Visual presentation raises trust in election polls

Elias Koch[1,2], Simon Munzert[1], Linus Hagemann[1], Hannah Rajski[3,4], Thomas Gschwend[3,4], Cornelius Erfort[5], and Lukas F. Stoetzer[5]

[1]Data Science Lab, Hertie School
[2]Research Training Group DYNAMICS
[3]University of Mannheim
[4]Mannheim Centre for European Social Research (MZES)
[5]University of Witten/Herdecke

October 27, 2026

Election polls are arguably the most visible signals of public opinion in modern democracies, yet how the design of poll reports shapes their appeal, perceived credibility, and informational value has not been systematically studied. In a preregistered visual conjoint experiment conducted during the 2025 German federal election campaign, 22,658 voters evaluated poll reports varying across format, reference values, and technical information, in 160 possible designs. All visual formats were rated as more attractive and trustworthy than text-only reports. Trust gains were greatest among respondents with lower data and polling literacy and less accurate beliefs about current polling, whereas the appeal of visualization was greatest among the most literate. Disclosing the margin of error made reports less attractive without affecting trust, and no format made expectations about the election outcome more accurate. Text, the dominant format in poll coverage, is thus the least effective at fostering trust and accessibility.

Introduction

Pre-election polls are a central feature of modern democratic politics (1). They provide the electorate with a regularly updated signal of electoral competition, which helps people form expectations about the behavior of others and adapt their voting accordingly (2–5). Polls have also become a central information tool for political elites, allowing them to adapt their behavior (6–9). Consumers of polls, however, rarely encounter polls as raw numbers but as a particular depiction of them chosen by media outlets. How that choice shapes the way polls are perceived and trusted is not sufficiently understood.

News reporting on election polls varies substantially in both visual format and presented content. Our media content analysis of newspaper coverage during the 2025 German federal election campaign (see SI Section B) shows considerable variation in the inclusion of contextual and technical information such as reference values, electoral thresholds, methodological details, and statistical uncertainty. At the same time, despite the recent rise of data journalism and data visualization becoming more mainstream, more than 90% of analyzed articles featuring polling information still relied exclusively on text-based reporting.

Decades of research in visualization science and cognitive psychology establish that the way quantitative information is presented to audiences matters: format determines which patterns are easy to extract and how much they are believed (10, 11). Graphic encodings can improve comprehension and recall relative to text and tables (12, 13) but also shape affective and credibility judgments: visual framing alters risk perception (14), chart design influences perceived bias and persuasiveness (15, 16), and—critically for polls—depictions of statistical uncertainty are read in systematically different ways depending on how they are drawn (17–19). This body of work, much of it developed in the human–computer interaction and visualization communities, has rarely been brought to bear on political information, where questions of trust and credibility are paramount.

In political science, by contrast, the role of polling information for political behavior has largely been treated as a substantive rather than a design problem (20–22). More recent work has turned to selected elements of information provision, such as polling uncertainty (23), electoral thresholds (5), or—in related work on election forecast communication—probabilities and chart types (24, 25). While those studies have generated valuable evidence on the potential influence of poll communication features on political behavior, there is no systematic, causal account of how a variety of relevant design decisions when communicating polls shapes audience perceptions.

Here, we bridge these literatures. We show that visual design systematically alters the perceived attractiveness and trustworthiness of poll reporting, even when the underlying results are held constant. We conducted a large preregistered experiment among German voters ($N = 22{,}658$) in the run-up to the 2025 federal election, in which participants viewed different representations of the same poll. To the best of our knowledge, our sample size makes this the largest experimental study on the presentation of election polls to date. We orthogonally varied key features of real-world reporting, including presentation format (text, table, bar chart, dot plot, line plot), the inclusion of reference values (none, previous poll, polling average, previous election result), and additional technical information (electoral thresholds, statistical uncer-

tainty, sample characteristics). Crossing these features factorially yields 160 distinct reports and lets us estimate the causal effect of each design choice while averaging over the others, at a scale that supports resolving how those effects differ across audiences.

We focus on three core dimensions that capture how individuals engage with quantitative information. For a poll to inform political behavior, its depiction must first draw attention and be perceived as appealing; it must be trusted as a credible representation of reality; and it must enable individuals to form expectations about relevant outcomes. We therefore measure three corresponding outcomes — perceived attractiveness, trustworthiness, and the quality of expectation formation regarding relevant political scenarios. Together, our outcomes capture whether and how information can meaningfully inform decision-making.

Our results show that visual presentations are consistently perceived as more attractive and more trustworthy than text-only formats. The two benefits fall unevenly across audiences: trust gains are largest among individuals with lower data literacy and those whose prior beliefs diverge most from the poll, while gains in attractiveness occur across the board but are largest among more literate readers—so that visualization serves an accessibility function, extending trust to those who typically struggle to interpret quantitative political information. At the same time, reporting the margin of error makes a poll report less attractive without changing how much it is trusted, and it does not help audiences form more realistic expectations.

We also explore in how far these findings extend beyond voters, to journalists and political candidates. In these substantially smaller samples, the evidence is necessarily tentative. Yet, the estimates are broadly consistent with the voter results, with some suggestive differences. In contrast to voters, for example, journalists appear skeptical of visualizations, whereas both journalists and candidates tend to find visual reporting more visually attractive. We further find suggestive evidence that candidates associate the reporting of statistical uncertainty with greater trustworthiness. Taken together, these patterns suggest that the effects of poll presentation may vary across those who communicate political information, those who consume it, and those whose electoral prospects are shaped by it.

These findings have implications beyond election polling. They demonstrate that trust in public information is shaped not only by data quality or transparency, but also by how information is designed and communicated. By providing causal evidence on how specific reporting choices influence trust and understanding, our study contributes to broader debates about the role of information in democratic governance and offers evidence-based guidance for communicating quantitative information in ways that support informed decision-making.

Design and Data

To systematically evaluate how choices in poll reporting influence audience perceptions, we fielded a large visual conjoint experiment among individuals eligible to vote in the run-up to the 2025 German federal election between January 24 and February 21, 2025. See *Materials and Methods* for more information on participant recruitment, sample characteristics, the survey instrument, collected outcome variables, statistical analyses, and the electoral context of our application case.

To complement the voter sample, we fielded the identical survey instrument among two

additional populations: political candidates running in the 2025 German Bundestag election and German newspaper journalists (see SI Section A.3.1). Due to the limited number of responses in both these samples (N = 108 newspaper journalists; N = 143 candidates), the corresponding analyses are exploratory and will only be interpreted with caution. Nevertheless, they provide valuable context for our findings from the voter experiment and, in the case of journalists, complement the results of our media content analysis. We therefore report these analyses complementary to the main results.

Varied Features The experiment employed a between-subjects visual conjoint design. Most participants viewed a single depiction (456 rated two) of an election poll drawn from a $5 \times 4 \times 2 \times 2 \times 2$ factorial that varies five design features (Figure 1), yielding 160 distinct visual configurations. Crossed with these design features, we additionally manipulated the poll's substantive content—the performance attributed to the respondent's preferred party—across three levels (its real level, above it, or below it), so that the full pool of vignettes comprises 480 conditions. This performance manipulation tests whether visual presentation attenuates desirability bias and is analyzed in a companion paper. Here, we focus exclusively on the five design features; all models include voting-intention and performance fixed effects, and we do not interpret the performance component (see *Materials and Methods*).

First, we varied the *presentation format of the poll*. Reflecting formats commonly used in real-world election poll reporting, we included textual and tabular presentations as well as bar charts, line plots, and dot plots. The distinction between text and graphics is grounded in a long line of work in visualization science: graphical encodings can facilitate the extraction of quantitative relationships compared to information transported in text and tables (10, 11). Following the cognitive framework for decision-making with visualizations developed by Padilla et al. [26], presentation format governs how much interpretive work is delegated to the perceptual system rather than to effortful symbolic reasoning: well-chosen visual encodings support near-immediate extraction of patterns such as rank order and relative magnitude, whereas the same quantities must be read off and mentally compared when presented as text. Format should therefore shape not only how accessible and attractive polling information appears, but also how readily, and how accurately, audiences can read electoral standings from it.

Importantly, the choice is not merely whether to visualize but how. Different chart types encode the same values through different visual channels—position, length, or aligned points—and are not interchangeable in how easily or accurately they are read (10, 11). Our three visual formats correspond to conventions we observe in the German polling news environment (see SI Section B) as well as in scientific reporting of polling-based information. Bar charts are by far the most common way polls are visualized, making them the natural visual baseline. Line plots, historically rarer, have gained ground with the rise of data journalism, where they are used chiefly to render poll-of-polls trends over time. Dot plots remain less familiar to general audiences but are well established in scientific reporting and election forecasting. Comparing these formats therefore lets us ask not only whether visualization helps, but whether the specific encoding a newsroom chooses—largely by convention rather than by evidence—affects how the underlying poll is perceived.

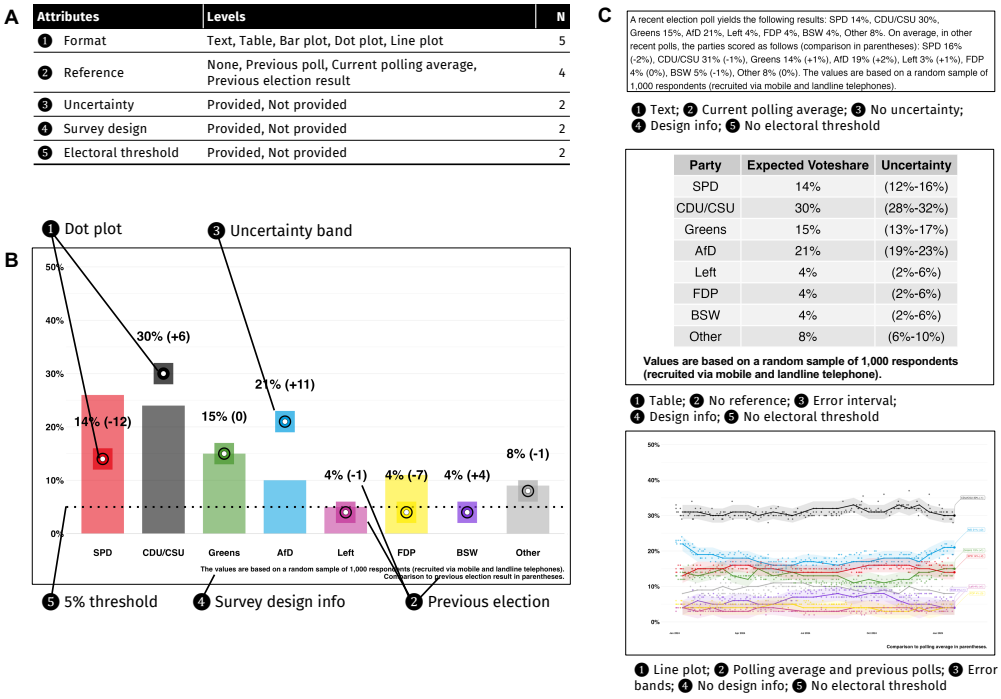


Figure 1: **Randomization setup of poll vignettes.** **(A)** Vignette attributes and levels, which were randomly combined into poll vignettes, resulting in $N_{\text{msg}} = 160$ unique design variations. **(B)** Illustration of poll vignette: A dot plot enriched with additional information; **(C)** Further examples of vignettes with different attribute combinations. See SI A for an overview of the vignette construction and additional examples. All vignettes were presented in German; English translations above are provided solely for illustrative purposes.

Second, we varied whether the poll was contextualized with a *reference value*, displaying the change relative to the previous election result, the previous poll by the same firm, the average of all current polls, or no reference point. Where a reference value was included, the depiction reported the comparison to that benchmark in parentheses. Reference values turn a static snapshot into a signal of movement, and political-behavior research suggests such signals matter for how polls are received. Bandwagon accounts, for example, argue that perceived electoral gains and losses can shift voter support and expected electoral outcomes (20, 21, 27, 28). Similarly, Bayesian learning models of campaign dynamics show that voters use new polls as information signals to update their beliefs about party strength (29). In this framework, a benchmark can determine whether a poll reads as consistent with previously reported results or as a marked departure, which could in turn condition how credible and informative audiences find the poll.

Third, we varied whether the poll reported the *statistical uncertainty* associated with the presented vote shares (margins of error of ± 2 percentage points). Whether and how to communicate this uncertainty is actively debated (5, 23). Experimental work outside the polling

domain finds that disclosing numerical uncertainty makes audiences appropriately less certain about a point estimate without necessarily eroding trust in the number or its source (30, 31). How uncertainty is rendered, however, is itself consequential: research in visualization science shows that different encodings of the same uncertainty produce systematically different inferences, and that conventional interval representations are frequently misread (17, 18, 32). We manipulate whether a margin-of-error accompanies the poll, holding its representation fixed; because making uncertainty salient may undercut the sense that a poll pins down the state of the race, we expected it to bear on trust and expectation quality even where it clarifies the underlying estimate.

Fourth, we varied whether the depiction referenced the *electoral threshold* (i.e., the five percent of the vote required for parliamentary representation). The threshold is a salient institutional cue in multi-party systems, where the viability of smaller parties and the resulting coalition arithmetic hinge on which parties clear it. Consistent with research on how voters rely on simple heuristic cues to navigate complex electoral choices (33, 34), displaying the threshold should help audiences judge which parties are at risk of falling out of parliament and thereby shape their expectations about its composition.

Finally, we varied whether the poll was accompanied by *methodological information* about the underlying sample, specifying that "The values are based on a random sample of 1,000 respondents (recruited via mobile and landline telephone)." Because the polling source itself was held constant across all vignettes, this manipulation isolates the effect of methodological transparency rather than source identity and variation in methodology. Consistent with research showing that respondents draw on such methodological quality cues when evaluating polls, sometimes in tension with partisan considerations (35–37), the disclosure of sample characteristics should shape how credible the poll appears—though not unambiguously, as transparency may equally foreground a poll's limitations rather than bolster its perceived quality.

The design varies the presence of each reporting element while fixing its concrete implementation. Format is the exception, spanning five types (text, table, bar, dot, line), so the visual-versus-text contrast rests on multiple implementations; the other four features each take a single form—a parenthetical reference value, a ±2-point margin, one threshold cue, and one methodological statement. Effects for these features should thus be read as conditional on the chosen implementation, most consequentially for uncertainty, where alternative encodings may shape comprehension and trust differently (32).

Our pre-analysis plan committed us to estimating average marginal component effects (AMCEs) and their heterogeneity for the reporting-design features at the center of this study, but it specified no directional hypotheses. Prior work offers no comprehensive basis for such predictions across these features. We therefore report the results in an explicitly exploratory register, consistent with our pre-registered commitment to flag such analyses as exploratory. The analysis describes, systematically and at scale, how presentation choices map onto audience evaluations. The estimates themselves remain causal, following from the randomized assignment of vignette features.

Outcomes Below each vignette, respondents answered three sets of questions capturing the perceived attractiveness of the poll presentation, its trustworthiness, and the quality of expectations formed based on the information provided. These dimensions reflect key stages in how individuals engage with political information: whether it captures attention, whether it is perceived as credible, and whether it can be used to form meaningful judgments about political outcomes.

Attractiveness was measured using items assessing how easy to understand, attractively designed, and helpful the poll presentation was perceived to be. Trustworthiness captured perceived accuracy of the poll as a representation of current vote intentions and expected election outcomes. Finally, the alignment of expectations with a forecast benchmark was derived following a two-step approach: First, respondents provided probabilistic assessments of three electoral scenarios: (a) the CDU/CSU becoming the strongest party, (b) the AfD becoming the second-strongest party, and (c) the FDP receiving less than 5% of the vote. In a second step, these expectations were then evaluated against the scenario probabilities generated by the *Zweitstimme* forecasting model (38), with alignment measured as the negative mean absolute error, so that higher values indicate more realistic expectations. These items were combined into three standardized indices corresponding to the respective dimensions (SI Section A). See SI for more information on the electoral setting of our application (SI Section A.4.5).

