

Journal of Humanities and Society

available at http: www.econjournals.com

Journal of Humanities and Society, 2025, 1(1), 13-23.

What Changes When People Step Away from the News? Civic Knowledge, Political Judgments and Well-Being in Two Experiments

Muddassar Sarfraz*

School of Management, Zhejiang Shuren University, Hanzhou, China. *Email: muddassar.sarfraz@gmail.com

Received: 29 August 2025 **Accepted:** 30 October 2025 **DOI:** https://doi.org/10.32479/johas.25424

ABSTRACT

Does a week away from the news cost readers factual knowledge, and does it change how they feel? We reanalysed two randomised experiments that asked US participants to avoid news for 1 week and Polish participants to follow more news for 2 weeks, estimating for each outcome the difference between the assigned groups and the smallest symmetric margin its 90% confidence interval would fit. Eighteen of 24 intervals lay within an illustrative five-point margin. Two scores raised measurement concerns. Among the remaining 22 comparisons, 17 met the margin individually. After correction across all 24 comparisons, four supported that conclusion: attitude strength, attitude importance, and mental and physical well-being in Poland. In the US news-break experiment, the estimated mental well-being difference was −2.36 points (90% interval, −5.33-0.61), leaving a five-point reduction compatible with the observations. The narrowest interval in Poland concerned a factual score whose four components reused one survey item. That interval says little about knowledge across topics.

Keywords: News Avoidance, Equivalence Testing, Field Experiments, Political Knowledge, Well-Being
JEL Classifications: L82, D83, C93

1. INTRODUCTION

A reader who puts the news aside for a week may hope for some relief and wonder what they will miss. Those concerns pose different questions for a study. A modest drop in factual knowledge may matter even if the evidence is too imprecise to establish whether there was a drop at all, while a range of estimates that spans zero can still exclude a loss large enough to affect the reader's decision.

News avoidance is now a subject of research in both journalism and the wider social sciences, and a special issue of Journalism Studies examined why some people step away from particular subjects or have difficulty sustaining a news habit (Andersen et al., 2024; Schäfer et al., 2024). Across the markets in the 2024 Digital News Report, 39% said that they sometimes or often avoided news, compared with 29% in 2017 (Newman et al., 2024). Those responses identify a practice, but to find out what happens when a person accustomed to following news is invited to stop, one needs a comparison with people who could have received the same invitation and instead continued as usual.

Wojcieszak et al. (2022) offered panel participants two quite different changes to their usual routines. In the United States, the instruction was to spend 1 week avoiding news. In Poland, it was to spend 2 weeks seeking out more of it. The original investigators reported mostly small or statistically inconclusive differences across civic, political and personal outcomes and examined compliance, prior news habits and Bayesian evidence, whereas the question we take back to their archive is how much uncertainty surrounds those differences on the scales participants actually completed.

This Journal is licensed under a Creative Commons Attribution 4.0 International License

We keep the experiments separate because combining them would require deciding how a week of trying to avoid news compares with a fortnight of trying to consume more, even before considering differences in recruitment or in the questions asked in each country. Our contribution is a retrospective comparison of the margins supported by each outcome, followed by a check of the response records and the scoring code. The participants and experimental assignments all come from the original studies.

2. NEWS AVOIDANCE AND THE DEMOCRATIC AUDIENCE

2.1. A News Break Has More Than One Meaning

For the people interviewed by Toff et al. (2024), reluctance to follow news could reflect distress, frustration or an expectation that someone else would pass on anything worth knowing. An experimental instruction captures a particular moment in that process when a person agrees to interrupt their usual routine for a week, a choice that tells us considerably less about a household in which the habit of following journalism has gradually fallen away without anyone deciding to stop.

The subject avoided matters too. Schäfer et al. (2024) distinguish general avoidance from a decision to turn away from particular topics. Andersen et al. (2026) examine how selective avoidance can vary both between people and within the same person over time. A week in which someone declines to read about a war but follows an election closely would be a poor instance of complete abstention, although an observer tracking visits to one newspaper might be unable to distinguish it from many other changes in that person's reading.

Other explanations take the conditions of access seriously. Lindtner and Nærland (2024) show how poverty shapes everyday news use in Norway. Hasell and Halversen (2024) study the difficulty of evaluating online information alongside fatigue and avoidance. Zeng et al. (2024) examine avoidance among users of a short-video platform, where a person opening an application for entertainment can turn away from a political report for reasons quite different from those of someone who has stopped following the news across an entire week. A person choosing rest, a person who finds a report difficult to evaluate and someone who lacks the time or resources to attend to it need not experience the same consequence.

2.2. Exposure Can Change Without a Visible Public Response

A person can hear a headline from a friend, read about it later and never share the story publicly. Dennis (2024) traced changing news routines among young people in England during the pandemic. Tenenboim (2025) asked why people restrict the visibility of their engagement with news. Together, the studies give a concrete reason to expect that a record of visits to a news domain may miss encounters with the very same story through a message or a conversation elsewhere, even when the person makes no public comment on it.

That qualification matters particularly for democratic claims. Lee (2025) examines intentional avoidance in a setting of democratic backsliding. Robertson (2025) describes declining news engagement across a set of markets followed over time. Neither a political context nor a decade-long trend can be reduced to the circumstances of two short experiments, but the original investigators could do something those longer histories cannot: offer different instructions by random assignment to participants in the same setting. We can ask whether the contrast produced by that offer is precise enough to distinguish a small average difference from an unresolved one.

2.3. Why the Size of a Difference Belongs in the Question

Consider a newsroom deciding whether to describe a short news break as harmless to readers' political knowledge. An estimated difference of minus two points does not settle the issue because an interval that extends from below minus five to above zero leaves open a loss large enough to matter as well as the possibility that there was no loss at all. An editor who regards a loss of that size as consequential needs both ends of the interval, including the endpoint extending furthest in the direction of a possible loss.

