.87
D — All Features	0.727%	−0.98%	0.061

Uninstall per-day rate: Control, After Unfilled, Any Size, All In-View and All Features

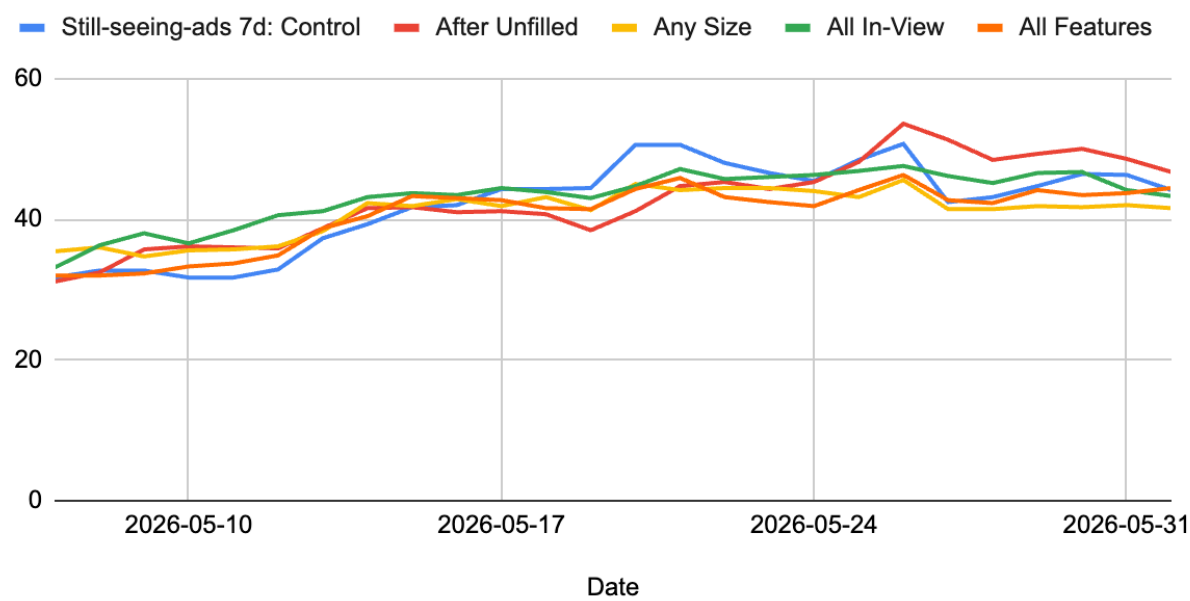


AA-related uninstall reasons

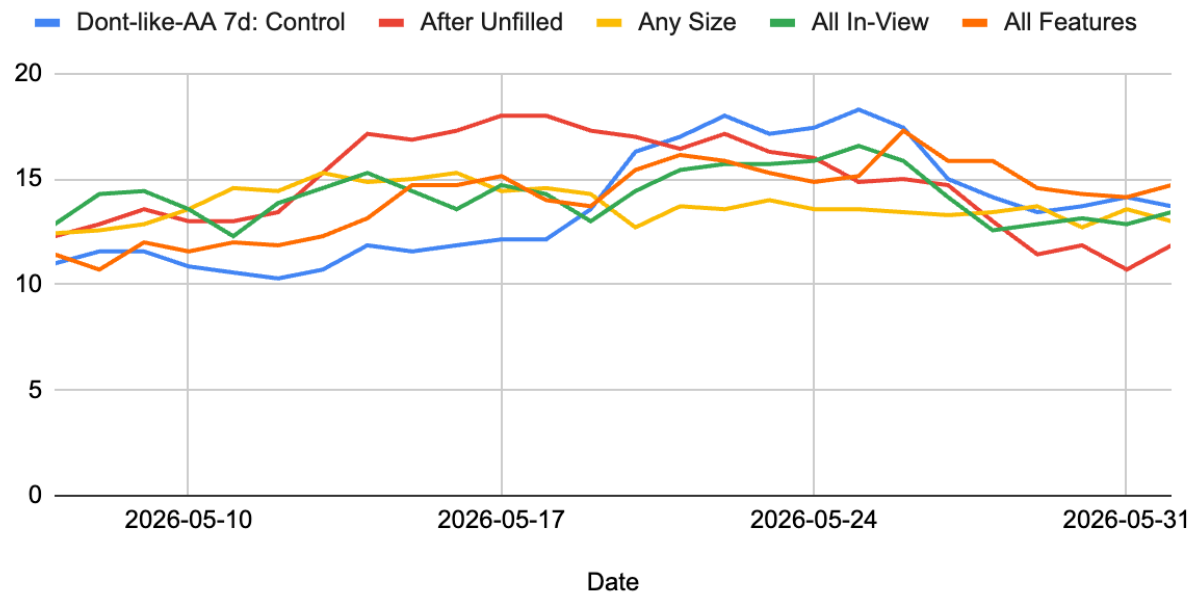
When users uninstall Adblock Plus, they can optionally indicate the reason they uninstalled the extension. Two of those responses relate to AA: “still seeing ads” and “I don’t like Acceptable Ads”. As can be seen in the Table and figures below, there were no differences between the ad-refresh conditions in the reasons that users uninstalled the extension.

Group	Still seeing ads (n)	Still seeing ads (%)	p-value (vs. control)	Don't like AA (n)	Don't like AA (%)	p-value (vs. control)
Control	1302	1.79		428	0.59	
After Unfilled	1336	1.84	0.523	456	0.63	0.354
Any Size	1297	1.8	0.913	428	0.59	0.905