Results

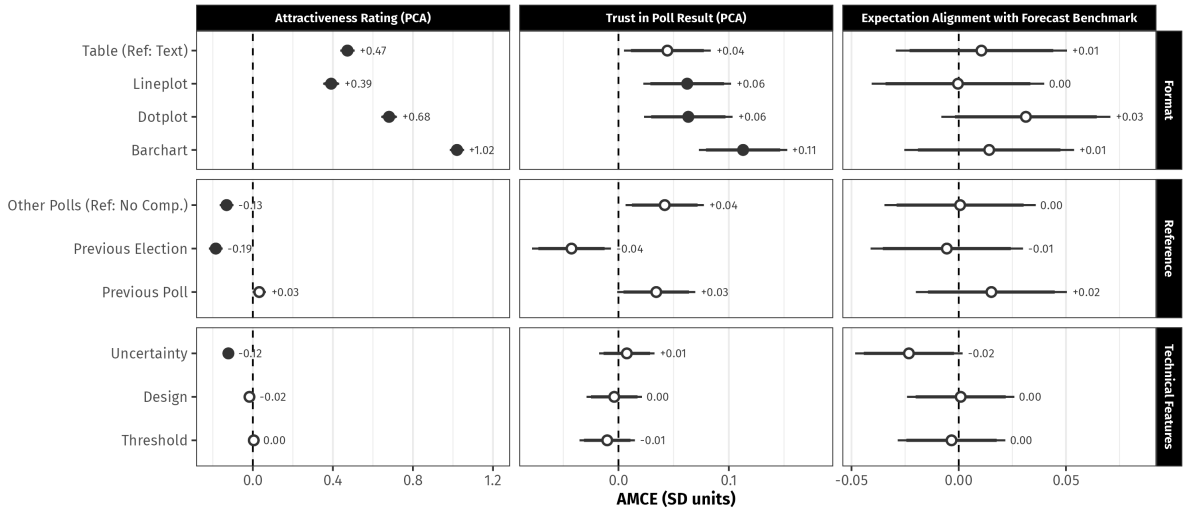


Figure 2: **Effects of poll reporting features on attractiveness, trust, and expectation alignment.** AMCEs in standard deviations of the respective outcome, estimated by OLS with voting-intention and performance fixed effects; 95% (narrow) and 90% (wide) confidence intervals, standard errors clustered by respondent. Solid dots mark effects that remain significant after Benjamini–Hochberg correction across all effects in the figure; hollow dots mark all other effects. Reference categories in parentheses; numbers of observations in SI Table A6.

Main Effects

Figure 2 reports the main AMCE estimates in standardized units of the varied features. Any type of visualization — bar charts, dot plots, or line graphs — is perceived as substantially more attractive and more trustworthy than text-based reporting of the same poll (effect sizes in raw slider and rating units are reported in SI Section A.7). The attractiveness gains through visualization are substantial: bar charts (+1.02 s.d.; 95% CI = [0.98, 1.06]; $p < 0.001$), dot plots (+0.68 s.d.; 95% CI = [0.64, 0.72]; $p < 0.001$), and line graphs (+0.39 s.d.; 95% CI = [0.35, 0.43]; $p < 0.001$) all outperform text-based reporting. More importantly, these formats are likewise rated more trustworthy (bar charts +0.11 s.d.; 95% CI = [0.07, 0.15]; $p < 0.001$; dot plots +0.06 s.d.; 95% CI = [0.02, 0.10]; $p < 0.01$; line graphs +0.06 s.d.; 95% CI = [0.02, 0.10]; $p < 0.01$). Even a plain tabular presentation is already rated more attractive (+0.47 s.d.; 95% CI = [0.44, 0.51]; $p < 0.001$) and more trustworthy (+0.04 s.d.; 95% CI = [0.00, 0.08]; $p < 0.05$) than text, although the trust gain from tables, unlike those from the visual formats, does not survive false-discovery-rate correction (SI Section A.8). Visual formats do not, however, help audiences form more realistic expectations about party competition: none of the formats shifts expectation alignment relative to text (all $|\beta| \leq 0.03$ s.d.; all $p > 0.10$).

The consequences of including reference values are more nuanced. Comparing current polling standings to the polling average (−0.13 s.d.; 95% CI = [−0.17, −0.10]; $p < 0.001$) or the previous election result (−0.19 s.d.; 95% CI = [−0.22, −0.15]; $p < 0.001$) decreases the perceived attractiveness of the report, while referencing the most recent poll by the same firm leaves attractiveness essentially unchanged (+0.03 s.d.; 95% CI = [−0.00, 0.07]; $p = 0.06$). In terms of trust, recent references tend to lend credibility: respondents report higher trust when the polling average is included (+0.04 s.d.; 95% CI = [0.01, 0.08]; $p < 0.05$), with a comparable, marginally significant shift for the previous poll (+0.03 s.d.; 95% CI = [−0.00, 0.07]; $p = 0.06$), whereas contrasting current standings with the previous election result decreases trust (−0.04 s.d.; 95% CI = [−0.08, −0.01]; $p < 0.05$). These trust effects are small and do not survive false-discovery-rate correction (SI Section A.8). One plausible interpretation is that recent reference values typically imply only minor deviations from the current poll, and observing that a new estimate falls within the range of existing measurements may reassure consumers of its validity. The election results, by contrast, diverge more markedly from current standings, potentially prompting skepticism about the accuracy of the reported figures. With regard to expectation quality, none of the reference value conditions produce statistically significant effects (all $|\beta| \leq 0.02$ s.d.; $p > 0.05$), suggesting that the inclusion of benchmarks shapes how polls are received but not how effectively they calibrate audience expectations about electoral outcomes.

Turning to the technical features of poll reports, the inclusion of uncertainty information is the only design element that produces a statistically significant effect, reducing the perceived attractiveness of the reporting (−0.12 s.d.; 95% CI = [−0.15, −0.10]; $p < 0.001$). There is suggestive evidence that uncertainty information also diminishes expectation quality, though this effect does not reach conventional significance levels (−0.02 s.d.; 95% CI = [−0.05, 0.00]; $p = 0.07$). Neither the reporting of design details nor the inclusion of the electoral threshold significantly moves any of the three outcome dimensions (all $|\beta| \leq 0.02$ s.d.; $p > 0.05$).

Heterogeneous Effects

If visualizations are effective by reducing the cognitive demands of interpreting poll information, their effects should be strongest among respondents for whom polls are hardest to evaluate. To explore these patterns, we examine three dimensions of heterogeneity: respondents' data literacy and their polling literacy, both of which may condition the capacity to process different reporting formats, and the accuracy of their prior beliefs about party support relative to the presented results, which may shape how receptive they are to the signal conveyed by new polling information. For ease of presentation, Figure 3 reports AMCEs separately for respondents at the low and high end of each moderator, leaving out a middle group so that the two groups are clearly distinguishable (see *Materials and Methods*). Formal tests of moderation use the full, continuous version of each moderator (SI Section A.6).

Data literacy moderates the two benefits of visualization in opposite directions (Figure 3A). The trust advantage is concentrated among low-literacy respondents: bar charts increase trust among respondents with low data literacy (+0.18 s.d.; 95% CI = [0.09, 0.27]; $p < 0.001$) but not among those with high data literacy (+0.02 s.d.; 95% CI = [−0.11, 0.14]; $p > 0.05$), and the difference between the two groups is significant ($p < 0.05$). Dot plots follow the same ordering (low literacy +0.10 s.d.; 95% CI = [0.01, 0.20]; $p < 0.05$ vs. high literacy +0.00 s.d.; 95% CI = [−0.12, 0.12]; $p > 0.05$), although here the group difference is not significant. A continuous specification that retains the full variation in literacy confirms the moderation for bar charts ($p < 0.05$), but not for the other formats (SI Section A.6). Attractiveness is moderated in the opposite direction. Both groups rate visual formats as more attractive than text, but the gains are larger among respondents with high data literacy: bar charts (high literacy +1.16 s.d.; 95% CI = [1.05, 1.27] vs. low literacy +0.83 s.d.; 95% CI = [0.75, 0.92]), tables (+0.52 vs. +0.35 s.d.), and especially line graphs (high literacy +0.64 s.d.; 95% CI = [0.52, 0.76] vs. low literacy −0.11 s.d.; 95% CI = [−0.20, −0.01]), which low-literacy respondents rate slightly below text. The group differences are significant for all three formats (all $p < 0.05$).

Polling literacy, measured by five true/false statements about how polls are conducted (for example, on sample size, weighting, and differences between polling firms), shows the same division (Figure 3B). Bar charts raise trust more among respondents with low polling literacy (+0.19 s.d.; 95% CI = [0.11, 0.26]; $p < 0.001$) than among those with high polling literacy (+0.08 s.d.; 95% CI = [0.02, 0.13]; $p < 0.05$). This difference is significant both in the group contrast ($p < 0.05$) and in the continuous specification ($p < 0.01$; SI Section A.6). For the other formats, the trust effects are small and similar in both groups. On attractiveness, respondents with high polling literacy again reward visual formats more: the gains from bar charts (high +1.04 s.d.; 95% CI = [0.99, 1.09] vs. low +0.91 s.d.; 95% CI = [0.84, 0.98]; $p < 0.01$ for the difference) and line graphs (high +0.47 s.d.; 95% CI = [0.41, 0.53] vs. low +0.19 s.d.; 95% CI = [0.11, 0.27]; $p < 0.001$) differ significantly between the groups, while the differences for dot plots and tables are not significant. None of the subgroup AMCEs for the political expectations is statistically distinguishable from zero.

A similar pattern emerges when respondents are grouped by the accuracy of their prior beliefs about party support (Figure 3C). The trust advantage is concentrated among respondents whose priors deviate most from the presented poll. In this low-accuracy group, all four formats raise

trust (bar charts +0.15 s.d.; 95% CI = [0.08, 0.21]; $p < 0.001$; line graphs +0.09 s.d.; 95% CI = [0.03, 0.16]; $p < 0.01$; tables +0.08 s.d.; 95% CI = [0.02, 0.15]; $p < 0.05$; dot plots +0.07 s.d.; 95% CI = [0.00, 0.14]; $p < 0.05$), whereas none does among respondents with highly accurate priors (largest estimate: bar charts +0.05 s.d.; 95% CI = [−0.02, 0.11]; $p > 0.05$). The group differences are significant for bar charts, line graphs, and tables (all $p < 0.05$), but not for dot plots. In the continuous specification, the moderation of the bar-chart effect falls just short of conventional significance ($p = 0.05$; SI Section A.6). Attractiveness is again moderated in the opposite direction: both groups rate visual formats as more attractive than text, but the gains are larger among respondents whose priors already align closely with current polling: bar charts (high accuracy +1.13 s.d.; 95% CI = [1.07, 1.19] vs. low accuracy +0.86 s.d.; 95% CI = [0.80, 0.93]), tables (+0.57 vs. +0.37 s.d.), and especially line graphs (high accuracy +0.63 s.d.; 95% CI = [0.56, 0.70] vs. low accuracy +0.11 s.d.; 95% CI = [0.04, 0.18]) are rated more attractive by respondents with accurate priors (all differences $p < 0.001$). Dot plots are again the exception, rated almost identically in both groups (+0.70 vs. +0.65 s.d.; $p > 0.05$ for the difference).

The heterogeneity thus follows the same logic for all three moderators: the subgroup for whom the poll poses a greater processing challenge gains most in trust, while the subgroup that engages with it more readily rewards visual presentation with higher attractiveness. The continuous specifications confirm the attractiveness moderation of bar charts and line graphs for all three moderators, and the trust moderation of bar charts for data literacy ($p < 0.05$) and polling literacy ($p < 0.01$).

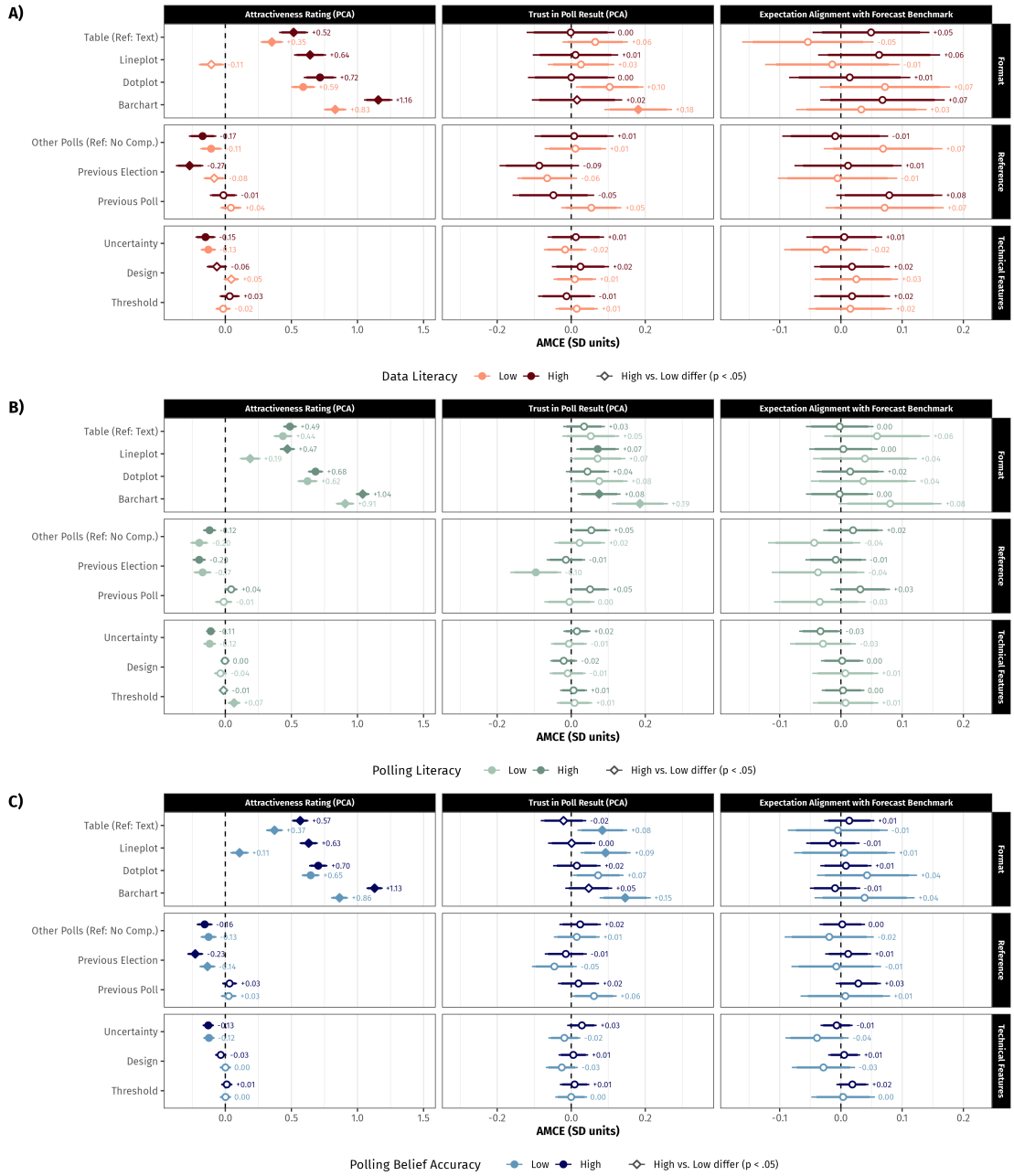


Figure 3: **Heterogeneous effects of poll reporting features by (A) data literacy, (B) polling literacy, and (C) prior-belief accuracy (voters).** Subgroup-specific AMCEs in standard deviations of the respective outcome, estimated by OLS with voting-intention and performance fixed effects; 95% (narrow) and 90% (wide) confidence intervals, standard errors clustered by respondent. For ease of presentation, groups contrast respondents at the low and high end of each moderator, leaving out a middle group (see *Materials and Methods*): low data literacy means at most one of three items correct (N = 4,390 respondents), high means all answered items correct (N = 2,179); low polling literacy means at most one of five statements correct (N = 6,074), high means at least three correct (N = 10,923); low and high belief accuracy are the top and bottom third of the mean absolute deviation between prior beliefs and the displayed poll (N = 7,801 and 7,157). Tests of moderation use the continuous moderators (SI Section A.6). Solid dots mark effects that remain significant after Benjamini–Hochberg correction across all effects within a panel; hollow dots mark all other effects. Reference categories are text (format) and no comparison (reference value). Diamonds mark features whose effect differs significantly in the direct comparison between the respective *high* and *low* group ($p < .05$).