Equivalence testing states that second question explicitly. Riesthuis (2024) relates the design of such tests to the smallest effect an investigator wants to distinguish, and here we report the margin the data would require and then compare it with several illustrative values, rather than assigning one social meaning to a point of political knowledge and a point of well-being. Five points is a reporting convention. Readers interested in a smaller change can read the required margin for each outcome directly from the tables.

3. DATA AND ANALYTICAL STRATEGY

3.1. Following Participants into the Analysis

We use version 1 of the replication archive released by Clemm von Hohenberg (2022), including the participant-level joined file and the code that constructs the survey scores. The two experiments were embedded in existing panels. In the US study, 803 people agreed to participate: 457 were assigned the news-break instruction and 346 to control, while in Poland, 445 of 939 consenting participants received the more-news instruction and 494 were assigned to control. Control participants received no corresponding instruction to change their news use (Wojcieszak et al., 2022).

For the main comparison we require an affirmative opt-in label, an assignment code and an observed outcome. The joined file contains assignment labels for 872 US records and 976 Polish records. Those totals include people outside the documented opt-in groups. The eligible US records with scores number 364 in treatment and 272 in control; the corresponding Polish counts are 397 and 402. All twelve main scores share those arm-specific counts. We retain the original assignment, irrespective of measured compliance. Because outcomes are missing, this is an assignment contrast among observed respondents. Interpreting it as the effect for everyone assigned requires assumptions about the absent scores.

Table 1 follows records through the joined data file, whereas the source article counts people who completed the post-test. The article reports 666 completions in the United States and 823 in

Poland. Before we require an explicit opt-in label, the file contains 637 and 823 observed mental scores, respectively. Applying that requirement leaves 636 and 799. The archive lets us verify the two fields separately, but it does not explain each discrepancy, so we also report what happens when we include every observed score that carries an assignment label.

3.2. What Was Assigned and What Was Observed

The American respondents were asked to avoid news across devices and services and to minimise conversations about current events, whereas Polish respondents were encouraged to use news more than usual, with examples from print, broadcast and online sources. Neither assignment guaranteed a particular number of stories read. Participants in either arm could encounter current events through friends or during unrelated browsing, and a person asked to read more could comply by choosing subjects they already knew well.

The US contrast is visible in the self-report index: 0.10 in the instructed group against 0.38 in control, and among respondents with browsing traces during the experiment, news-domain visits averaged 3.86 versus 5.22 per day. In Poland the corresponding visits averaged 5.65 versus 4.61, whereas the self-report means differed by only 0.01. The Polish browsing means come from just 155 treated and 104 control respondents, a subset that can describe exposure in those available histories but supplies a poor denominator for estimating how much one extra news visit might have changed a person's political judgment.

Table 2 also records a difference before the US instruction took effect. Among those with a prior trace, the treatment arm averaged 4.94 daily news-domain visits and the control arm 7.14. Different numbers of people contributed usable traces during the experiment, which means that subtracting the prior arm difference from the later one would confound a change in reading with a change in whose browser history entered the calculation. We use assignment to define the groups throughout.

3.3. Outcome Register and Assignment Contrasts

The archived analysis names thirteen outcomes. We retain twelve, including self-perceived and factual knowledge, support for compromise, strength and importance of attitudes, several evaluations of political opponents, perceived polarisation, and two well-being scores. Political participation is excluded from both experiments because the US post participation field is populated even for rows without the other post outcomes. It would silently introduce a different follow-up population. We report numerical results for all twelve retained fields, with qualifications for two specific constructions discussed below.

The mean contrast is the instructed arm minus its control in the same study. Scores retain their archived direction. A negative difference on the thermometer, for example, indicates a lower recorded score under the instruction, and assigning a political value to that change would require more information than its sign alone supplies. For unadjusted comparisons we estimate a Welch standard error from the sample variance in each observed arm, using Welch–Satterthwaite degrees of freedom for the intervals, where T is the instruction arm and C the control arm.

$$d = \bar{Y}_T - \bar{Y}_C, s = \sqrt{s_T^2 / n_T + s_C^2 / n_C} \quad (1)$$

Every contrast receives a two-sided 95% interval for a difference and a two-sided 90% interval for equivalence, with the latter required to fall wholly between the chosen negative and positive margins if both one-sided tests are to reject at a level of 0.05. Define m as the larger absolute endpoint of the 90% interval. A symmetric margin must be strictly greater than m for the test to pass. This single number allows readers to choose a margin quite different from ours without re-estimating the model.

$$L = d\text{-}t_{0.95,v}s,\ U = d\text{+}t_{0.95,v}s,\ m = max(|L|,|U|) \quad (2)$$

For each equivalence test we report the larger of the two one-sided P-values and, at the five-point margin, apply Holm's correction

Table 1: From assignment labels to observed response scores

Study	Arm	Listed	Opted in	Observed	Missing	Retained %
US	Control	359	346	272	74	78.6
US	Instruction	513	457	364	93	79.6
PL	Control	500	494	402	92	81.4
PL	Instruction	476	445	397	48	89.2

Listed includes records outside the documented opt-in groups. Observed refers to the primary outcome rows in the joined archive, not to all people reported as completing the survey

Table 2: How much news the two arms encountered

Indicator	Study	Control n	Control mean	Treat. n	Treat. mean	Δ
Prior use, channel-days	US	346	27.53	457	27.34	−0.19
Prior use, channel-days	PL	494	29.90	445	29.16	−0.74
Self-report during	US	273	0.38	364	0.10	−0.28
Self-report during	PL	402	0.49	397	0.50	0.01
Daily news visits, before	US	334	7.14	443	4.94	−2.21
Daily news visits, before	PL	474	5.86	430	5.48	−0.38
Daily news visits, during	US	266	5.22	349	3.86	−1.36
Daily news visits, during	PL	104	4.61	155	5.65	1.04
Political visits, during	US	266	7.03	349	5.50	−1.53
Political visits, during	PL	104	3.48	155	3.82	0.34

Prior use sums days of use across nine channels (0-63). Self-report during is the archive's unflipped 0-1 news-use index. Visit counts are means per day for available browsing histories. Different rows may describe different respondents

to the family of 24 tests, retaining pointwise 95% intervals for the ordinary difference tests. Since the same respondents answered several questions, eight outcomes meeting a margin do not constitute eight independent replications. We use unrounded endpoints for the decisions. Near the threshold, the displayed value can round to the margin itself.