All In-View	1337	1.84	0.487	451	0.62	0.434
All Features	1287	1.79	0.959	437	0.61	0.654

Still-seeing-ads 7d: Control, After Unfilled, Any Size, All In-View and All Features



Dont-like-AA 7d: Control, After Unfilled, Any Size, All In-View and All Features



AA opt-out rates

Opt-out events:

No group showed a statistically significant change in AA opt-out events relative to control. Group C (All In-View) was directionally higher (7.9%, $p = 0.081$) in events, below the significance threshold, but noted here for transparency and worth monitoring at rollout.

Group	Opt-outs	Rate per 1M	vs Control	p-value
Control	1018	102.78		
After Unfilled	1044	105.35	+2.5%	0.575
Any Size	962	97.11	-5.50%	0.207
All In-View	1099	110.89	+7.9%	0.081
All Features	1067	107.7	+4.8%	0.286

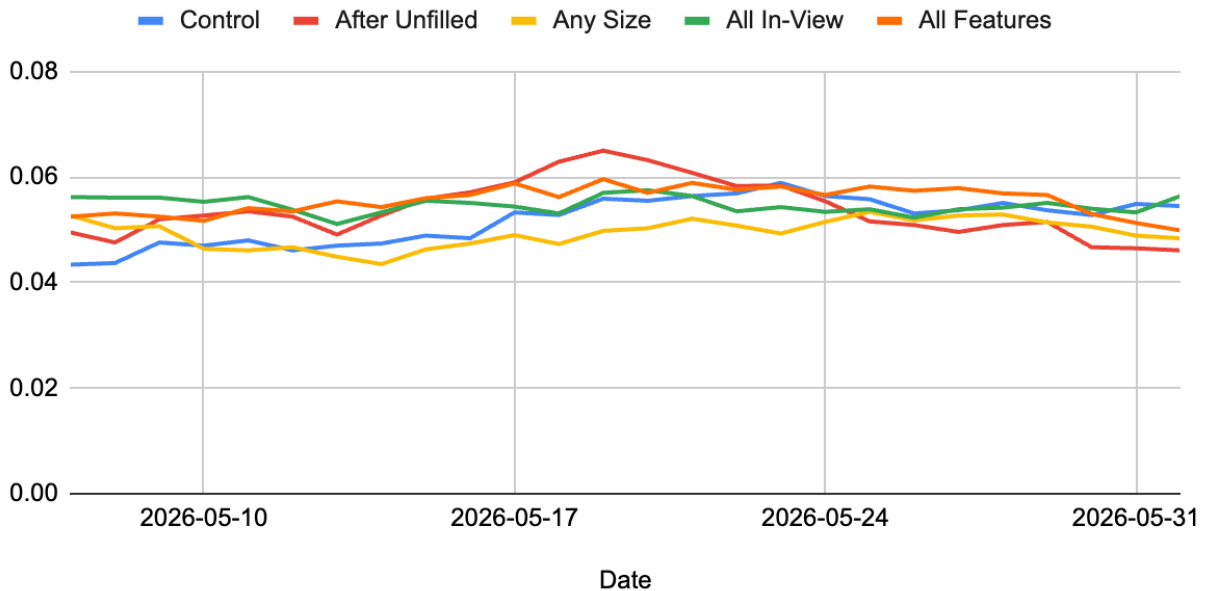
Opt-out rate overall:

We also tracked the overall opt-out rate between the experiments during the trial:

Variant Flag	AA Active Rate	% Diff from Control*	Significant ($\alpha=0.05$)
Control	94.89%	0	0.2432
After Unfilled	94.88%	-0.01%	0.026
Any Size	94.88%	—	—
All In-View	94.88%	0	0.524
All Features	94.88%	-0.01%	0.0056

*note, differences refer to rounding

AA Opt-out per day rate: Control, After Unfilled, Any Size, All In-View and All Features



AA-related opt-out reasons

It was not possible to connect the AA opt-out survey responses page to the A/B test in this test experience, so we are not able to evaluate AA opt-out reasons in relation to the A/B test group. However, as shown in the AA opt-out rates above, very few users opted-out of AA during the trial period. Future tests will aim to include this outcome as part of the experiment.

Issue report submissions

Reported as distinct reporters per group; raw submission counts are not used, as they can be dominated by individual repeat-submitters or automated reports (see Data Reliability). As can

be seen in the Table, there were no meaningful differences in the number of issue reports submitted to the extension across the groups.

Group	Distinct reporters	Rate per 1M	vs Control	p-value
Control	411	41.8		
After Unfilled	408	41.48	-0.80%	0.911
Any Size	414	42.1	+0.7%	0.918
All In-View	414	42.09	+0.7%	0.923
All Features	420	42.71	+2.2%	0.757

6.2 Moderate behavioural cues

Page views per session

Page views per session did not have any group showing a meaningful difference. By grouping user behaviour together there was no significant change within the MDE; variations were at the noise threshold.

Variant	Total AA Pageviews	Total AA Sessions	Aggregate AA PV / AA Session	% Diff from Control	p-value	Exceeds MDE	Assessment
Control	64,661,187	31,925,195	2.0254	—	—	—	—
A — After Unfilled	64,520,537	31,873,928	2.0242	-0.07%	0.6596	✗	Not significant
B — Any Size	64,563,563	31,971,366	2.0194	-0.30%	0.06104	✗	Not significant
C — All In-View	64,595,145	31,879,561	2.0262	0.0073	7.86E-05	✗	Significant but within MDE (treated as noise)

D — All Features	64,347,088	31,809,390	2.0229	0.0021	0.2711	x	Not significant
------------------	------------	------------	--------	--------	--------	----------	-----------------

6.3 Revenue

Relaxing ad refresh constraints delivered a clear, significant revenue gain in the groups that combine behaviours. After Unfilled and Any Size alone did not clear the pre-specified revenue threshold (4%) and are treated as noise. All In-View and All Features both showed significant, robust revenue lifts. The increase in publisher revenue was driven by more auctions and impressions per Acceptable Ads page view. The larger increase in Groups C and D is likely the result of more ads being shown (e.g., bid opportunities and successful bids), compared to Group A and B which may not have resulted in many more additional ad units being shown. The revenue gain was broadly distributed across publishers with the majority seeing similar size positive effects, with larger publishers receiving greater revenue as scaled with their traffic.