Effects on journalists and candidates

Finally, we compare these patterns among voters with evaluations provided by journalists and political candidates. Given the smaller elite subsamples (N = 108 newspaper journalists; N = 143 candidates), these comparisons are necessarily exploratory; the full results are reported in SI Section C.1.

Effects are broadly consistent across samples, yet some notable patterns stand out. First, while visualizations consistently increase perceived trustworthiness among voters, journalists rate bar charts (−0.63 s.d.; 95% CI = [−1.16, −0.10]; $p < 0.05$) and dot plots (−0.62 s.d.; 95% CI = [−1.17, −0.07]; $p < 0.05$) as less trustworthy than text, with a weaker effect in the same direction for line graphs (−0.46 s.d.; 95% CI = [−0.94, 0.02]; $p = 0.06$). Given the small sample, none of these effects survives false-discovery-rate correction (SI Section A.8). The skepticism towards visualization is notable in light of our descriptive evidence that newspaper coverage of election polls remains dominated by text-based formats: newspaper journalists may evaluate as more credible the very format they are most accustomed to producing.

Like voters, however, both journalists and candidates perceive visual representations as more attractive than text-based reporting, most clearly for bar charts (journalists +0.69 s.d.; 95% CI = [0.11, 1.27]; $p < 0.05$; candidates +1.15 s.d.; 95% CI = [0.78, 1.52]; $p < 0.001$) and line graphs (journalists +0.58 s.d.; 95% CI = [0.03, 1.12]; $p < 0.05$; candidates +0.93 s.d.; 95% CI = [0.42, 1.44]; $p < 0.001$). Further, the trust effect of communicating statistical uncertainty is largest among political candidates (+0.23 s.d.; 95% CI = [−0.07, 0.52]; $p = 0.13$; voters +0.01 s.d.; journalists +0.09 s.d.), although it does not reach conventional significance. Finally, journalists, unlike voters, rate the inclusion of the electoral threshold as more attractive (+0.39 s.d.; 95% CI = [0.00, 0.77]; $p < 0.05$; voters +0.00 s.d.; $p > 0.05$; candidates +0.29 s.d.; $p = 0.06$).

Discussion

How citizens encounter political information depends not only on its substance, but also on how that information is communicated. Using a large-scale visual conjoint experiment conducted during the 2025 German federal election campaign, we show that design choices in poll reporting systematically shape how audiences evaluate election polls. Across every visual format we tested, graphical presentations were perceived as more attractive and more trustworthy than text-only reporting of the same numbers.

These findings are notable because text-based reporting remains the dominant format in contemporary German election coverage: our content analysis shows that fewer than one in ten articles visualize the polls they report. The format that audiences find the least attractive and least trustworthy is thus the one they overwhelmingly encounter. This gap between prevailing practice and audience response is the central practical implication of our study.

The advantage of visualization, however, is neither uniform across audiences nor formats. Gains in *trust* are concentrated among the readers for whom a poll is hardest to process—those with lower data literacy, and those whose prior beliefs diverge most sharply from the reported standings. Gains in *attractiveness*, by contrast, accrue most to the readers already best equipped

to engage with a poll: the statistically literate and those whose priors are already accurate. Visualization therefore does double duty. It rewards fluent readers with a more engaging artifact while extending credibility to those readers who might otherwise struggle to extract a trustworthy signal. An important implication of this pattern is that visualization of polling results can help narrow the gap in who can benefit from polling information. At the same time, our results should not be read as a blanket endorsement of "more visualization" no matter how: communicating statistical uncertainty reduced perceived attractiveness and showed suggestive signs of degrading expectation quality, indicating that the benefit lies in specific design choices rather than in visual embellishment as such.

Our study makes three main contributions. First, it provides systematic causal evidence that the mode of presentation itself is a determinant of how citizens evaluate political information. While prior work has emphasized how accuracy, source credibility, or partisan alignment shape trust in polls, our findings show that design choices—particularly the choice between textual and visual formats—are independently consequential. Even when the underlying data are held constant, changing the representational form meaningfully shifts both perceived credibility and accessibility. This suggests that communication design is not just a peripheral concern in political information environments but a core component of how information acquires authority.

Second, the study advances understanding of heterogeneity in information processing by showing that design effects are systematically moderated by data literacy and prior polling beliefs. Rather than functioning as a uniform "visual advantage," graphical presentation confers trust where numeracy and prior accuracy are lowest, and appeal where they are highest. This speaks to debates about informational inequality in democratic societies (39, 40) by highlighting communication design as a potential corrective mechanism.

Third, the results bear on best practices in journalistic data communication and complement recent findings on effective journalistic communication through visualization (41–43). The widespread dominance of text-based reporting in our descriptive evidence stands in tension with the experimentally observed advantages of visual formats.

Several limitations warrant consideration. First, although we examine multiple reporting formats, each feature is represented in only a single implementation. In practice, there is a wide range of options for communicating elements such as uncertainty, threshold information, or sampling details. Future research, particularly work aimed at developing accessible and feasible reporting practices, should compare alternative design choices within each feature category. Second, our elite comparisons rest on small samples ($N = 108$ journalists; $N = 143$ candidates) such that patterns merit confirmation at scale. Third, we study a multi-party system with a proportional threshold. While this setting reflects a substantial share of Western democracies, party systems with fewer competing actors, such as those typically found in presidential systems, may structure voter perceptions and media practices differently. The effects observed here may therefore not generalize fully to two-party or highly candidate-centered environments.

Finally, our design isolates the effect of how a given poll is presented, not considering anything that surrounds that encounter in the real world. The aggregate influence of poll communication on the public, however, operates through a longer chain. Which polls citizens see at all is shaped by exposure and selection: audiences choose outlets and stories in line with their predispositions,

so the reports we randomize would, in the field, reach people non-randomly (44, 45). And the poll itself is only one element of a report, which also carries headlines, verbal framing, and surrounding narrative that independently move opinion (46, 47). Our estimates should thus be read as the causal effect of presentation *conditional on* exposure to the report, not as the total real-world impact of poll communication, which additionally depends on who is reached and how the numbers are framed in prose.

Taken together, our findings demonstrate that the presentation of polling information is not merely cosmetic. Communication design shapes whether citizens perceive polls as credible, accessible sources of political information. As election polling continues to occupy a central role in democratic information environments, identifying reporting practices that improve both comprehension and trust represents an important task for scholars, journalists, and democratic institutions alike.

Materials and Methods

The research received ethical approval from the Research Ethics Committee at the Hertie School (Reference: 20250102-141). The study was pre-registered at the Open Science Framework (OSF) on January 24, 2025 (`10.17605/OSF.IO/W7QZF`). The analytical procedure and the deviations from the pre-analysis plan are documented in SI Section A.1.

Participants

Participants were recruited between January 24 and February 21, 2025, via an online access panel by Bilendi ($N = 10{,}564$) and social media ads on Meta platforms (*Facebook* and *Instagram*; $N = 12{,}094$), following best-practice targeting principles by Neundorf & Öztürk [48]; all participants were eligible to vote in the federal election on February 23, 2025, and completed the same *Qualtrics* survey (see SI Section A.3). In total, 22,976 participants rated 23,444 poll vignettes (456 rated two). For 318 of them, the vignette ratings could not be linked to their survey responses, so their recruitment channel and vote intention are unknown; they are excluded from all analyses. The analysis sample therefore comprises 22,658 respondents (10,564 from the online panel and 12,094 from social media) who rated 23,114 vignettes.

Every respondent provided informed consent at the beginning of the survey. At the end of the survey, every respondent received the current standing in the polls, presented as a bar chart, alongside information on where to find current, high-quality polling data on the German election.

Experimental setup

Campaign Environment The experiment was fielded during the campaign for the 2025 German federal election, which was called as a snap election following the collapse of the so-called traffic-light coalition led by Chancellor Olaf Scholz. The coalition comprised the Social Democratic Party (SPD), the Greens (Bündnis 90/Die Grünen), and the liberal Free Democratic Party (FDP). These incumbent parties competed against the Christian Democrats (CDU/CSU), led by the eventual chancellor Friedrich Merz, the far-right Alternative for Germany (AfD), the left party (Die Linke), and the newly formed Bündnis Sahra Wagenknecht (BSW).

Presented Poll As the baseline, we used a poll conducted by Forschungsgruppe Wahlen, one of the leading polling firms in Germany, and published on 10 January 2025. The poll reported vote shares of 30% for the CDU/CSU, 14% for the SPD, 15% for the Greens, 4% for the FDP, 4% for the Left, 21% for the AfD, and 4% for the BSW. The remaining 8% were attributed to other parties and were not displayed.

As described in above, we also manipulated the reported support for respondents' preferred party. Its vote share was set to be lower, unchanged, or higher than in the reference poll (Forschungsgruppe Wahlen, 10 January 2025): shares of 10% or more were adjusted by ± 5 percentage points, while smaller

shares were multiplied by 0.5 or 1.5. The remaining parties were then rescaled proportionally to sum to 100%. Because the magnitude of the manipulation varied with respondents' preferred party and its baseline support, all analyses include party and performance fixed effects to account for differences in the resulting poll to exclusively focus on the variations in the depiction of the polls.

Measurement

Outcomes We measured three groups of outcome variables capturing the attractiveness, the trustworthiness, and the quality of the expectations derived from the respective poll. See SI Section A.4 for the documentation of how the variables were collected in the survey environment. Each group of outcomes was later combined into joint measures.

Attractiveness Rating (PCA) was measured using four trade-off items helping participants to describe a poll, which were displayed in random order. Respondents were asked to indicate whether they perceived the depiction a) to be designed in an attractive or unattractive way, b) to comprise too much or too little information, c) to be helpful or not helpful, and d) to be easy or not easy to understand. The index uses items (a), (c), and (d); item (b) is excluded because its endpoints capture the amount of information rather than a positive or negative evaluation (SI Section A.4.2).

Trust in Poll (PCA) was measured using two items. Participants were asked to indicate in how far they perceived the reported poll to reflect a) the *current distribution of voting intentions in the electorate* and b) their own expectations for the outcome of the federal election. Both answers were collected on a sliding scale ranging between 0 (*not at all an accurate reflection*) and 100 (*represented a fully accurate reflection*). Again, this outcome was operationalized as a latent factor score derived from a principal component analysis (PCA) with one factor.

Expectation Alignment with Forecast Benchmark was measured following a two step approach. In the survey, participants were asked to indicate the likelihood of three scenarios, namely the CDU/CSU being the strongest party, the AfD being the second strongest party and the FDP receiving less than 5% of the votes on election day. Here, we used a 0-100 percent scale, labeled with "very unlikely" and "very likely" at the two respective ends. In a second step, we determined in how far the indicated likelihood was a realistic assessment. For this purpose, we compared participants' responses with the corresponding scenario probabilities estimated by the *Zweitstimme* election forecasting model (38). We use the forecast from January 10, 2025, the day the baseline poll was published. The outcome is the negative mean absolute error, so that higher values indicate better alignment. For a more precise documentation of the computation of the expectation alignment outcome, see SI Section A.4.2.

Covariates We further measured three moderators: respondents' data literacy, their polling literacy, and the accuracy of their prior beliefs about current party standings.

Data Literacy was measured in the second survey wave (February 13-23, 2025), a median of ten days after the experiment, using three items from the INSPIRE inventory (49) that capture respondents' ability to match quantitative information to its correct graphical representation (DVL4), to extract politically relevant information from a visual data display (DVL5), and to reason about stochastic quantities (DVL9). The score is the share of answered items answered correctly and is available for 9,550 respondents who took part in the second wave. For the heterogeneity analyses, we contrast respondents who answered at most one third of the items correctly (46%) with those who answered all items correctly (23%) and leave out respondents with two correct answers (31%). Results for the continuous measure are reported in SI Section A.6; see SI Section A.4.3 for the full item wording.

Belief Accuracy was measured by comparing respondents' prior expectations about party support with the polling figures displayed in their assigned vignette. Prior to the treatment, participants were asked to estimate the current polling standing of each party. For each respondent, we then computed the mean absolute deviation between these elicited priors and the presented poll values across all eight parties (CDU/CSU, SPD, AfD, Greens, FDP, Left, BSW, and Other), such that smaller values indicate more accurate priors, i.e. a weaker required belief update. For the heterogeneity analyses, we contrast vignettes in the bottom third of this deviation (high belief accuracy) with those in the top third (low belief accuracy) and leave out the middle third.

Polling Literacy was measured with five statements about how election polls work, and respondents were asked to select all statements they considered correct. Three of the statements are true (polls conducted via social media can be very inaccurate because parts of the population do not use social media; weighting responses by characteristics such as age, gender, and place of residence can improve a poll's accuracy; some polling firms systematically report higher or lower support for certain parties),

and two are false (polls with many participants always reflect public opinion better than polls with few participants; a representative poll requires at least 3,000 respondents). A statement counts as answered correctly if it is true and was selected, or false and was not selected. Respondents who selected none of the statements were not assigned a score; the score is available for 22,326 respondents. For the heterogeneity analyses, we contrast respondents with at most one correct answer (27%) with those with at least three (49%), leaving out respondents with two correct answers (24%).

Statistical Analysis

We estimate average marginal component effects (AMCEs) with ordinary least squares (OLS) regressions of each outcome on all five design features. All models include fixed effects for voting intention and the performance condition, which is analyzed in a separate study. Standard errors are clustered by respondent, since some respondents rated two vignettes, and effects are reported in standard deviations of the respective outcome in each sample. Consistent with the pre-analysis plan, we rely on classical frequentist inference and the conventional $p < 0.05$ threshold (narrow error bars in coefficient plots); coefficient plots additionally show 90% confidence intervals (wide error bars). As a robustness check not specified in the pre-analysis plan, we apply Benjamini–Hochberg false discovery rate correction within each coefficient plot; solid dots mark effects that remain significant after this correction (SI Section A.8). For the heterogeneity analyses, we divide each moderator into thirds and contrast the bottom and top third, leaving out the middle third. Because data and polling literacy take only a few distinct values, each score value is assigned as a whole to one third, so respondents with the same score always fall into the same group; the resulting groups are therefore not exact thirds. We estimate the models separately in both groups for presentation and test whether an effect differs between them by interacting all features with the group indicator in a pooled model (diamonds in Figure 3). Formal tests of moderation interact each format with the continuous moderator (SI Section A.6).

In the experimental analyses, we leave the data unweighted. Under randomization, demographic imbalance does not generally bias treatment effect estimates (50, 51). Deviations between sample and population should therefore matter little for the causal quantities of interest.

Media Content Analysis

To identify relevant newspaper articles, we conducted a search in *LexisNexis*. We restricted the search to articles published between January 1, 2025, and February 24, 2025, and included the keywords "Bundestagswahl" (federal election) and "Umfrag*" (poll). This procedure yielded 2,219 potentially relevant articles. Four trained research assistants screened all articles and determined whether they reported on an election poll. In 661 articles where this criterion was met, the coders extracted additional information on how the poll results were presented, following a detailed coding scheme documented in SI Section B.

Acknowledgments

This project has benefited from excellent feedback by audiences at the *European Political Science Society*, the *Data Science Lab* at the *Hertie School*, the *Social Science Data Lab* at the *Mannheim Centre for European Social Research* (MZES), the *Zweitstimme.org Workshop on Elections, Public Opinion, and Elite Behaviour* at the *Hertie School*, and the *International Association for the Study of German Politics* (IASGP). We thank Anna Aigner, Stine Bartram, Marietta Sujin Hüttemann, Till Kolloge, Nicolas Reichardt, Tom Rößger, Leon Schmidt, Lisa-Marie Zikesch, and David Zingher for their excellent research assistance.

Funding: This contribution was generously funded by the German Research Foundation (DFG; Grant Number 529544178).

Author contributions: E.K., S.M.: Conceptualization, Methodology, Formal analysis, Writing – original draft; E.K., S.M., H.R., T.G.: Investigation; H.R., T.G.: Data curation; All authors: Writing – review and editing; S.M.: Supervision; E.K., S.M.: Project administration; S.M., T.G., L.F.S.: Funding acquisition.

Competing interests: The authors declare that they have no competing interests.