3.4. Baseline Adjustment and Missing Responses

Where the joined file provides a baseline measure of the same outcome, we regress its post score on assignment and baseline score, with one common baseline slope within each model and heteroskedasticity-consistent standard errors (HC3) with residual degrees of freedom. The two well-being measures and US perceived knowledge have no available baseline of the same name. For these rows the sensitivity analysis uses the assignment term alone and still applies HC3. These specifications approximate the logic of conditioning on earlier responses; they are separate estimates, not a claim to reproduce every model in the source article.

To examine missingness, suppose the unobserved treatment mean exceeds the observed treatment mean by h points while the unobserved control mean falls h points below the observed control mean, so that equation (3) gives the resulting full-arm contrast when r denotes the observed fraction of each consenting arm. We change h from −10 to +10. The exercise imposes a pattern on missing means and gives no probability to any point on the line.

$$d(h) = d + h[(1-r_T) + (1-r_C)] \quad (3)$$

4. RESULTS

4.1. The Outcome Distributions Behind the Estimates

In the US control group the mental score averaged 36.36, against 34.00 among those asked to avoid news (Table 3). The standard deviations were 22.86 and 21.96. Thus a difference of a little over two points arose in groups whose individual responses were spread much more widely. The factual-knowledge means were 67.28 and 65.32. To learn whether knowledge and well-being changed together for the same individuals, we would need to analyse the association between their scores, a question that these separate comparisons of the average level in each arm do not address.

Poland offers a revealing contrast between two kinds of knowledge. The self-perceived knowledge means were 54.94 and 54.27, with standard deviations above 22 points (Table 4). The archived factual score averaged almost 49 in both arms and its standard deviations were about 7.4, a difference that makes the construction of the factual score, examined below, especially important to the interpretation of its apparently precise estimate. Carrying the two measures under a common heading of political knowledge would give the quieter of the two instruments more authority than it deserves.

The central estimates for support for compromise are below zero in both studies, −0.87 under the US news break and −1.64 under the Polish encouragement. Since the two assignments asked people to move their news use in opposite directions, the two negative signs cannot both be read as evidence that removing news reduces support for compromise. The appropriate comparison is always with the control arm in the same experiment.

4.2. Which US Differences can the Break Experiment Exclude?

Table 5 supports two statements about the US study. All twelve 95% intervals include zero, and eight of the twelve 90% intervals lie wholly within the five-point margin. Perceived knowledge meets that margin: its estimate is −1.38 and its interval runs from −4.51 to +1.75. Factual knowledge just misses. The estimate of −1.96 has a lower endpoint of −5.21, leaving a five-point loss compatible with the unadjusted comparison.

The US mental and physical scores give different answers at the same margin. For the mental score, a difference of −2.36 has a 90% interval from −5.33 to +0.61. Five points is too narrow. For the physical score, the interval around −1.44 runs from −3.62 to +0.74 and fits comfortably inside the margin, although a summary that mentions only the absence of a significant difference from control would make the two findings sound alike.

The malevolence result warrants attention for the opposite reason. Its 90% interval reaches −4.97, just inside the margin, and the unadjusted equivalence P = 0.0485. This is a fragile classification. A slightly smaller margin would reverse the decision. Its adjusted test also fails, so the contrast belongs among those for which the answer depends on whether an editor wants a claim about one political judgment or about all twelve outcomes examined together.

4.3. What the Polish Experiment Adds

Ten Polish contrasts fell within the five-point band (Table 6 and Figure 1). The two outside it concerned political judgments: the feeling thermometer (−2.89, 90% interval −5.05–−0.73) and perceived polarisation (−3.05, 90% interval −5.24–−0.87). Both pointwise 95% intervals also fell below zero. Here the data give some evidence about the sign of the difference while leaving open the possibility that its magnitude reached five points.

The Polish mental-score contrast, −0.20, puts both endpoints within three points of zero. For the physical score the difference is +0.52, and a symmetric margin just above 2.80 points suffices. The US intervals for those same outcomes are wider or further from zero. Instruction, time window, recruitment and follow-up composition all change between the two studies, which makes the difference in precision a poor basis for concluding that a person in Poland would respond differently from a person in the United States to the same request.

The Polish factual-knowledge estimate was almost exactly zero. Its 90% interval, from −0.88 to +0.85, was narrower than that for any other score. Why did this instrument distinguish such small differences? Taken by itself, the width of that interval would make factual knowledge appear the best measured of the Polish outcomes, yet the frequency distribution and the scoring script in Section 4.7 give a specific reason to qualify that reading.

4.4. The Margin and the Number of Claims

Across the 24 study–outcome contrasts, eighteen pass at five points before multiplicity adjustment. Only five pass after Holm

Table 3: Outcome distributions in the United States

Score	Control n	Mean	SD	Treat. n	Mean	SD
Perceived knowledge	272	62.38	23.44	364	61.00	24.09
Factual knowledge	272	67.28	24.40	364	65.32	24.77
Compromise	272	66.82	23.26	364	65.95	26.39
Attitude strength	272	70.32	17.41	364	69.26	19.13
Attitude importance	272	89.28	26.88	364	86.74	26.35
Feeling thermometer	272	68.10	20.99	364	69.35	20.65
Out-group stupidity	272	43.50	28.02	364	43.89	28.99
Lack of understanding	272	38.33	22.15	364	35.84	22.23
Malevolence	272	60.92	26.77	364	59.51	27.11
Perceived polarisation	272	64.64	19.89	364	63.35	19.86
Mental well-being	272	36.36	22.86	364	34.00	21.96
Physical well-being	272	20.50	16.43	364	19.07	16.60

Published score units retained without clipping. In this and later tables, “Factual knowledge” in Poland is the archived composite, which uses the same source item 4 times