Group	Revenue lift vs. control	p-value	Assessment
A — After Unfilled	+0.6%	0.17	Not significant
B — Any Size	+2.1%	<0.001	Significant but within meaningful effect threshold (treated as noise)
C — All In-View	+4.5%	<0.001	Significant
D — All Features	+16.5%	<0.001	Significant, robust to multiple test methods

7. Summary of Findings

Group	Strong cues: negative effect?	Moderate cues:	Revenue uplift	Preliminary assessment
-------	-------------------------------	----------------	----------------	------------------------

		negative effect?		
A — After Unfilled	No	No	+0.6% (not significant)	No increase in negative user experience detected; revenue gain not significant. No AAC action required.
B — Any Size	No	No	+2.1% (within meaningful effect threshold — treated as noise)*	No increase in negative user experience detected; revenue gain insufficient to recommend in isolation.
C — All In-View	No (opt-out directionally higher; below significance — monitor)	No	+4.5% (significant)	No detectable increase in negative user experience; meaningful revenue gain. Request AAC approval.
D — All Features	No	No	+16.5% (significant, robust)	No detectable increase in negative user experience; strongest revenue outcome. Request AAC approval.

**Revenue uplifts are assessed against a pre-specified meaningful effect threshold of 4%, the minimum change considered practically significant rather than noise.*

8. Comparison with 2022 Survey Findings

Group B — Any Size. This condition allowed a refreshed ad to return a different size from the original. The 2022 survey tested specific fixed sizes in the side rail — 300×250 and 160×600, and found both acceptable, but did not evaluate what happens when the size changes on refresh. Group B's behavioural data shows no signal of increased disruption: uninstall rate was marginally lower than control (−0.82%, $p = 0.12$) and the revenue uplift of +2.1%, while statistically significant, fell below the meaningful effect threshold.

Group C — All In-View. The 2022 survey found a single side-rail unit acceptable; multi-unit refresh was not tested, and multi-ad conditions on non-scrollable pages were found inconclusive. Group C showed no increase in uninstalls (−0.09%, $p = 0.87$) and a significant revenue uplift of +4.5%. A directional increase in AA opt-outs (+8.9%, $p = 0.061$) was observed but did not reach significance; this is noted as a signal to monitor at rollout.

Groups A and D. No comparable 2022 conditions exist for the unfilled-bid trigger or the combined feature set. Neither showed an increase in uninstalls or opt-outs relative to control; Group D produced the strongest revenue outcome of all groups (+16.5%).

9. Limitations

The following limitation should be considered when interpreting findings from this report.

Page views per session. This metric cannot be tested for statistical significance in the current data, as user-level page view data does not yet exist; only aggregated per-website-day data is available. It is included as a descriptive indicator only.

10. Conclusion and Recommendations

Across every user-experience metric measured, treatment users behaved the same as control users. There were no significant increases in uninstalls, AA opt-outs, issue reports, or reduction in page views per session under any group. Relaxing the constraints related to in-view ad-refresh delivered a material revenue gain for publishers with no measurable cost to the user experience.

Group	Request	Rationale
-------	---------	-----------

A — After Unfilled	No AAC action required	Does not deviate from the current AA Standard
B — Any Size	Request more data	No increase in negative user experience detected, but revenue gain did not clear the meaningful effect threshold; insufficient basis for an AA Standard change in isolation
C — All In-View	Preliminary acceptance	Significant revenue gain (+4.5%); no detectable increase in negative user experience; opt-out directional signal warrants monitoring.
D — All Features	Preliminary acceptance	Strongest revenue outcome (+16.5%); no detectable increase in negative user experience across any metric; robust to multiple statistical tests