Data and materials availability: All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Replication data and code will be available at the Harvard Dataverse.

Use of AI tools: The authors used Claude Opus 5.5 (Anthropic) for assistance with developing and reviewing analysis code and with editing the manuscript text. All code, analyses, and text were reviewed and verified by the authors, who take full responsibility for the content of this article.

References

1. W. Jennings & C. Wlezien. "Election polling errors across time and space". In: *Nature Human Behaviour* 2.4 (2018), pp. 276–283. DOI: 10.1038/s41562-018-0315-6.
2. J. O. Dahlgaard, J. H. Hansen, K. M. Hansen & M. V. Larsen. "How Election Polls Shape Voting Behaviour". In: *Scandinavian Political Studies* 40.3 (2017), pp. 330–343. DOI: 10.1111/1467-9477.12094.
3. T. W. G. van der Meer, A. Hakhverdian & L. Aaldering. "Off the Fence, Onto the Bandwagon? A Large-Scale Survey Experiment on Effect of Real-Life Poll Outcomes on Subsequent Vote Intentions". In: *International Journal of Public Opinion Research* 28.1 (2016), pp. 46–72. DOI: 10.1093/ijpor/edu041.
4. A. Mehrabian. "Effects of poll reports on voter preferences". In: *Journal of Applied Social Psychology* 28.23 (1998), pp. 2119–2130. DOI: 10.1111/j.1559-1816.1998.tb01363.x.
5. W. Krause & C. Gahn. "Self-Fulfilling Prophecies at the Voting Booth: Evidence from Polling Effects around Electoral Thresholds". In: *Journal of Politics* (2026). DOI: 10.1086/742632.
6. M. M. Pereira. "Responsive Campaigning: Evidence from European Parties". In: *The Journal of Politics* 82.4 (2020), pp. 1183–1195. DOI: 10.1086/707827.
7. H. H. Pedersen & H. B. Seeberg. "Using Social Media to Respond to Negative Polls: Politicians' Issue Responsiveness on Facebook". In: *The Journal of Politics* (2024). DOI: 10.1086/733007.
8. Ž. Poljak. "The impact of public approval on the use of negativity throughout the electoral cycle". In: *West European Politics* 46.5 (2023), pp. 947–970. DOI: 10.1080/01402382.2022.2095487.
9. E. Koch. "Opposition parties seek more conflict with the government when losing electoral support". In: *West European Politics* (2025). DOI: 10.1080/01402382.2025.2524944.
10. W. S. Cleveland & R. McGill. "Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods". In: *Journal of the American Statistical Association* 79.387 (1984), pp. 531–554. DOI: 10.1080/01621459.1984.10478080.
11. S. L. Franconeri, L. M. Padilla, P. Shah, J. M. Zacks & J. Hullman. "The Science of Visual Data Communication: What Works". In: *Psychological Science in the Public Interest* 22.3 (2021), pp. 110–161. DOI: 10.1177/15291006211051956.

12. S. S. J. Alhadad. “Visualizing Data to Support Judgement, Inference, and Decision Making in Learning Analytics: Insights from Cognitive Psychology and Visualization Science”. In: *Journal of Learning Analytics* 5.2 (2018), pp. 60–85. DOI: 10.18608/jla.2018.52.5.

13. K. Eberhard. “The effects of visualization on judgment and decision-making: a systematic literature review”. In: *Management Review Quarterly* 73.1 (2023), pp. 167–214. DOI: 10.1007/s11301-021-00235-8.

14. Z. Hildon, D. Allwood & N. Black. “Impact of format and content of visual display of data on comprehension, choice and preference: a systematic review”. In: *International Journal for Quality in Health Care* 24.1 (2012), pp. 55–64. DOI: 10.1093/intqhc/mzr072.

15. E. Holder & C. X. Bearfield. “Polarizing Political Polls: How Visualization Design Choices Can Shape Public Opinion and Increase Political Polarization”. In: *IEEE Transactions on Visualization and Computer Graphics* 30.1 (2024), pp. 1446–1456. DOI: 10.1109/TVCG.2023.3326512.

16. A. V. Pandey, A. Manivannan, O. Nov, M. Satterthwaite & E. Bertini. “The Persuasive Power of Data Visualization”. In: *IEEE Transactions on Visualization and Computer Graphics (Proc. IEEE VIS)*. Vol. 20. 12. 2014, pp. 2211–2220. DOI: 10.1109/TVCG.2014.2346419.

17. J. Hullman, P. Resnick & E. Adar. “Hypothetical Outcome Plots Outperform Error Bars and Violin Plots for Inferences about Reliability of Variable Ordering”. In: *PLOS ONE* 10.11 (2015), e0142444. DOI: 10.1371/journal.pone.0142444.

18. A. Kale, F. Nguyen, M. Kay & J. Hullman. “Hypothetical Outcome Plots Help Untrained Observers Judge Trends in Ambiguous Data”. In: *IEEE Transactions on Visualization and Computer Graphics* 25.1 (2019), pp. 892–902. DOI: 10.1109/TVCG.2018.2864909.

19. C. Stokes, C. X. Bearfield & M. A. Hearst. “The Role of Text in Visualizations: How Annotations Shape Perceptions of Bias and Influence Predictions”. In: *IEEE Transactions on Visualization and Computer Graphics* 30.10 (2024), pp. 6787–6800. DOI: 10.1109/TVCG.2023.3338451.

20. C. Marsh. “Back on the bandwagon: The effect of opinion polls on public opinion”. In: *British Journal of Political Science* 15.1 (1985), pp. 51–74.

21. D. C. Mutz. “Mechanisms of momentum: does thinking make it so?” In: *The Journal of Politics* 59.1 (1997), pp. 104–125.

22. J. Großer & A. Schram. “Public opinion polls, voter turnout, and welfare: An experimental study”. In: *American Journal of Political Science* 54.3 (2010), pp. 700–717.

23. W. Krause & C. Gahn. “Should we include margins of error in public opinion polls?” In: *European Journal of Political Research* 63.3 (2023), pp. 1082–1107. DOI: 10.1111/1475-6765.12633.

24. S. J. Westwood, S. Messing & Y. Lelkes. “Projecting confidence: How the probabilistic horse race confuses and demobilizes the public”. In: *The Journal of politics* 82.4 (2020), pp. 1530–1544.

25. F. Yang, M. Cai, C. Mortenson, H. Fakhari, A. D. Lokmanoglu, J. Hullman, S. Franconeri, N. Diakopoulos, E. C. Nisbet & M. Kay. “Swaying the Public? Impacts of Election Forecast Visualizations on Emotion, Trust, and Intention in the 2022 U.S. Midterms”. In: *IEEE Transactions on Visualization and Computer Graphics* 30.1 (2024), pp. 23–33. DOI: 10.1109/TVCG.2023.3327356.

26. L. M. Padilla, S. H. Creem-Regehr, M. Hegarty & J. K. Stefanucci. “Decision making with visualizations: a cognitive framework across disciplines”. In: *Cognitive Research: Principles and Implications* 3.1 (2018), p. 29. DOI: 10.1186/s41235-018-0120-9.

27. I. McAllister & D. T. Studlar. “Bandwagon, underdog, or projection? Opinion polls and electoral choice in Britain, 1979-1987”. In: *The Journal of Politics* 53.3 (1991), pp. 720–741.

28. M. Barnfield. “Momentum in the polls raises electoral expectations”. In: *Electoral Studies* 84 (2023), p. 102656. DOI: 10.1016/j.electstud.2023.102656.

29. L. F. Stoetzer, L. Leemann & R. Traunmüller. “Learning from Polls During Electoral Campaigns”. In: *Political Behavior* (2024).

30. A. M. van der Bles, S. van der Linden, A. L. J. Freeman, J. Mitchell, A. B. Galvao, L. Zaval & D. J. Spiegelhalter. “Communicating uncertainty about facts, numbers and science”. In: *Royal Society Open Science* 6.5 (2019), p. 181870. DOI: 10.1098/rsos.181870.

31. A. M. van der Bles, S. van der Linden, A. L. J. Freeman & D. J. Spiegelhalter. “The effects of communicating uncertainty on public trust in facts and numbers”. In: *Proceedings of the National Academy of Sciences* 117.14 (2020), pp. 7672–7683. DOI: 10.1073/pnas.1913678117.

32. L. M. Padilla, M. Kay & J. Hullman. “Uncertainty Visualization”. In: *Wiley StatsRef: Statistics Reference Online*. Wiley, 2021. DOI: 10.1002/9781118445112.stat08296.

33. R. R. Lau & D. P. Redlawsk. “Advantages and Disadvantages of Cognitive Heuristics in Political Decision Making”. In: *American Journal of Political Science* 45.4 (2001), pp. 951–971.

34. M. F. Meffert & T. Gschwend. "Polls, coalition signals and strategic voting: An experimental investigation of perceptions and effects". In: *European Journal of Political Research* 50.5 (2011), pp. 636–667.

35. Á. Stefkovics & Z. Kmetty. *Trust in Survey Results. A Cross-country Replication Experiment*. 2023. DOI: 10.31219/osf.io/crwq7.

36. Á. Stefkovics & D. Gere. "Trust in Election Polls. The Role of the Source Institution and the Results of the Poll". In: *International Journal of Public Opinion Research* 37.2 (2025). DOI: 10.1093/ijpor/edaf018.

37. S. Stadtmüller, H. Silber & C. Beuthner. "What Influences Trust in Survey Results? Evidence From a Vignette Experiment". In: *International Journal of Public Opinion Research* 34.2 (2022), edac012. DOI: 10.1093/ijpor/edac012.

38. C. Erfort, L. F. Stoetzer, T. Gschwend, E. Koch, S. Munzert & H. Rajski. "The Zweitstimme Forecast for the German Federal Election 2025: Coalition Majorities and Vacant Districts". In: *PS: Political Science & Politics* (2025), pp. 1–12. DOI: 10.1017/S1049096525000150.

39. M. Prior. *Post-broadcast democracy: How media choice increases inequality in political involvement and polarizes elections*. Cambridge University Press, 2007.

40. P. Van Aelst, J. Strömbäck, T. Aalberg, F. Esser, C. De Vreese, J. Matthes, D. Hopmann, S. Salgado, N. Hubé, A. Stepińska, et al. "Political communication in a high-choice media environment: A challenge for democracy?" In: *Annals of the international communication association* 41.1 (2017), pp. 3–27.

41. E. Link, J. Henke & W. Möhring. "Credibility and enjoyment through data? Effects of statistical information and data visualizations on message credibility and reading experience". In: *Journalism studies* 22.5 (2021), pp. 575–594.

42. P. Mena. "Reducing misperceptions through news stories with data visualization: The role of readers' prior knowledge and prior beliefs". In: *Journalism* 24.4 (2023), pp. 729–748.

43. Y. De Haan, S. Kruikemeier, S. Lecheler, G. Smit & R. Van der Nat. "When does an infographic say more than a thousand words? Audience evaluations of news visualizations". In: *Journalism Studies* 19.9 (2018), pp. 1293–1312.

44. N. J. Stroud. "Media Use and Political Predispositions: Revisiting the Concept of Selective Exposure". In: *Political Behavior* 30.3 (2008), pp. 341–366. DOI: 10.1007/s11109-007-9050-9.

45. D. Steppat, L. Castro Herrero & F. Esser. “Selective Exposure in Different Political Information Environments: How Media Fragmentation and Polarization Shape Congruent News Use”. In: *European Journal of Communication* 37.1 (2022), pp. 82–102. DOI: 10.1177/02673231211012141.

46. D. Chong & J. N. Druckman. “Framing Theory”. In: *Annual Review of Political Science* 10 (2007), pp. 103–126. DOI: 10.1146/annurev.polisci.10.072805.103054.

47. M. A. Cacciatore, D. A. Scheufele & S. Iyengar. “The End of Framing as We Know It . . . and the Future of Media Effects”. In: *Mass Communication and Society* 19.1 (2016), pp. 7–23. DOI: 10.1080/15205436.2015.1068811.

48. A. Neundorf & A. Öztürk. “How to improve representativeness and cost-effectiveness in samples recruited through meta: A comparison of advertisement tools”. In: *PLOS ONE* 18.2 (2023). DOI: 10.1371/journal.pone.0281243.

49. S. Ramirez-Ruiz & S. Munzert. *Measuring scientific evidence consumption literacy for public policy: Development and validation of the INSPIRE inventory*. 2026.

50. L. W. Miratrix, J. S. Sekhon, A. G. Theodoridis & L. F. Campos. “Worth Weighting? How to Think About and Use Weights in Survey Experiments”. In: *Political Analysis* 26.3 (2018), pp. 275–291. DOI: 10.1017/pan.2018.1.

51. A. Deaton & N. Cartwright. “Understanding and misunderstanding randomized controlled trials”. In: *Social Science & Medicine* 210 (2018), pp. 2–21. DOI: 10.1016/j.socscimed.2017.12.005.

Visual presentation raises trust in election polls

Supplementary Information (SI)

Contents

A. Extended Materials and Methods **2**

A.1. Pre-Registration . . . 2

A.2. Deviations from the pre-registration . . . 2

A.3. Participant Recruitment . . . 4

A.3.1. Voter Sample . . . 4

A.3.2. Elite Samples . . . 12

A.4. Experimental Design . . . 15

A.4.1. Construction of Vignettes . . . 16

A.4.2. Outcome Measures . . . 19

A.4.3. Covariate Measures . . . 21

A.4.4. Survey Instrument . . . 23

A.4.5. Context of the electoral campaign and Polling Landscape . . . 26

A.4.6. Statistical Power . . . 30

A.5. Regression Tables . . . 31

A.6. Heterogeneous effects using continuous moderators . . . 35

A.7. Anchoring effect sizes in raw outcome units . . . 39

A.8. False discovery rate correction . . . 42

A.9. Marginal means . . . 44

A.10.Replication Materials . . . 45

B. Media Content Analysis **47**

B.1. Methodological Approach & Selection Criteria . . . 47

B.2. Coding Instructions . . . 48

B.3. Results . . . 50

C. Further Analyses **51**

C.1. Comparing voters to political candidates and journalists . . . 51

C.2. Ranking all 160 poll presentations . . . 53

A. Extended Materials and Methods

A.1. Pre-Registration

This study adheres to a pre-registered research plan that was submitted to the Open Science Framework (OSF) prior to the beginning of data collection on January 24, 2025 (see `https://doi.org/10.17605/OSF.IO/W7QZF`). The pre-registration documents the core elements of the research design, including the data collection strategy, exemplary survey vignettes, and a schematic overview of the experimental features and their planned variation (see `https://osf.io/w7qzf/files/v6exc`).

Our pre-analysis plan registered a single directional hypothesis: that visual depictions of polling results—particularly those displaying historical trends—attenuate desirability bias, i.e., reduce the degree to which favorable results inflate trust in the reported poll. Testing this hypothesis requires the experimentally manipulated party-performance component of the design (whether a respondent's preferred party was shown performing above, below, or at its baseline level). Because that manipulation is the focus of a companion paper, the confirmatory test of the desirability-bias hypothesis is reported there.

The present paper reports a distinct set of estimands: the average marginal component effects of the reporting-design features (format, reference value, uncertainty, threshold, and sample information) together with their heterogeneity across respondent subgroups. Our pre-analysis plan committed us ex ante to estimating these effects. It specifies fully specified AMCE models for all attribute levels as well as interactions with respondent characteristics, but registered no directional predictions for them. Consistent with our pre-registered commitment to flag such analyses as exploratory, we treat the main-effect and heterogeneity results reported here as exploratory throughout. As stated in the manuscript, the analytic stance concerns the specification of directional hypotheses only; it does not affect the causal interpretation of the estimates, which follows from the randomized assignment of vignette features.

A.2. Deviations from the pre-registration

While the implementation of the study closely followed the pre-registered protocol, a small number of deviations from the original plan proved necessary. Importantly, none of these deviations alter the fundamental structure of the design or the logic of inference. In the interest of full transparency and replicability, we document all deviations from the pre-registration below:

1. **Measurement**: In order to reduce survey length within acceptable limits, several items

were removed. Consequently, only three of the six originally planned scenarios designed to measure the effectiveness of poll presentations in conveying realistic expectations about party competition were retained. Specifically, respondents were not asked to estimate the likelihood of scenarios in which (a) BSW or (b) Die Linke received less than five percent of the vote, despite these items being specified in the pre-registration.