Table 4: Outcome distributions in Poland

Score	Control n	Mean	SD	Treat. n	Mean	SD
Perceived knowledge	402	54.94	22.51	397	54.27	23.64
Factual knowledge	402	48.88	7.41	397	48.87	7.45
Compromise	402	63.07	27.56	397	61.43	26.55
Attitude strength	402	72.06	17.85	397	71.20	17.83
Attitude importance	402	72.86	17.06	397	72.51	16.84
Feeling thermometer	402	71.56	18.67	397	68.67	18.45
Out-group stupidity	402	46.83	27.95	397	45.86	25.56
Lack of understanding	402	38.59	21.23	397	40.09	19.78
Malevolence	402	69.33	22.44	397	68.57	21.80
Perceived polarisation	402	71.75	19.09	397	68.69	18.38
Mental well-being	402	41.88	21.28	397	41.68	19.48
Physical well-being	402	22.17	19.65	397	22.68	19.53

Published score units retained without clipping. In this and later tables, “Factual knowledge” in Poland is the archived composite, which uses the same source item 4 times

Table 5: Assignment contrasts and interval precision in the United States

Outcome	Δ	SE	95% low	95% high	90% low	90% high	δ needed
Perceived knowledge	−1.38	1.90	−5.11	2.35	−4.51	1.75	4.51
Factual knowledge	−1.96	1.97	−5.83	1.90	−5.21	1.28	5.21
Compromise	−0.87	1.98	−4.75	3.01	−4.13	2.38	4.13
Attitude strength	−1.06	1.46	−3.92	1.80	−3.46	1.34	3.46
Attitude importance	−2.53	2.14	−6.73	1.66	−6.05	0.99	6.05
Feeling thermometer	1.24	1.67	−2.04	4.53	−1.51	4.00	4.00
Out-group stupidity	0.38	2.28	−4.09	4.86	−3.37	4.14	4.14
Lack of understanding	−2.49	1.78	−5.99	1.00	−5.42	0.44	5.42
Malevolence	−1.41	2.16	−5.65	2.82	−4.97	2.14	4.97
Perceived polarisation	−1.30	1.59	−4.43	1.83	−3.92	1.33	3.92
Mental well-being	−2.36	1.80	−5.90	1.18	−5.33	0.61	5.33
Physical well-being	−1.44	1.32	−4.04	1.16	−3.62	0.74	3.62

Δ is instruction minus control. δ needed is the largest absolute 90% endpoint. A symmetric margin must exceed that number to establish pointwise equivalence

adjustment (Table 7). All five are Polish: factual knowledge, attitude strength, attitude importance and the two well-being scores. Once the problematic factual composite is set aside, four results remain that have both a direct interpretation on the archived score and familywise support for the specified margin. No US result meets that standard. The correction was applied to all 24 tests, including the two fields later flagged for interpretation, so dropping those fields did not give the other outcomes an easier test.

Using the control-arm standard deviation gives another reading of the same intervals. In the United States, the margin required for self-perceived knowledge is about 0.19 control standard deviations, and the required margin for mental well-being is about 0.23, whereas in Poland the two polarisation measures associated with pointwise differences need margins near 0.27. There is no reason to expect one standard deviation of anger, factual recall and perceived division to carry the same social meaning. These conversions describe dispersion, with the observed control SD treated as a fixed ruler rather than estimating a new set of standardised effects whose uncertainty would also depend on how well that standard deviation was measured in each control group.

At three points none of the raw US contrasts met the margin, while five Polish contrasts did. At five points the counts rose to eight and ten. At seven points all twenty-four intervals fitted. Figure 2 shows where each additional outcome enters the count as the margin

Table 6: Assignment contrasts and interval precision in Poland

Outcome	Δ	SE	95% low	95% high	90% low	90% high	δ needed
Perceived knowledge	−0.67	1.63	−3.87	2.54	−3.36	2.02	3.36
Factual knowledge	−0.01	0.53	−1.05	1.02	−0.88	0.85	0.88
Compromise	−1.64	1.91	−5.40	2.12	−4.79	1.51	4.79
Attitude strength	−0.86	1.26	−3.33	1.62	−2.93	1.22	2.93
Attitude importance	−0.35	1.20	−2.70	2.00	−2.33	1.63	2.33
Feeling thermometer	−2.89	1.31	−5.47	−0.31	−5.05	−0.73	5.05
Out-group stupidity	−0.98	1.89	−4.69	2.74	−4.10	2.14	4.10
Lack of understanding	1.50	1.45	−1.35	4.34	−0.89	3.89	3.89
Malevolence	−0.76	1.56	−3.83	2.32	−3.33	1.82	3.33
Perceived polarisation	−3.05	1.33	−5.66	−0.45	−5.24	−0.87	5.24
Mental well-being	−0.20	1.44	−3.03	2.63	−2.58	2.18	2.58
Physical well-being	0.52	1.39	−2.20	3.24	−1.77	2.80	2.80

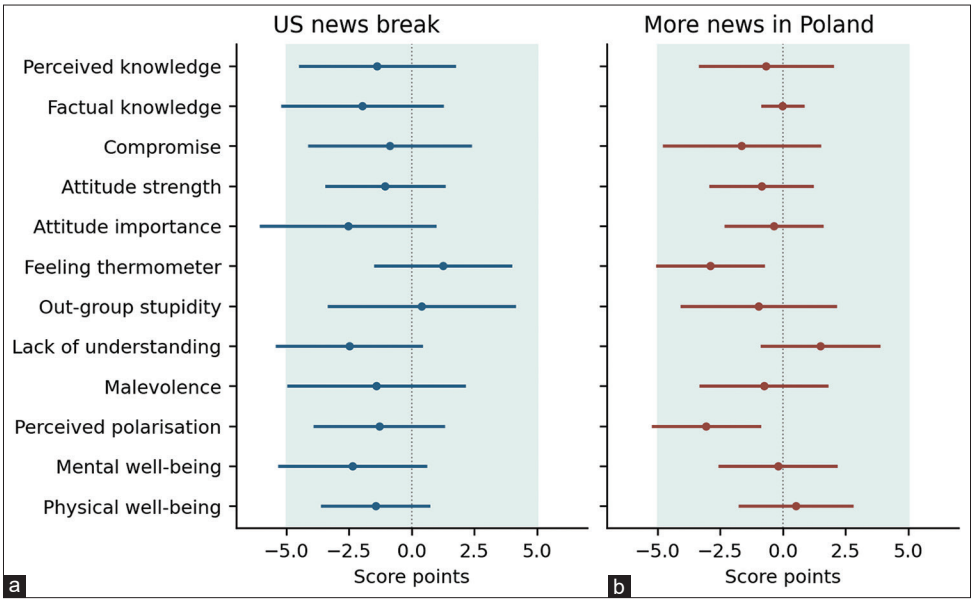
Δ is instruction minus control. δ needed is the largest absolute 90% endpoint. A symmetric margin must exceed that number to establish pointwise equivalence