2. **Response Scales**: The pre-analysis plan specified that the trust and scenario-probability items would be collected on 0–10 scales. In the fielded instrument, these items were instead measured using 0–100 sliders (see SI Section A.4.4). This change concerns only the granularity of the response format and not the substantive content of the items; because all outcomes are standardized prior to analysis, it does not affect the interpretation of the reported effects.

3. **Outcome Scope**: The pre-analysis plan listed a fourth block of outcomes capturing general attitudes towards election polls and polling, to be aggregated into a separate index with its own control-group design. These items are not analyzed in the present paper, which focuses on the three vignette-specific outcome dimensions (attractiveness, trust, and expectation quality). They are not analyzed as part of this project.

4. **Analysis**: We do not report analyses on the experimentally manipulated performance of the respondent's preferred party. These results are analytically distinct from the present contribution and will be reported in a separate paper.

5. **Outcome Construction**: The pre-analysis plan listed five perception items; the fielded instrument contained four (SI Section A.4.2), and the item on the amount of information is excluded from the attractiveness index because its endpoints do not capture a positive or negative evaluation (see *Dimensionality of the outcome indices*). The wording of the second trust item also differs slightly from the pre-registered version.

6. **Inference**: Standard errors are clustered by respondent, and we add Benjamini–Hochberg false discovery rate correction as a robustness check (SI Section A.8).

A.3. Participant Recruitment

A.3.1. Voter Sample

Between January 24 and February 21, 2025, 22,976 participants took part in the survey. Of these, 10,564 were recruited via an online access panel (Bilendi) and 12,094 through social media ads on Facebook and Instagram; for the remaining 318, the vignette ratings could not be linked to their survey responses, and they are excluded from the analyses (see *Materials and Methods*). Both groups of participants were directed into the same survey hosted using Qualtrics. The majority of responses were collected by February 12, 2025.

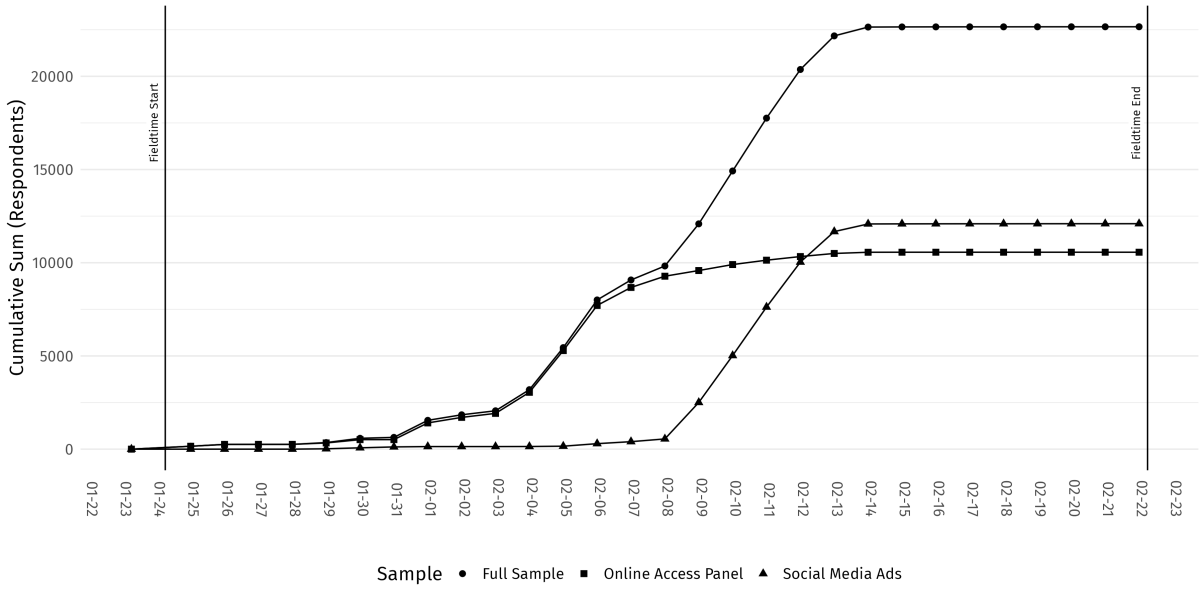


Figure A1: Cumulative sum of completed participations over the course of the field time.

As depicted in Figure A1, the majority of the online access panel observations were collected in the first half of the field period, while the social media ads recruited the majority of its participants in the second. In this section, we outline the recruitment strategies in both contexts and describe the sample characteristics.

Recruitment Strategies The online access panel component of the sample was drawn from the provider *Bilendi*. The recruitment strategy applied gender quotas and required a minimum of 20 respondents from each of the 299 electoral districts in the 2025 German federal election.

The social media component of the sample was recruited following the best-practice principles proposed by (1, 2). In the Facebook Ad Manager, we set up 64 conversion campaigns, enabling us to target every combination of the following three features: German state (s), gender (g), and age (a).

$$Budget_{s,g,a} = \frac{\frac{\text{German Citizens}_s}{71{,}761{,}258}}{4} \tag{1}$$

The total campaign budget was allocated to states in proportion to their share of the German population (Equation 1). Within each state, the budget was distributed evenly across the four gender–age groups. The budget was reset each day, and recruitment restarted until the daily spending limit was again reached.

Table A1: Allocation of resources across states

State (s)	**Population$_s$**	**German Citizens$_s$**	**Budget Share$_s$**
Baden-Württemberg	11,339,260	9,239,843	0.13
Bayern	13,435,062	11,278,833	0.16
Berlin	3,782,202	2,901,363	0.04
Brandenburg	2,581,667	2,387,544	0.03
Bremen	691,703	540,345	0.01
Hamburg	1,910,160	1,527,631	0,02
Hessen	6,420,729	5,173,792	0,07
Mecklenburg-Vorpommern	1,629,464	1,515,144	0.02
Niedersachsen	8,161,981	7,156,757	0.10
Nordrhein-Westfalen	18,190,422	15,253,631	0.21
Rheinland-Pfalz	4,174,311	3,582,879	0.05
Saarland	994,424	847,490	0.01
Sachsen	4,089,467	3,757,065	0.05
Sachsen-Anhalt	2,180,448	2,006,167	0.03
Schleswig-Holstein	2,965,691	2,646,955	0.04
Thüringen	2,122,335	1,945,819	0.03
Germany	84,669,326	71,761,258	1

Ad Design Each of the 64 *state* × *gender* × *age* brackets was supplied with four different ad designs which the algorithm could vary freely to optimize conversions (i.e. survey completions). The Meta algorithm optimizes the use of ads for maximum conversion. We ensured that the design of the ads was kept as neutral as possible (in line with recommendations by 3) and that the ads did not differ too greatly in this respect.



Figure A2: Wide Format Ads



Figure A3: Square Format Ads



Figure A4: Story Format Ads

Sample Characteristics Figure A2 compares key demographic characteristics across the pooled sample, the separate social media and Bilendi components, and benchmark values derived from German census data and the Federal Returning Officer (Bundeswahlleiterin).

			Sample Share Compared to Benchmark		
	Group	**Benchmark**	**Full Sample**	**Bilendi Sample**	**Social Media Sample**
Age	15-24 y.o.	11.7%	6.5% (-5.2 pp)	4.4% (-7.3 pp)	8.4% (-3.3 pp)
	25-34 y.o.	14.4%	15.5% (+1.0 pp)	14.4% (-0.1 pp)	16.4% (+2.0 pp)
	35-44 y.o.	15.5%	19.5% (+4.0 pp)	21.7% (+6.2 pp)	17.5% (+2.0 pp)
	45-54 y.o.	14.6%	17.8% (+3.2 pp)	19.2% (+4.6 pp)	16.5% (+2.0 pp)
	55-64 y.o.	18.3%	26.0% (+7.6 pp)	26.1% (+7.8 pp)	25.9% (+7.5 pp)
	65 y.o. or older	25.5%	14.8% (-10.8 pp)	14.2% (-11.3 pp)	15.2% (-10.3 pp)
Gender	Female	50.9%	48.4% (-2.4 pp)	51.9% (+1.1 pp)	45.4% (-5.5 pp)
	Male	49.1%	51.6% (+2.4 pp)	48.1% (-1.1 pp)	54.6% (+5.5 pp)
Education	High	36.6%	61.1% (+24.5 pp)	53.3% (+16.7 pp)	68.0% (+31.4 pp)
	Medium	29.5%	30.6% (+1.1 pp)	35.9% (+6.4 pp)	26.0% (-3.5 pp)
	Low	30.3%	7.5% (-22.9 pp)	10.4% (-20.0 pp)	4.9% (-25.4 pp)
	Still in school	3.6%	0.8% (-2.8 pp)	0.4% (-3.1 pp)	1.1% (-2.4 pp)
Turnout	Yes	82.5%	93.7% (+11.2 pp)	88.0% (+5.5 pp)	98.7% (+16.2 pp)
Vote Choice	CDU/CSU	28.6%	17.3% (-11.3 pp)	21.2% (-7.4 pp)	13.9% (-14.7 pp)
	AfD	20.8%	23.8% (+3.0 pp)	19.9% (-0.9 pp)	27.2% (+6.4 pp)
	SPD	16.4%	10.6% (-5.8 pp)	14.3% (-2.1 pp)	7.4% (-9.0 pp)
	Greens	11.6%	19.4% (+7.8 pp)	13.2% (+1.6 pp)	24.7% (+13.1 pp)
	Left Party	8.8%	10.8% (+2.0 pp)	6.1% (-2.7 pp)	14.9% (+6.1 pp)
	BSW	4.9%	4.2% (-0.7 pp)	6.2% (+1.3 pp)	2.4% (-2.5 pp)
	FDP	4.3%	4.1% (-0.2 pp)	3.7% (-0.6 pp)	4.4% (+0.1 pp)
	Other	4.5%	9.9% (+5.4 pp)	15.5% (+11.0 pp)	5.1% (+0.6 pp)
N (respondents)			22,658	10,564	12,094

Table A2: Sample composition compared to population benchmarks. Shares for the pooled sample and the two subsamples, with the difference from the benchmark in percentage points (pp) in parentheses. Green cells mark groups over-represented relative to the benchmark, red cells under-represented groups. Age, gender, and education benchmarks are from the microcensus (population aged 15 and older); turnout and vote choice benchmarks are the official results of the 2025 federal election (second vote) from the Federal Returning Officer. Turnout in the sample is the share of respondents rating their likelihood of voting at 9 or 10 on a 10-point scale or reporting that they had already voted by mail.

The comparisons reveal partly substantial discrepancies between the target population and the sample(s). Especially in the social media component of the sample, highly educated respondents and supporters of the Greens and the AfD are over-represented, while respondents aged 65 and older, CDU/CSU voters, and respondents with low formal education are underrepresented (see SI Section A.3.1). At the same time, low levels of education, intention of voting abstention and voters of the CDU/CSU are under-represented in our sample.

The majority of these discrepancies stem from biases in the social media component. On

the one hand, the demographic features our stratification targeting strategy accounted for (i.e., gender and age) only showed marginal deviations which are comparable to this in the online access panel. On the other hand, the deviations from the population benchmarks with regard to levels of education, as well as turnout and voting intentions are up to twice as severe as in the online access panel.

Randomization Statistics Within the voters-only sample, the randomization is well balanced across all experimental dimensions. No condition exhibits substantial over- or under-representation. Importantly, even the smallest experimental formats contain more than 4,500 vignettes, ensuring large cell sizes throughout (Figure A5). This combination of balanced assignment and substantial group sizes indicates that the randomization procedure performed as intended and provides strong statistical power.

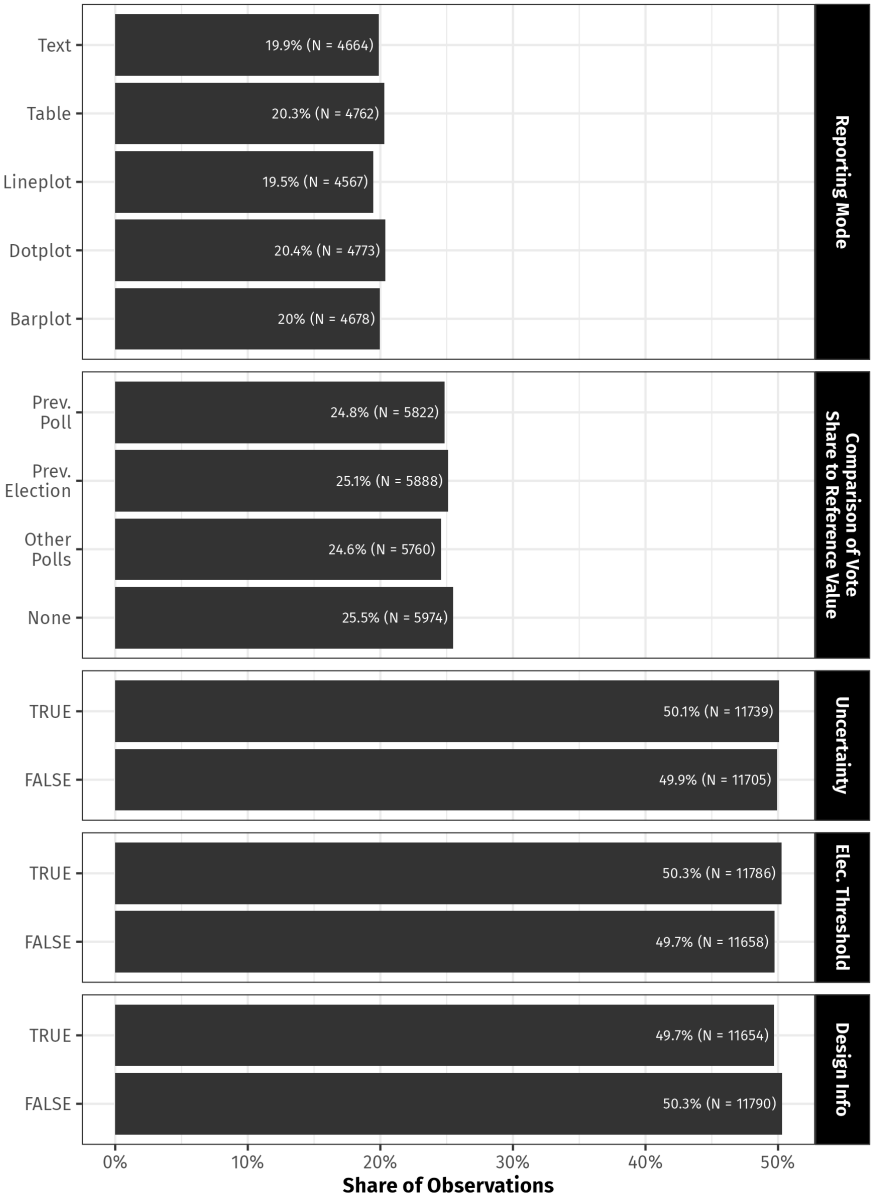


Figure A5: Randomization statistics for voter sample. Shares of voter sample with absolute numbers in parentheses.

Table A3 reports the composition of each experimental condition by age, gender, education, region, and vote intention. The conditions are nearly identical in composition, with differences of at most 2.5 percentage points, which further indicates that the randomization was successful.

Table A3: Respondent characteristics by experimental condition (voters). Cells show the share (%) of rated vignettes in each condition that fall into each category, among vignettes with non-missing values on that characteristic.

Respondent characteristics by experimental condition (voters)																			
		Format					Reference				Uncertainty		Design Info		Threshold		Preferred Party Performance		
		Text	**Table**	**Lineplot**	**Dotplot**	**Barchart**	**None**	**Other Polls**	**Previous Election**	**Previous Poll**	**No**	**Yes**	**No**	**Yes**	**No**	**Yes**	**Real**	**Below**	**Above**
Sample	N (vignettes)	4,664	4,762	4,567	4,773	4,678	5,974	5,760	5,888	5,822	11,705	11,739	11,790	11,654	11,658	11,786	7,765	7,850	7,829
Age	18-24	6.6	6.4	6.5	6.3	7.0	6.1	6.0	6.7	7.3	6.6	6.5	6.4	6.7	6.4	6.7	6.5	6.4	6.7
	25-34	16.2	15.4	15.2	15.2	15.3	16.1	15.0	15.0	15.7	15.1	15.8	15.7	15.3	16.0	14.9	15.2	15.0	16.2
	35-44	19.3	18.8	19.6	20.0	19.8	19.4	20.0	19.7	18.9	19.3	19.7	19.6	19.4	19.2	19.8	19.8	19.7	19.0
	45-54	17.2	17.6	18.3	18.1	17.7	17.2	18.6	17.5	17.8	18.1	17.5	17.7	17.9	17.8	17.8	17.1	17.9	18.4
	55-64	26.2	26.6	25.1	25.6	26.0	26.2	26.2	26.7	24.5	26.1	25.7	26.0	25.8	25.9	25.9	26.3	26.4	25.0
	65+	14.5	15.2	15.4	14.7	14.2	14.9	14.2	14.3	15.7	14.9	14.7	14.5	15.0	14.6	14.9	15.1	14.6	14.7
Gender	Male	50.5	51.4	50.2	52.7	51.4	51.1	51.1	51.1	51.7	51.0	51.5	50.6	52.0	51.4	51.1	51.0	51.6	51.2
	Female	49.1	47.8	49.2	46.7	48.0	48.4	48.3	48.4	47.5	48.4	47.9	48.8	47.5	47.9	48.4	48.4	47.8	48.3
	Diverse	0.4	0.8	0.6	0.6	0.6	0.5	0.7	0.5	0.8	0.6	0.5	0.6	0.6	0.7	0.5	0.6	0.6	0.5
Education	Low	7.2	7.4	7.7	7.5	7.1	7.5	7.2	7.3	7.5	7.4	7.4	7.1	7.6	7.3	7.5	7.4	7.7	7.0
	Medium	30.4	30.9	31.6	31.1	30.3	30.8	30.5	30.3	31.7	30.8	30.9	30.4	31.3	30.7	31.0	30.6	31.6	30.4
	High	61.7	60.8	59.9	60.6	61.9	61.0	61.4	61.6	59.9	61.1	60.9	61.7	60.2	61.2	60.8	61.2	59.8	61.9
	Still in school	0.7	0.9	0.9	0.8	0.7	0.7	0.8	0.8	0.9	0.7	0.9	0.7	0.9	0.8	0.8	0.8	0.8	0.7
Region	West	77.1	77.8	76.6	76.7	77.3	76.4	77.8	77.5	76.8	76.8	77.3	77.2	77.0	76.9	77.3	76.8	77.7	76.7
	East (incl. Berlin)	22.9	22.2	23.4	23.3	22.7	23.6	22.2	22.5	23.2	23.2	22.7	22.8	23.0	23.1	22.7	23.2	22.3	23.3
Vote Intention	CDU/CSU	17.1	17.5	17.4	16.6	17.1	16.9	17.9	17.3	16.4	16.9	17.4	16.7	17.6	17.2	17.1	17.2	16.8	17.4
	AfD	23.9	24.1	24.1	24.7	23.7	23.5	24.1	23.6	25.2	24.2	23.9	24.0	24.2	23.9	24.2	23.3	24.8	24.1
	SPD	10.5	10.5	10.5	10.9	10.6	10.6	10.2	11.0	10.5	10.6	10.5	10.8	10.4	10.6	10.6	10.6	10.4	10.8
	Greens	19.8	19.2	19.1	18.8	19.4	19.6	19.5	18.9	19.1	19.0	19.6	19.7	18.9	19.1	19.4	19.5	19.4	18.9
	Left Party	10.8	10.6	10.8	10.8	11.2	11.1	10.6	11.0	10.6	10.9	10.7	10.8	10.9	10.9	10.7	11.1	10.3	11.1
	BSW	4.2	4.4	4.8	4.1	4.1	4.1	4.3	4.5	4.3	4.4	4.3	4.4	4.3	4.4	4.2	4.4	4.2	4.4
	FDP	4.1	3.7	4.1	4.4	4.1	4.5	3.9	3.9	4.1	4.2	4.0	4.1	4.1	4.1	4.1	4.0	4.4	3.9
	Other	9.6	10.0	9.2	9.7	9.9	9.8	9.5	9.8	9.6	9.8	9.6	9.7	9.7	9.8	9.6	9.9	9.7	9.4

Cells show the share (%) of rated vignettes in each condition that fall into each category, among vignettes with non-missing values on that characteristic. N = 23,444 vignettes from 22,976 respondents.

A.3.2. Elite Samples

Recruitment Journalists were recruited by email parallel to the voter study. To construct the sampling frame, we began from the population of German regional newspapers, identified through two sources: entries in Wikidata and the Wikipedia listing of German regional newspapers (`https://de.wikipedia.org/wiki/Liste_deutscher_Regionalzeitungen`, accessed on November 07, 2024). We focused on regional rather than national outlets because they constitute the large majority of print titles in Germany and employ the editors most likely to report polling during a campaign. Four research assistants visited the website of each outlet and collected all publicly available editorial contact addresses, prioritizing politics, domestic-affairs, and editorial desks over general or subscription inboxes where a distinction could be drawn. In total, they screened 547 outlet websites and compiled 2,590 individual contacts. We sent an invitation email to all 2,590 contacts, containing a short description of the study and a personalized link to the survey. Of these, 298 provided complete responses, yielding a response rate of 11.5% (298/2,590). The poll-reporting instrument analyzed here was fielded to a randomly assigned subset of participants, comprising 110 of the completed responses. After excluding two respondents with incomplete vignette data, the final journalist sample comprises 108 respondents (see SI Section C.1 for the analytic sample).

Political candidates were recruited via email as well. Due to the snap election, the nomination deadline for candidates was quite late in the campaign (January 28). Thus, student assistants searched for the publicly available names and e-mail addresses of district candidates from major parties. Once the nominations were officially announced, the collected names were verified. We included all parties represented in parliament, as well as the BSW. The study was in the field from February 11 to 22. Unfortunately, it was not possible to retrieve email addresses from all candidates. Overall, 1,414 e-mails were sent. We asked the candidates or their staff members to fill in a short survey. We received 268 responses, which translates into a response rate of 18.9%. 38 of them (14%) were filled in by members of staff. Response rates differed somewhat across parties: from the AfD, 35 of 187 invited candidates responded (18.7%; 295 candidates total); from B'90/Die Grünen, 54 of 246 (22.0%; 297 total); from the BSW, 2 of 14 (14.3%; 35 total); from the CDU, 28 of 198 (14.1%; 252 total); from the CSU, 2 of 46 (4.35%; 47 total); from Die Linke, 54 of 207 (26.1%; 297 total); from the FDP, 48 of 238 (20.2%; 299 total); and from the SPD, 45 of 278 (16.2%; 299 total).

Of the 268 respondents, 143 candidates completed the poll-reporting experiment and enter the analyses reported here—a few more than the number who also completed the remainder of the survey.

Sample Distribution Figure A6 shows the party affiliation of the 143 candidates in the experimental sample. While the overall sample size is modest, the major German parties are well balanced: each party except the BSW is represented by 20 to 30 candidates.

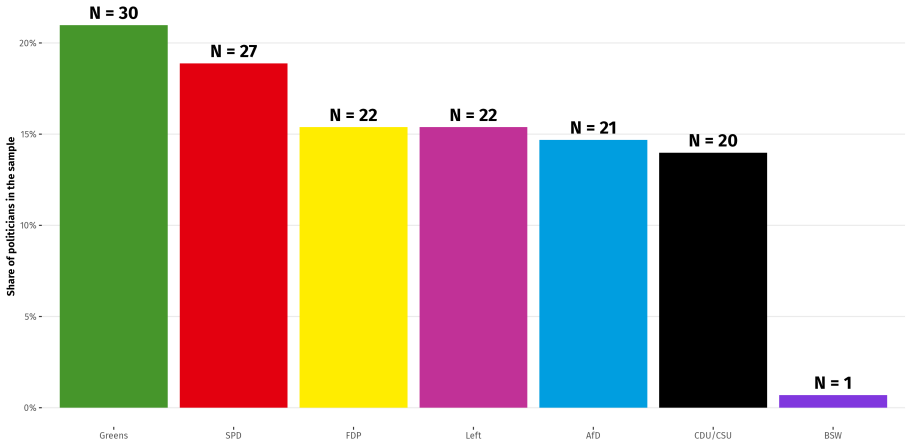


Figure A6: Distribution of Candidates across parties.

The majority of journalists in our sample are based in the major federal states—Lower Saxony, Bavaria, Baden-Württemberg, and North Rhine–Westphalia (see Figure A7). These are both the most populous states and those with particularly large numbers of print journalists, making this distribution neither surprising nor problematic, as it likely mirrors the overall population of print editors in Germany. Berlin is the only notable exception, appearing to be substantially under-represented in the sample.

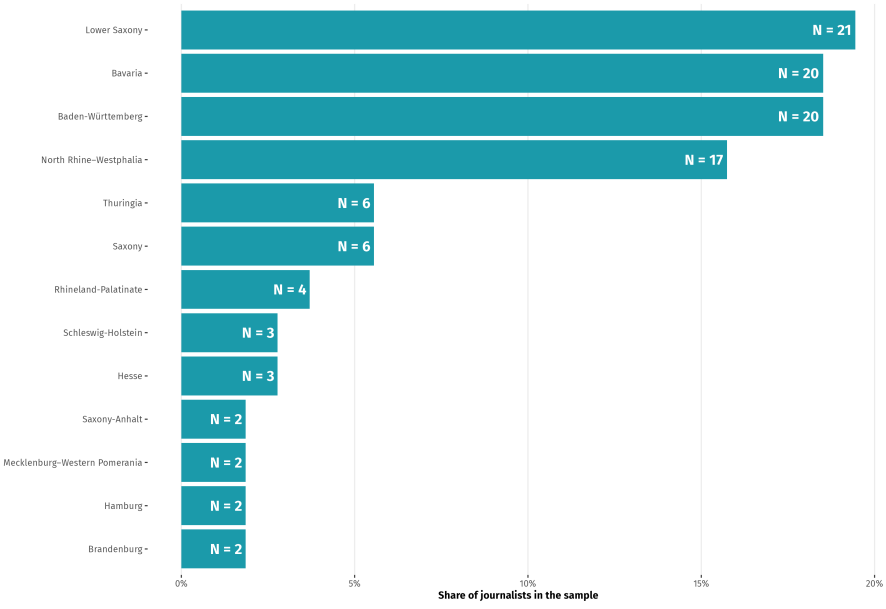


Figure A7: Distribution of journalists across German states.

Randomization Statistics Although case numbers in the elite samples are, as expected, not comparable to those in the voter sample, we also observe balanced distributions across experimental treatment conditions. This shows that randomization was successful and mitigates concerns about systematic imbalances that could bias treatment effect estimates.

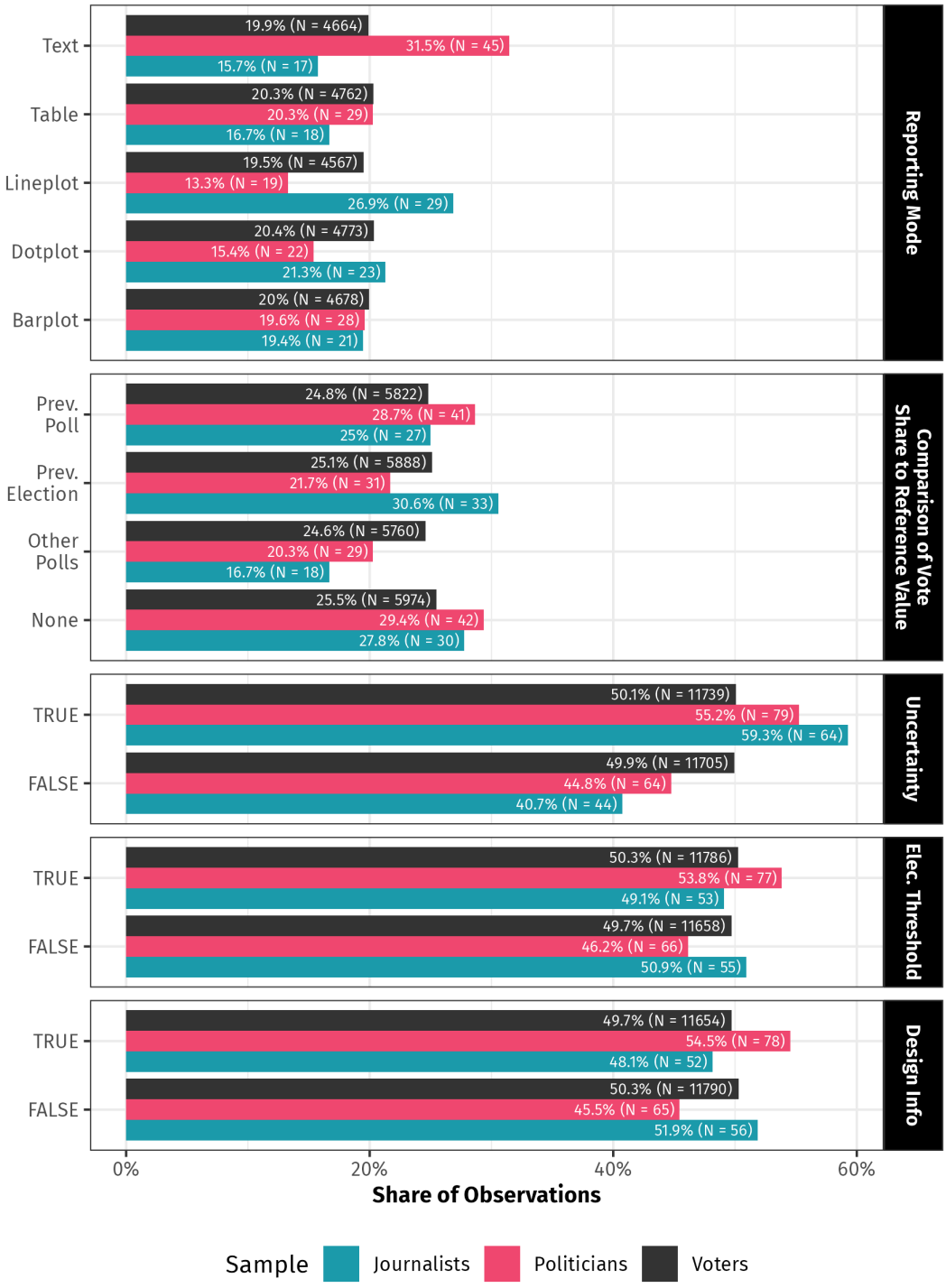


Figure A8: Randomization statistics for all three samples. Shares of the respective treated subsample with absolute numbers in parentheses.

A.4. Experimental Design

This section provides an overview of the experimental setup of this study. Within the experimental environment, respondents were presented with a randomly assigned vignette depicting a reporting of an election poll (see Figure A9 and SI Section A.4.1). Below the vignette, we measured three groups of outcomes (SI Section A.4.2). For a documentation of the implementation of this design in Qualtrics, please refer to SI Section A.4.4.

Data collection took place between January 24 and February 21, 2025. There might be the concern that polls could be trusted less due to the potential for changes in the poll during field time. As depicted in SI Section A.4.5, however, the polls were highly stable during field time.

Calculations regarding statistical power are presented in SI Section A.4.6.