Table 7: Outcome-specific equivalence tests at five score points

Outcome	US p	US Holm	US δ/SD	PL p	PL Holm	PL δ/SD
Perceived knowledge	0.0287	0.2579	0.19	0.0041	0.0653	0.15
Factual knowledge	0.0617	0.3392	0.21	<0.0001	<0.0001	0.12
Compromise	0.0186	0.2044	0.18	0.0398	0.3187	0.17
Attitude strength	0.0035	0.0649	0.20	0.0005	0.0113	0.16
Attitude importance	0.1243	0.3392	0.23	<0.0001	0.0013	0.14
Feeling thermometer	0.0125	0.1623	0.19	0.0543	0.3392	0.27
Out-group stupidity	0.0216	0.2165	0.15	0.0170	0.2037	0.15
Lack of understanding	0.0796	0.3392	0.24	0.0080	0.1198	0.18
Malevolence	0.0485	0.3392	0.19	0.0034	0.0649	0.15
Perceived polarisation	0.0102	0.1432	0.20	0.0712	0.3392	0.27
Mental well-being	0.0716	0.3392	0.23	0.0005	0.0101	0.12
Physical well-being	0.0036	0.0649	0.22	0.0006	0.0127	0.14

Holm adjusts 24 tests together. δ/SD divides the minimum margin by the observed control-arm SD. It is a descriptive conversion, without uncertainty in the SD

Figure 1: (a and b) Estimated effects against a five-point margin. Segments are pointwise 90% Welch intervals. Shading spans −5-+5. Positive means a higher stored score under the instruction



widens. A reader interested in a three-point change can therefore use the figure directly, without adopting our five-point convention.

4.5. What Changes after Adjustment?

Baseline adjustment places eleven US intervals and all twelve Polish intervals inside the five-point band (Table 8 and Figure 3). The American mental score remains outside it. There is no baseline measure of that outcome to account for any of its post-test dispersion. In contrast, the required US margin for factual knowledge falls from 5.21 to 4.45 points when the earlier factual score is used, bringing the estimate inside the band even though the estimated difference between the two arms remains negative. Improved precision, rather than a reversal of sign, changes the five-point decision.

Table 8: Baseline-adjusted effects and minimum margins

Outcome	US Δ	US δ	US n	PL Δ	PL δ	PL n
Perceived knowledge	−1.38	4.52	636	−1.05	2.63	799
Factual knowledge	−2.27	4.45	636	−0.27	1.10	799
Compromise	−0.03	2.25	636	0.45	3.11	799
Attitude strength	−0.71	2.52	636	0.63	2.24	799
Attitude importance	−1.52	4.15	636	0.75	2.20	799
Feeling thermometer	−0.50	2.22	636	−2.83	4.56	799
Out-group stupidity	−0.07	3.34	636	−1.90	4.77	799
Lack of understanding	−1.74	3.84	636	1.32	3.40	799
Malevolence	−0.03	2.08	636	−0.24	2.03	799
Perceived polarisation	−2.85	4.72	636	−0.95	2.58	799
Mental well-being	−2.36	5.33	636	−0.20	2.58	799
Physical well-being	−1.44	3.62	636	0.52	2.80	799

OLS includes the corresponding baseline measure where available. Intervals use HC3 standard errors and residual degrees of freedom. Mental and physical scores, and US perceived knowledge, have no baseline covariate

Adjustment matters particularly for perceived polarisation in Poland. Including the earlier polarisation score brings the minimum margin down from 5.24 to 2.58 points; the adjusted comparison would meet even a three-point criterion. The calculation uses the same baseline slope in both arms, and we report the unadjusted estimate alongside it because choosing between models after seeing which one meets a preferred margin would make that margin easier to satisfy.

The two screened columns in Table 9 remove the Polish factual score and the US attitude-importance score from substantive interpretation. The reason differs by field. The Polish scoring code collapses several purported items into one response, whereas the American importance field extends beyond the scale implied by its general rescaling rule. If those two fields are set aside, seventeen of the twenty-two remaining contrasts fit at five points, which gives an outcome-by-outcome summary of how much the experiments can rule out before any correction for looking at several outcomes. That count does not imply familywise equivalence for seventeen civic or personal consequences. Table 7 is the appropriate reference for a simultaneous claim.

4.6. How Consequential could the Missing Scores be?

US score retention is 79.6% in treatment and 78.6% in control. Polish retention is 89.2% in treatment and 81.4% in control. Similar retention rates, where they occur, do not ensure that the people missing from the two arms would have given similar responses, which is why equation (3) asks what would follow under a specific departure from that assumption. At h = 10 it adds 4.17 points to each US contrast and 2.94 points to each Polish contrast (Table 10). The larger American shift arises from the greater sum of its two missing fractions.

For the US mental score, the scenario contrasts range from −6.53 to +1.81 (Figure 4). They straddle zero even though the observed mean difference favours the negative direction, and in Poland the same exercise yields a range from −3.14 to +2.74, which also accommodates both signs despite its narrower width. These are neither confidence limits nor statements about what nonrespondents actually felt. Their value is to specify how much unobserved divergence the interpretation would have to tolerate,

Figure 2: The result depends on what counts as a small difference. Each step marks an interval that fits inside a slightly wider margin. These are pointwise counts of archived scores

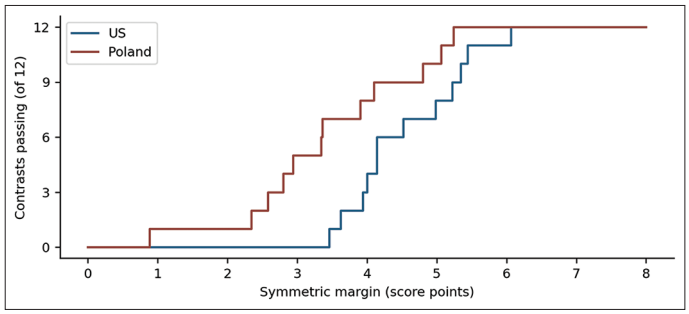


Figure 3: What earlier measures buy in precision. Points below the diagonal require a smaller margin after adjustment. Each point represents one study–outcome comparison

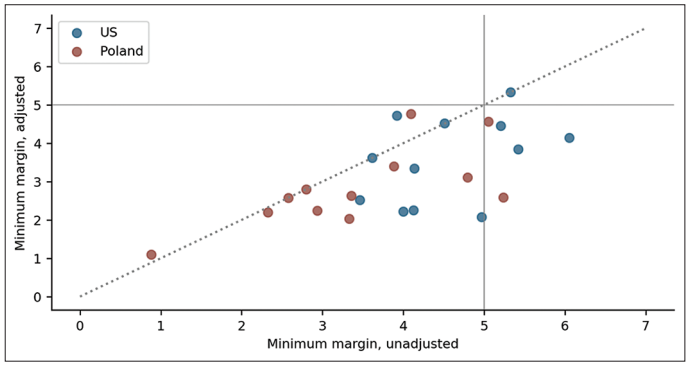
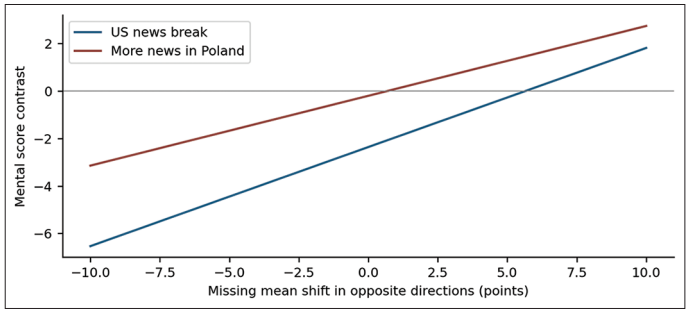


Figure 4: Sensitivity of the mental score to missing responses. The horizontal coordinate specifies an assumption, not an estimate. Lines cross zero when assignment-arm means become equal under the scenario



and the assumed shifts are stated in score points so that a reader can substitute a different value of h in equation (3).

Removing the opt-in requirement adds twenty-four Polish treatment records and one US treatment record with observed outcomes (Table 11). It adds no controls. For Polish compromise, those additional treatment records move the contrast from −1.64 to −1.92 and the minimum margin from 4.79 to 5.03, enough to reverse the decision at five points without any change to the responses of the original controls. The thermometer moves the other way and now fits inside five points. These two borderline results depend on which treatment records are admitted, even with the score and the interval calculation held fixed.

Table 9: How many contrasts fit each symmetric margin?

Margin	US raw/12	PL raw/12	US adj./12	PL adj./12	US screened/11	PL screened/11
1	0	1	0	0	0	0
2	0	1	0	1	0	0
3	0	5	4	8	0	4
4	4	8	7	10	4	7
5	8	10	11	12	8	9
6	11	12	12	12	11	11
7	12	12	12	12	11	11
8	12	12	12	12	11	11

Pointwise counts, no multiplicity adjustment. Screened columns exclude US attitude importance and Polish factual knowledge because of their scale diagnostics

Table 10: Missing-response sensitivity of the mean contrast

Outcome	US −10	US 0	US +10	PL −10	PL 0	PL +10
Perceived knowledge	−5.55	−1.38	2.79	−3.61	−0.67	2.27
Factual knowledge	−6.14	−1.96	2.21	−2.96	−0.01	2.93
Compromise	−5.05	−0.87	3.30	−4.58	−1.64	1.30
Attitude strength	−5.24	−1.06	3.11	−3.80	−0.86	2.09
Attitude importance	−6.71	−2.53	1.64	−3.29	−0.35	2.59
Feeling thermometer	−2.93	1.24	5.42	−5.83	−2.89	0.05
Out-group stupidity	−3.79	0.38	4.56	−3.92	−0.98	1.97
Lack of understanding	−6.67	−2.49	1.68	−1.45	1.50	4.44
Malevolence	−5.59	−1.41	2.76	−3.70	−0.76	2.19
Perceived polarisation	−5.47	−1.30	2.88	−5.99	−3.05	−0.11
Mental well-being	−6.53	−2.36	1.81	−3.14	−0.20	2.74
Physical well-being	−5.61	−1.44	2.74	−2.42	0.52	3.46

In +10 the missing treatment mean is ten points above its observed mean and the missing control mean ten below its observed mean. The −10 column reverses those departures. Values are scenario contrasts, not intervals

Table 11: Sensitivity to the archived opt-in field

Outcome	US opt-in Δ	US all Δ	US all δ	PL opt-in Δ	PL all Δ	PL all δ
Perceived knowledge	−1.38	−1.47	4.61	−0.67	−1.59	4.26
Factual knowledge	−1.96	−1.94	5.18	−0.01	0.05	0.89
Compromise	−0.87	−0.88	4.13	−1.64	−1.92	5.03
Attitude strength	−1.06	−1.07	3.46	−0.86	−0.83	2.88
Attitude importance	−2.53	−2.51	6.03	−0.35	−0.33	2.28
Feeling thermometer	1.24	1.19	3.94	−2.89	−2.72	4.83
Out-group stupidity	0.38	0.39	4.14	−0.98	−0.80	3.87
Lack of understanding	−2.49	−2.48	5.40	1.50	1.15	3.50
Malevolence	−1.41	−1.40	4.95	−0.76	−0.42	2.96
Perceived polarisation	−1.30	−1.27	3.89	−3.05	−2.71	4.86
Mental well-being	−2.36	−2.45	5.42	−0.20	0.07	2.42
Physical well-being	−1.44	−1.47	3.64	0.52	−0.02	2.25