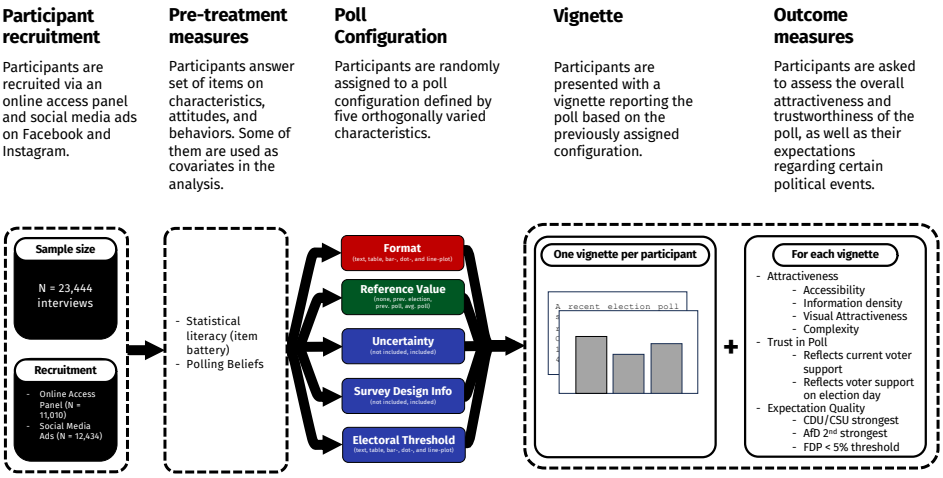


Figure A9: Overview of study design. Subjects were randomly assigned to a random poll configuration varying the format, the reference value, as well as the inclusion of statistical uncertainty, survey design information, and the electoral threshold. Respondents were presented a poll vignette based on these five characteristics. Below the vignette, they were asked to indicate their opinion regarding the attractiveness of the poll, the level of trust they place in it, and how likely they perceived certain political scenarios based on the information provided.

A.4.1. Construction of Vignettes

We used the election poll from the polling firm *Forschungsgruppe Wahlen*, published on 10 January 2025 as the baseline poll. This poll reported the distribution of voting intentions summarized in Table A4.

Party	**Vote Share**
SPD	14%
CDU/CSU	30%
AfD	21%
Greens	15%
BSW	4%
Left	4%
FDP	4%
Other	8%

Table A4: Election poll from *Forschungsgruppe Wahlen*, published on January 10, 2025 (Fieldtime: January 7-9, 2025; N = 1,433)

The depiction was varied along five types of characteristics, namely the format as well as the inclusion of reference values, uncertainty, the electoral threshold, and sample information.

We created 160 vignettes representing the same election poll in different ways.

1. **Format**: Respondents were randomly assigned to one of five reporting formats: a purely textual description, a table, or one of three graphical formats (bar plot, dot plot, or line plot). Each format conveyed the same underlying vote-share information, but differed in visual structure and cognitive demands, allowing us to isolate the effect of presentation mode on respondents' perceptions.

2. **Reference Value**: We further varied the presence and type of reference values used to contextualize the reported vote shares. Depending on the condition, respondents either saw no reference value or were provided with a comparative benchmark, such as the party's result in the previous election, the current average across recent polls, or the result of the previous poll conducted by the same firm. This manipulation allows for assessing how comparative cues shape the interpretation of polling information.

3. **Uncertainty**: We varied whether uncertainty information was included in the poll presentation. In vignette configurations where this was the case, reported vote shares were accompanied by margins of error (± 2 percentage points), which were naturally integrated

into the display depending on the reporting format. In other conditions, uncertainty information was omitted, mirroring common media practices in which point estimates are presented without explicit statistical context.

4. **Electoral Threshold**: The vignettes further differed with respect to the inclusion of the electoral threshold. In conditions where the threshold was shown, respondents were explicitly presented with the minimum vote share required for parliamentary representation. In other versions, this institutional benchmark was omitted, requiring respondents to interpret the reported vote shares without direct reference to electoral viability.

5. **Sample Information**: Finally, we manipulated whether information about the poll's sample was provided. Some vignettes included basic methodological details about the sample on which the poll was based, such as its size and composition, while others presented the polling results without any sample information. This variation captures differences in how transparently polling organizations communicate the basis of their estimates.

The twenty conditions where only one of the four latter options were varied in each of the five different formats are depicted in Figure A10. The vignettes were generated in R using *ggplot2* (4). The script replicating the data generation can be found in the replication archive of this article.

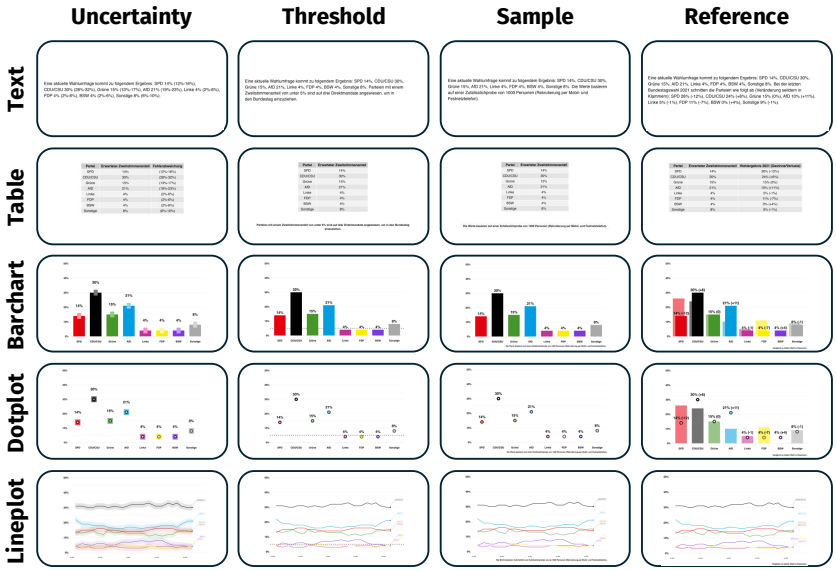


Figure A10: The figure illustrates the main experimental dimensions: reporting formats (rows), technical features (columns 1–3), and reference values (right column). The twenty examples shown represent only a subset of the full design. In total, the experiment includes 160 unique vignettes, generated by orthogonally varying all formats, technical features, and four reference categories (of which only one is displayed here for illustration).

As highlighted in the manuscript and in section elaborating on the pre-registration (Section A.1), note that for each of the seven major parties (SPD, CDU/CSU, AfD, Greens, Left, BSW, FDP) in this election campaign, there was another varied dimension where the preferred party either performed well above, well below, or as reported in the baseline poll.

A.4.2. Outcome Measures

To assess how voters assess the respective polls, we collected three groups of outcomes. For a more detailed elaboration on the collection of these items and the design of the survey instrument, please refer to SI Section A.4.4.

- **Attractiveness Rating**: The first group of outcomes describes the overall attractiveness of a poll. "Die Darstellung..." ("The depiction...")
  - enthält deutlich zu viele Informationen ↔ enthält deutlich zu wenig Informationen
    - entails too much information ↔ entails too little information
  - ist ansprechend gestaltet ↔ ist nicht ansprechend gestaltet
    - is designed attractively ↔ is not designed attractively
  - ist sehr hilfreich ↔ ist überhaupt nicht hilfreich
    - is very helpful ↔ is not helpful at all
  - ist sehr einfach zu verstehen ↔ ist sehr schwer zu verstehen
    - is very easy to understand ↔ is very hard to understand
- **Trust in Poll Result**: The second pair of outcomes describes the level of trust respondents place in the poll's ability to reflect correct information.
  - Was denken Sie, wie stark stimmen die berichteten Umfragewerte der unterschiedlichen Partein mit den aktuellen Stimmenanteilen in der Wahlbevölkerung überein?
    - How well do you think the reported poll numbers for the various parties correspond to the current share of votes among the electorate?
  - Was denken Sie, wie stark stimmen die berichteten Umfragewerte der unterschiedlichen Parteien mit dem von Ihnen erwarteten Ausgang der Bundestagswahl überein?
    - How well do you think the reported poll results for the various parties correspond to your expectations for the outcome of the federal election?
- **Expectation Quality**: The final set of items captures how likely participants consider certain political events based on the information provided by the poll. Wenn Sie die Informationen aus der Umfrage berücksichtigen, für wie wahrscheinlich halten Sie die folgenden Szenarien in der Bundestagswahl 2025? (Taking into account the information from the survey, how likely do you think the following scenarios are in the 2025 general election?)

- Die CDU/CSU wird stärkste Kraft
  - The CDU/CSU becomes the strongest party
- Die AfD wird zweitstärkste Kraft
  - The AfD becomes second strongest party
- Die FDP erhält weniger als 5% der Stimmen
  - The FDP receives less than 5% of the votes

We constructed three composite outcome measures corresponding to the conceptual dimensions outlined above. For the attractiveness dimension, we aggregated the three evaluative items (excluding the item on the amount of information) into a single index using Principal Component Analysis (PCA), which captures the common variance underlying respondents' evaluations of the poll's design, clarity, and informational value.

An analogous PCA-based procedure was applied to the trustworthiness items, yielding a unified measure of respondents' confidence in the poll's ability to accurately reflect both current electoral support and the expected election outcome.

Finally, to evaluate the accuracy of respondents' expectations inferred from the presented poll, we benchmarked their assessments against contemporaneous predictions from the election forecasting model developed by Erfort et al. [5]. Specifically, we calculated the model-implied probabilities for each scenario based on the share of simulation draws as of January 10, 2025—the day the poll was published. Respondent performance was then summarized using the mean absolute error across the three expectation items, providing a continuous measure of how closely individual judgments aligned with model-based forecasts.

Dimensionality of the outcome indices We retain a single component for both outcome indices (Table A5). For the attractiveness index, only the first eigenvalue exceeds one (1.81). This follows the pre-registered eigenvalue criterion and is confirmed by a parallel analysis, where the 95th percentile of eigenvalues from random data of the same size is 1.02. The first component explains 60% of the variance, all three items load on it between 0.74 and 0.80, and Cronbach's α is 0.67. The rating item on the amount of information is excluded because its endpoints capture too much vs. too little information rather than a positive or negative evaluation. Empirically, it is essentially uncorrelated with the other three ratings (r between -0.01 and 0.10). With this item included, a second component only marginally exceeds the eigenvalue criterion (1.01) and is defined entirely by this single item (loading 0.99), so it does not represent a common dimension. For the trust index, the two items correlate at $r = 0.76$; the first component has an

eigenvalue of 1.76 and explains 88% of the variance ($\alpha = 0.86$). Both outcomes are therefore well summarized by a single component.

Table A5: Eigenvalues of the item correlation matrices.

Eigenvalues of the item correlation matrices					
	N	Component 1	Component 2	Component 3	Component 4
Attractiveness (3 items)	22,852	1.81	0.65	0.54	
Attractiveness incl. information item (4 items)	22,823	1.81	1.01	0.65	0.53
Trust (2 items)	23,192	1.76	0.24		

Eigenvalues of the item correlation matrices underlying the PCAs, computed on all rated vignettes of voters and elites. Components with eigenvalues above one are retained. The information item is excluded from the attractiveness index.

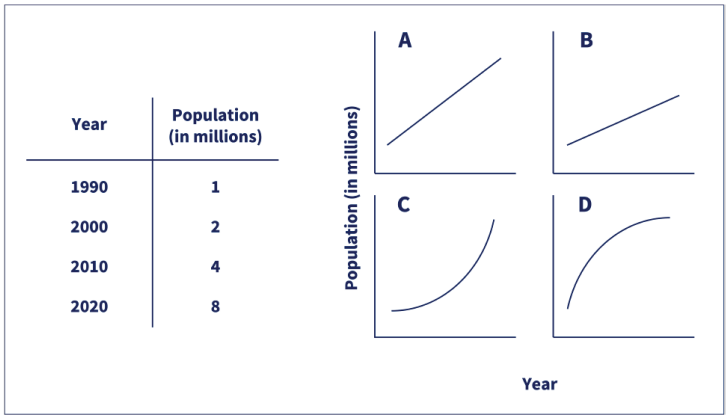
A.4.3. Covariate Measures

Data Literacy To assess respondents' data literacy in the context of election poll consumption, we draw on a set of indicators from the INSPIRE inventory (6). Specifically, we use items DVL4, DVL5, and DVL9 from the INSPIRE inventory.

We select three items from the INSPIRE inventory to capture respondents' capabilities in statistical reasoning and data interpretation. First, item DVL4 (Figure A11) measures respondents' ability to correctly match quantitative information with its corresponding graphical representation. Second, item DVL5 (Figure A12) captures the ability to extract relevant information on political issues from visual data displays. Third, item DVL9 (Figure A13) assesses respondents' stochastic reasoning skills.

For each respondent, we compute the share of correctly answered items among the items answered. For the heterogeneity analyses, respondents with a share of at most one third form the low group and respondents who answered all items correctly the high group; respondents with two of three correct answers are left out (see *Materials and Methods*). The continuous share is used for the formal moderation tests (SI Section A.6).

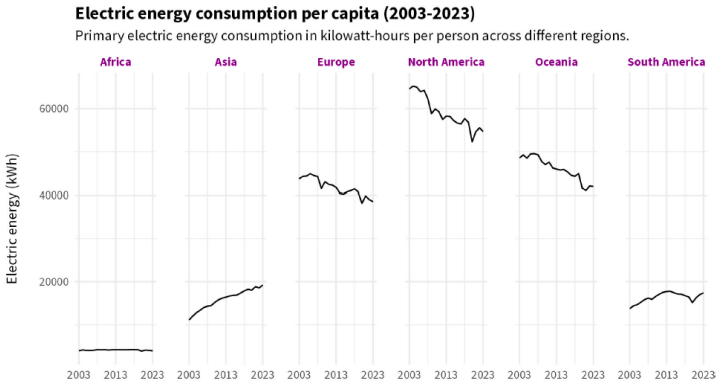
DVL4. Visual Data Representation 1
You have a table tracking population over time. Which graph best represents the data?



A. A
B. B
C. **C ✓**
D. D

Figure A11: DVL4

DVL5. Visual Information 1
Examine the graph. What can be learned?



A. A global decrease in electric energy consumption
B. Asia will surpass Europe in per capita use by 2050
C. **North America is the highest per capita electric consumer ✓**
D. Renewables caused a decrease in consumption in Europe, North America, and Oceania

Figure A12: DVL5

DVL9 (e). Complementary Probability 3
In a city, 40% of the population is vaccinated against a disease. If a person is selected at random, what is the probability that the person IS NOT vaccinated against the disease?

A. 0.2
B. 0.4
C. **0.6 ✓**
D. That is impossible to tell

Figure A13: DVL9

Belief Accuracy To measure how far the presented poll deviates from respondents' prior beliefs about party support, we first elicited, prior to the treatment, respondents' expectations about

the current polling standings of all parties (Figure **??**). The item was implemented in Qualtrics and constrained such that the reported vote shares sum to 100. We then compute the mean absolute deviation between these priors and the vote shares shown in the respondent's vignette across all eight parties (CDU/CSU, SPD, AfD, Greens, FDP, Left, BSW, and Other):

$$\text{Belief Inaccuracy}_{iv} = \frac{1}{8} \sum_{p=1}^{8} \left| \pi_{ip}^{\text{prior}} - \pi_{vp}^{\text{poll}} \right| \tag{2}$$

where π_{ip}^{prior} is respondent i's prior for party p and π_{vp}^{poll} is the share of party p in vignette v shown to respondent i. Smaller values indicate more accurate priors. The measure is computed per vignette, so that respondents who rated two vignettes are compared with the poll shown in each. For the heterogeneity analyses, vignettes in the bottom third of this deviation form the high-accuracy group and those in the top third the low-accuracy group. The continuous models additionally include the square of the deviation to allow for nonlinear moderation (SI Section A.6).

A.4.4. Survey Instrument

The survey instrument consisted of four elements which were consistently displayed in the same order to all participants (for exemplary illustrations, see Figures A14 and A15).

- **Vignette**: The first element of the survey instrument was a vignette presenting participants with an election poll. Vignette configurations were assigned using the Qualtrics *Loop & Merge* functionality, which randomly allocated each participant to a unique URL in a GitHub repository. Each URL pointed to a corresponding `.png` file containing the corresponding vignette, which was then dynamically embedded into the survey interface.

- **Attractiveness Items**: The attractiveness items were collected using a radio button battery. The different items in this battery were displayed in random order.

- **Trust Items**: The degree to which the respondents believed the poll reported the distribution of voting intentions and voter support on election result were collected using sliders ranging between 0 and 100 with labels at every tenth step. To indicate the substantive meaning of the two extremes, the two ends were labeled with "stimmen gar nicht überein" (do not correspond at all) and "stimmen voll und ganz überein" (exactly correspond).

- **Expectation Items**: The three scenario items were collected using the same sliders as for the trust items, where we replaced the labels of the two ends of the scale to "sehr unwahrscheinlich" (very unlikely) and "sehr wahrscheinlich" (very likely) respectively.



Figure A14: Example Vignette presenting a table with no reference values. Neither sample features, nor the electoral threshold or uncertainty is being reported.