All denotes all rows with an assignment label and an observed score; opt-in also requires the archive's affirmative opt-in label. The all-row analysis adds one US and 24 Polish score records

Two decisions about a data file thus have different consequences. Requiring an observable opt-in record specifies whose responses enter the comparison; allowing the missing mean to differ from the observed mean specifies a counterfactual value for responses that never entered it. Table 11 changes the former. Table 10 changes the latter. Combining the two would suggest a degree of knowledge about the missing records that the archive cannot provide.

4.7. An audit of the Most Precise Result

Of the 799 Polish factual scores in the main analysis, 781 are exactly 50 and eighteen are zero (Table 12). The script explains the absence of intermediate values. It draws on the same prime-minister item for each of the four components, awarding credit twice for response 3 and twice for response 1, so either answer earns two of four possible credits while a respondent who chooses either of the other categories earns none. The pre-test uses the same construction.

The Polish factual interval gives a precise estimate for that two-valued field, but using it to infer how people fared across four independent questions would require four distinct answers, a distinction the scoring script discarded before any arm means were calculated. The US factual score has five observed levels between zero and 100, consistent with an average of four binary items. We interpret it separately. A comparison of the two countries' factual knowledge would require comparable questions and scoring rules.

The US attitude-importance score raises a different issue. Among consenting respondents with observed values, 211 have scores above 100 and the highest is 133.33. The processing code applies

Table 12: Range and number of distinct stored outcome scores

Outcome	US min	US max	US values	PL min	PL max	PL values
Perceived knowledge	0.0	100.0	30	0.0	100.0	31
Factual knowledge	0.0	100.0	5	0.0	50.0	2
Compromise	0.0	100.0	25	0.0	100.0	25
Attitude strength	0.0	100.0	31	0.0	100.0	26
Attitude importance	0.0	133.3	42	0.0	100.0	25
Feeling thermometer	0.0	100.0	336	0.1	100.0	524
Out-group stupidity	0.0	100.0	65	0.0	100.0	43
Lack of understanding	0.0	100.0	89	0.0	100.0	83
Malevolence	0.0	100.0	31	0.0	100.0	28
Perceived polarisation	0.0	100.0	24	0.0	100.0	24
Mental well-being	0.0	100.0	35	0.0	100.0	37
Physical well-being	0.0	100.0	32	0.0	100.0	32

Distinct values refer to exact stored numeric values, not item-response options. The results use the stored units. Endpoints can differ across the experiments

a general 1-7 transformation to the composite, so we preserve the recorded values so that the published file remains recognisable, but do not interpret five points here as five percent of a validated scale. Before using this instrument for substantive comparison, one would need to check the original item response options and their mappings. A blanket cap at 100 would simply impose a new scoring rule on the 211 observations.

5. DISCUSSION

5.1. What the Evidence Permits a Newsroom to Say

For the US factual score, baseline adjustment changes the answer to the five-point question. The unadjusted interval admits a five-point loss after a week away from news; the adjusted one excludes it. Neither comparison measures a reader's knowledge after months without journalism. The instruction lasted 1 week, and participants could still learn about current events through conversation or on a device outside the browser record.

The evidence is more specific for some outcomes than for others. After adjustment across the full family of equivalence tests, the Polish more-news experiment supports differences smaller than five points on the two well-being scores as well as on attitude strength and importance, with all four conclusions restricted to consenting respondents for whom a post-intervention score is available in the public archive. The mental-score estimate in the American break experiment remains compatible with an effect larger than five points in one direction. A newsroom treating all of these results as evidence that time spent with news makes no difference would erase the strongest distinction the data allow.

An individual deciding to avoid news might also care about a sensation the study did not measure, since irritation at repeated alerts, a sense of having lost control of one's attention, or relief after disabling notifications may matter over a much shorter interval than the survey uses. Hasell and Halversen (2024) connect evaluative difficulty with fatigue. The aggregate mental score here bundles experiences across items. Its mean could move little even when one part of an individual's day changes. The archive supports neither a clinical recommendation about taking breaks nor an inference about how any one participant would benefit.

5.2. The Audience Behind an Experimental Instruction

Someone who encountered a headline despite trying to avoid the news stayed in the US instruction group, since retaining that person preserves the original assignment while restricting the group to successful avoiders would select participants on a choice they made after assignment. Dennis (2024) and Tenenboim (2025) also help explain why successful avoidance is hard to establish from browsing alone. A conversation or a private exchange can carry news without a recorded visit to a news domain.

Voluntary withdrawal is more heterogeneous still. Selectively avoiding reports of conflict differs from losing access to news through a combination of cost, time and social circumstances. Toff et al. (2024) take those situated accounts seriously. The experiments by Wojcieszak et al. (2022) make their own contribution precisely because they change the offer of exposure while leaving much of everyday life in place. That strength comes with a limit on interpretation: an effect of assigning the break cannot be divided by the difference of 1.36 daily news-domain visits and called the effect of missing a single visit.

The two national settings should be read asymmetrically. The Polish treatment asked for more news and the US treatment for less; the Polish self-report contrast was small, and observed trace coverage also differed. Even if both countries had used the same calendar window, a country comparison would require similar instruments and comparable treatment delivery. The Polish knowledge composite alone is sufficient reason to resist pooling, so we use Poland to ask a parallel question about an increase in news use, not to stand in for a counterfactual United States with the instruction reversed.

5.3. What Would Make the Next Study More Informative?

A new experiment could ask readers or editors which differences would alter an actual decision before selecting a margin. On a four-item factual scale scored in 25-point steps per person, a five-point group difference corresponds to one-fifth of a correct item on average, a conversion that has a clear interpretation in the American file because each of its four factual items contributes separately to the composite. A five-point change in a composite

mental score has no comparable interpretation without information on its items and on the change that readers regard as consequential. A margin chosen for numerical convenience should remain visible as such.

A short follow-up with a sample of people who missed the full questionnaire would help distinguish selection from a change in the outcome itself, provided researchers could reach those respondents without reintroducing the news exposure the intervention was designed to alter. A simpler improvement concerns record keeping: separate fields for random assignment, consent and the receipt of a usable post-test score would let later analysts tell exactly where each record dropped out.

The Polish factual score suggests a check that requires no new respondent. Before interpreting the interval, tabulate the derived values and trace each component back to its source question. In this case a variable with four component labels had only two observed scores. Anyone using the same script can reproduce the resulting interval, but treating its narrow width as evidence about four aspects of political knowledge would require four independently scored responses of the kind described by the outcome label.

5.4. Limits of this Secondary Analysis

The unadjusted comparisons describe people who supplied an outcome and had a documented opt-in, and the archive does not establish that missing outcomes are ignorable, that every treatment recipient followed the instruction, or that the recruited panels represent the full news audiences of either country. We examine sensitivity to alternative selection rules and missing means but cannot recover the missing responses. The calculations likewise give average effects; they place no useful bound on how much particular individuals may have changed in opposite directions.

The margins and instrument checks were chosen retrospectively. They form a way to interrogate the public evidence, not a preregistered confirmatory test. Holm adjustment controls familywise error across the stated 24 five-point tests under their modelling assumptions; it does not turn the counts at every other margin, the baseline-adjusted estimates or the missing-response scenarios into confirmatory findings. Reporting a series of margins allows the interpretation to remain open to readers who value a smaller difference or require a stronger standard of evidence.

Finally, differences here are in recorded score units. Only two studies and two short interventions are represented, so the conclusions concern what their archives can sustain, not the long-term civic consequences of industry change or the experiences of people who have stopped following journalism altogether. For those populations, the interviews and audience surveys that motivated the analysis still ask questions these experiments cannot answer.

6. CONCLUSION

In the US experiment, a claim that a week away from news changed none of the outcomes would go further than the data allow. Eight of twelve 90% intervals fit within five points, but the unadjusted intervals for factual knowledge and mental well-being still included a five-point reduction. Ten Polish intervals fit, including one based on a repeated factual item. Among the twenty-two interpretable contrasts across both experiments, seventeen met the five-point margin individually and four did so after correction across the full family.

For a reader considering a short news break, the American experiment gives a firmer bound on some political judgments than on possible changes in mental well-being, whereas the broader civic question concerns the consequences of a much longer withdrawal from journalism among people who may never have agreed to take part in an experiment. A 1-week interruption and a 2-week increase in news use leave that question open.

REFERENCES

Andersen, K., Shehata, A., Skovsgaard, M., Strömbäck, J. (2026), Selective news avoidance: Consistency and temporality. Communication Research, 53(3), 291-318.

Andersen, K., Toff, B., Ytre-Arne, B. (2024), Introduction: What we (don't) know about news avoidance. Journalism Studies, 25(12), 1367-1384.

Clemm Von Hohenberg, B. (2022), Replication Data for: Null Effects of News Exposure: A Test of the (un)Desirable Effects of a 'News Vacation' and 'News Binging' (Version 1) [Data Set]. Cambridge: Harvard Dataverse.

Dennis, J. (2024), How news found the avoiders: The changing news routines of infodemically vulnerable young people in England during Covid-19. Journalism Studies, 25(12), 1516-1534.

Hasell, A., Halversen, A. (2024), Feeling misinformed? The role of perceived difficulty in evaluating information online in news avoidance and news fatigue. Journalism Studies, 25(12), 1441-1459.

Lee, F.L.F. (2025), Politically driven intentional news avoidance under democratic backsliding. The International Journal of Press/Politics, 30(4), 1001-1019.

Lindtner, S.S., Nærland, T.U. (2024), News avoidance and poverty: Intersectional marginalization in the Norwegian "media welfare state". Journalism Studies, 25(12), 1498-1515.

Newman, N., Fletcher, R., Robertson, C.T., Ross Arguedas, A., Nielsen, R.K. (2024), Reuters Institute Digital News Report 2024. United Kingdom: Reuters Institute for the Study of Journalism.

Riesthuis, P. (2024), Simulation-based power analyses for the smallest effect size of interest: A confidence-interval approach for minimum-effect and equivalence testing. Advances in Methods and Practices in Psychological Science, 7(2), 1-12.

Robertson, C.T. (2025), People are Turning Away from the News. Here's Why it May Be Happening. Reuters Institute for the Study of Journalism. Available from: https://reutersinstitute.politics.ox.ac.uk/news/people-are-turning-away-news-heres-why-it-may-be-happening

Schäfer, S., Betakova, D., Lecheler, S. (2024), Zooming in on topics: An investigation of the prevalence and motives for selective news avoidance. Journalism Studies, 25(12), 1423-1440.

Tenenboim, O. (2025), Under-the-radar engagement: How and why news users limit their public expression. Journal of Computer-Mediated Communication, 30(1), zmae024.

Toff, B., Palmer, R., Nielsen, R.K. (2024), Avoiding the News: Reluctant Audiences for Journalism. New York: Columbia University Press.

Wojcieszak, M., Clemm Von Hohenberg, B., Casas, A., Menchen-Trevino, E., De Leeuw, S., Gonçalves, A., Boon, M. (2022), Null effects of news exposure: A test of the (un)desirable effects of a 'news vacation' and 'news binging'. Humanities and Social Sciences Communications, 9, 413.

Zeng, R., Guo, S., Evans, R. (2024), Intentional news avoidance on short-form video platforms: A moderated mediation model of psychological reactance and relative entertainment motivation. Humanities and Social Sciences Communications, 11, 248.