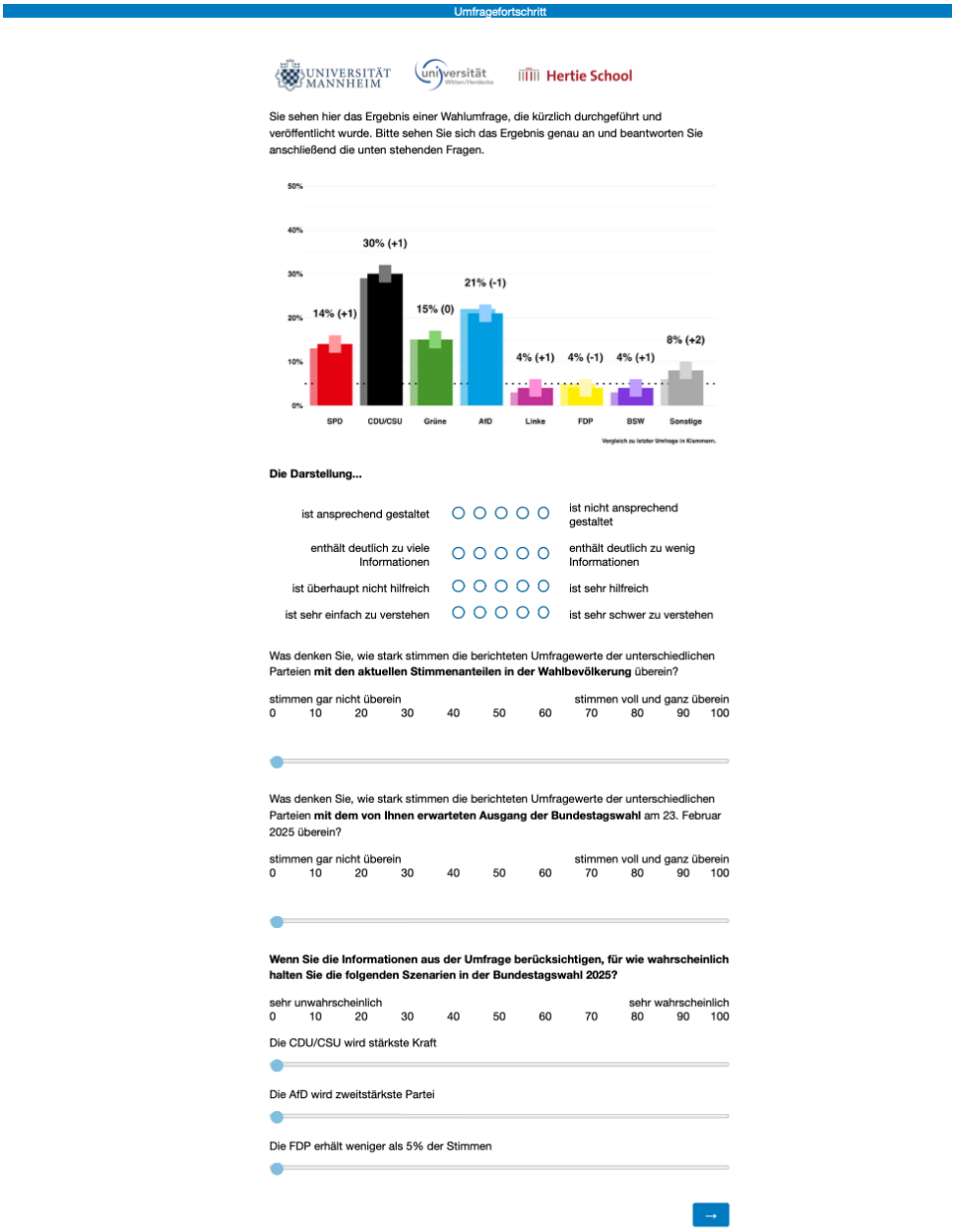


Figure A15: Example Vignette presenting a barchart using the previous poll as a reference value. Sample features are not reported, but there is both a reference to the electoral threshold as well as the uncertainty associated with the presented vote shares.

A.4.5. Context of the electoral campaign and Polling Landscape

The study was conducted during the campaign for the 2025 German federal snap election. The election was called following the collapse of the governing coalition between the Social Democratic Party (SPD), Alliance 90/The Greens, and the Free Democratic Party (FDP), commonly referred to as the *Ampel* (traffic-light) coalition, in November 2024. After disagreements over fiscal and economic policy, Chancellor Olaf Scholz dismissed Finance Minister Christian Lindner (FDP), prompting the FDP to leave the governing coalition. The government subsequently lost a vote of confidence in the Bundestag, leading to the dissolution of parliament and the scheduling of an early federal election for February 23, 2025 (7).

Throughout the field period, public opinion polls consistently indicated that the Christian Democratic Union/Christian Social Union (CDU/CSU) would emerge as the strongest party, followed by the Alternative for Germany (AfD). By contrast, greater uncertainty surrounded smaller parties, particularly the Free Democratic Party (FDP), as well as the Bündnis Sahra Wagenknecht (BSW), whose polling numbers both hovered around the 5% electoral threshold required for parliamentary representation.

The polling performances of the competing parties remained highly stable throughout the field period of the study. The parties' support in the baseline poll of January 10, on which all vignettes were based, was largely maintained in published polls throughout data collection. Although individual polls exhibited minor day-to-day fluctuations, the overall trends suggest that party support changed little over the course of the survey period (see Figure A16).

The only notable exception is the Left Party (*Die Linke*), whose polling support increased by approximately 2–4 percentage points over the observational period. Importantly, however, the vast majority of survey responses were collected by February 12, 2025 (see SI Section A.3). Consequently, most respondents participated while publicly available polling remained highly stable, limiting the likelihood that short-term changes in party support materially influenced the experimental results.

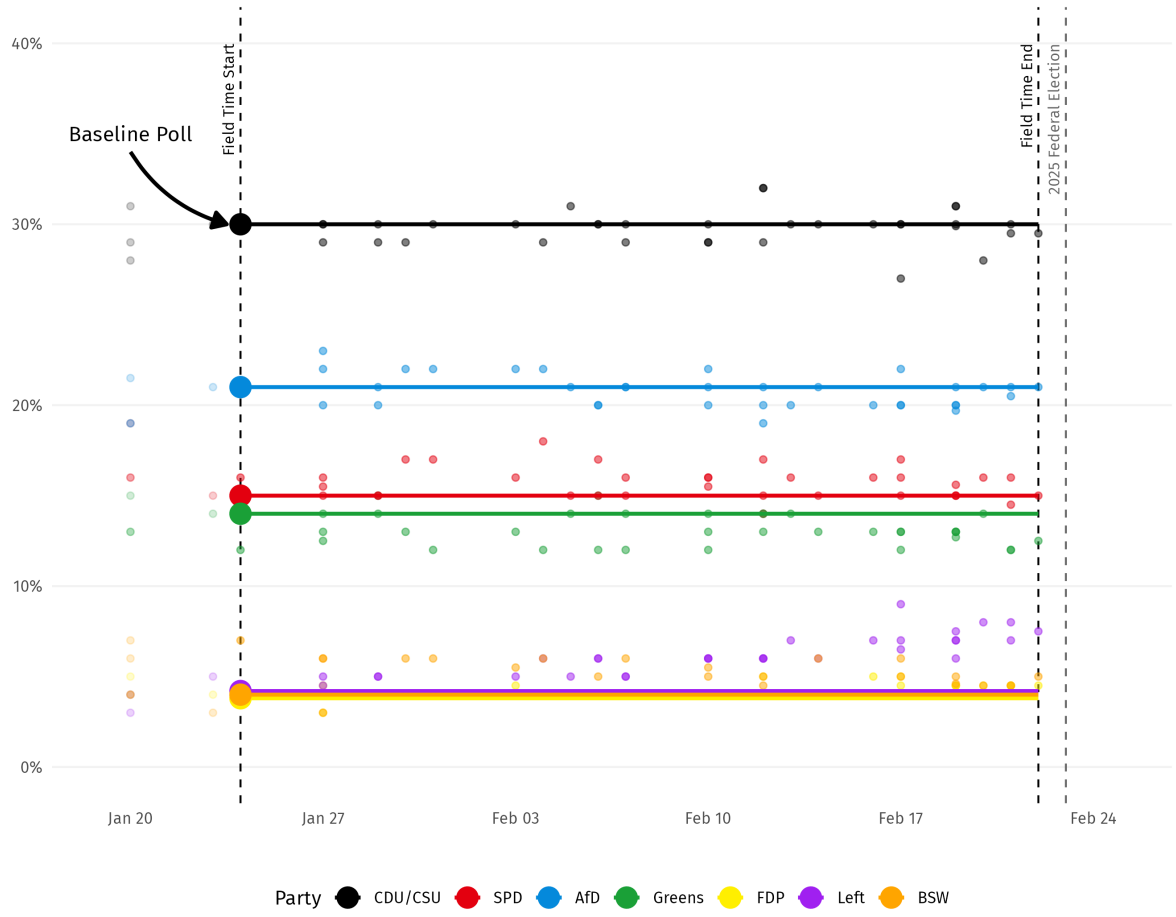


Figure A16: Illustration of polling landscape during field time. Polling data from Politico's *Poll of Polls*.

Heterogeneity over field time We fielded the voter survey between January 24 and February 21, 2025, and almost all responses came in before February 13. The campaign was still running during this time, but the forecast benchmark we use for expectation alignment is fixed (draws from January 10, the day the baseline poll was published). If the polling environment changed while we were in the field, the treatment effects on expectation alignment could depend on when people took part. To check this, we split the field period into three weeks (January 24–30: 681 vignettes; January 31–February 6: 8,602; February 7–21: 13,831), estimated the main models separately for each week, and tested whether the feature × week interactions are jointly zero. One complication is that field week and recruitment channel overlap: most online panel respondents took part in the second week, almost all social media respondents in the third. We therefore also ran the analysis for the online panel alone, which covers all three weeks.

The effects on expectation alignment do not change over the field period, either in the full sample ($F = 1.02$, $p = .43$) or in the online panel ($F = 0.98$, $p = .48$). The same holds for trust ($p = .13$ and $p = .35$). For attractiveness, the format effects differ between weeks in the

full sample ($F = 22.8$, $p < .001$), but not within the online panel ($F = 1.13$, $p = .31$ for all features). We take this to mean that the differences in the full sample come from the change in sample composition across weeks (Figures A17 and A18).

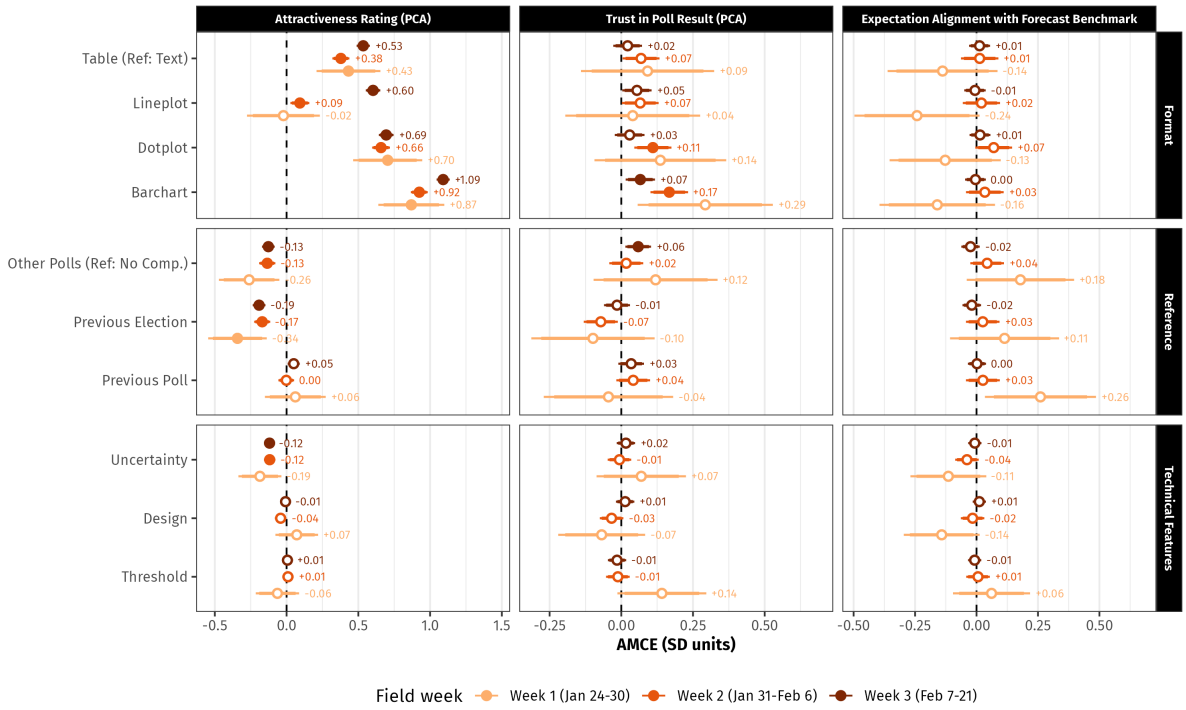


Figure A17: AMCEs by field week, all voters. Estimates in SD units with 95% and 90% confidence intervals. Solid dots mark effects that remain significant after FDR correction.

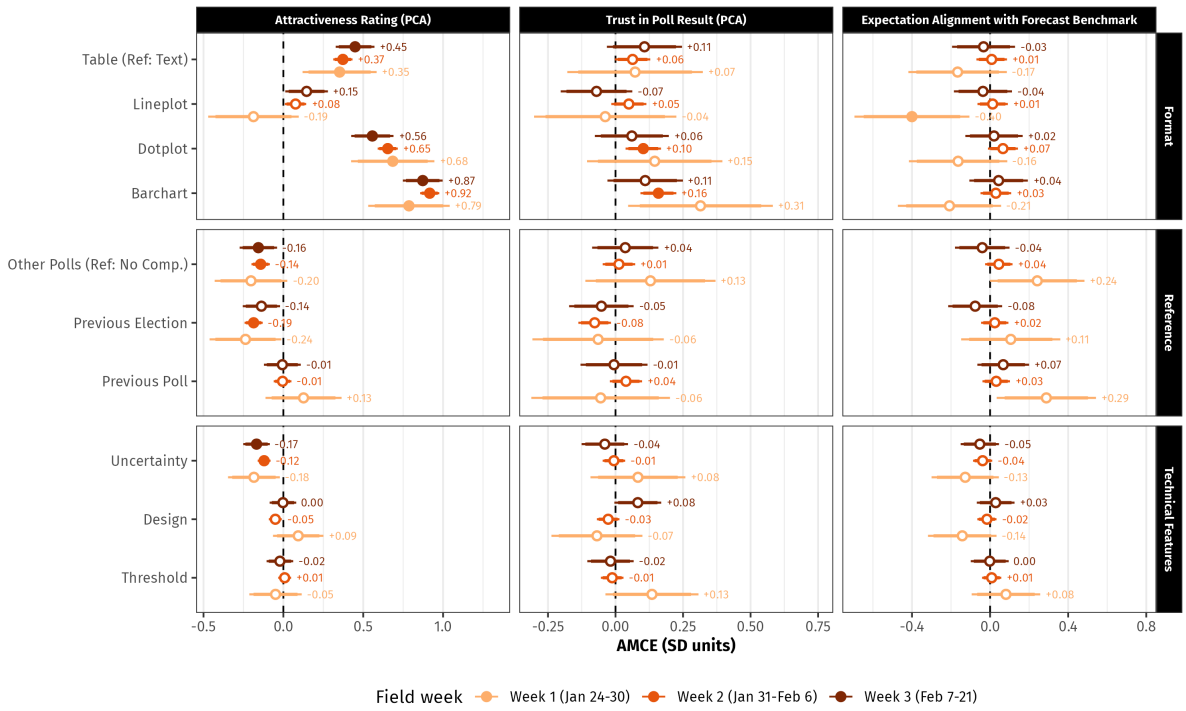


Figure A18: AMCEs by field week, online panel only. Estimates in SD units with 95% and 90% confidence intervals. Solid dots mark effects that remain significant after FDR correction.

A.4.6. Statistical Power

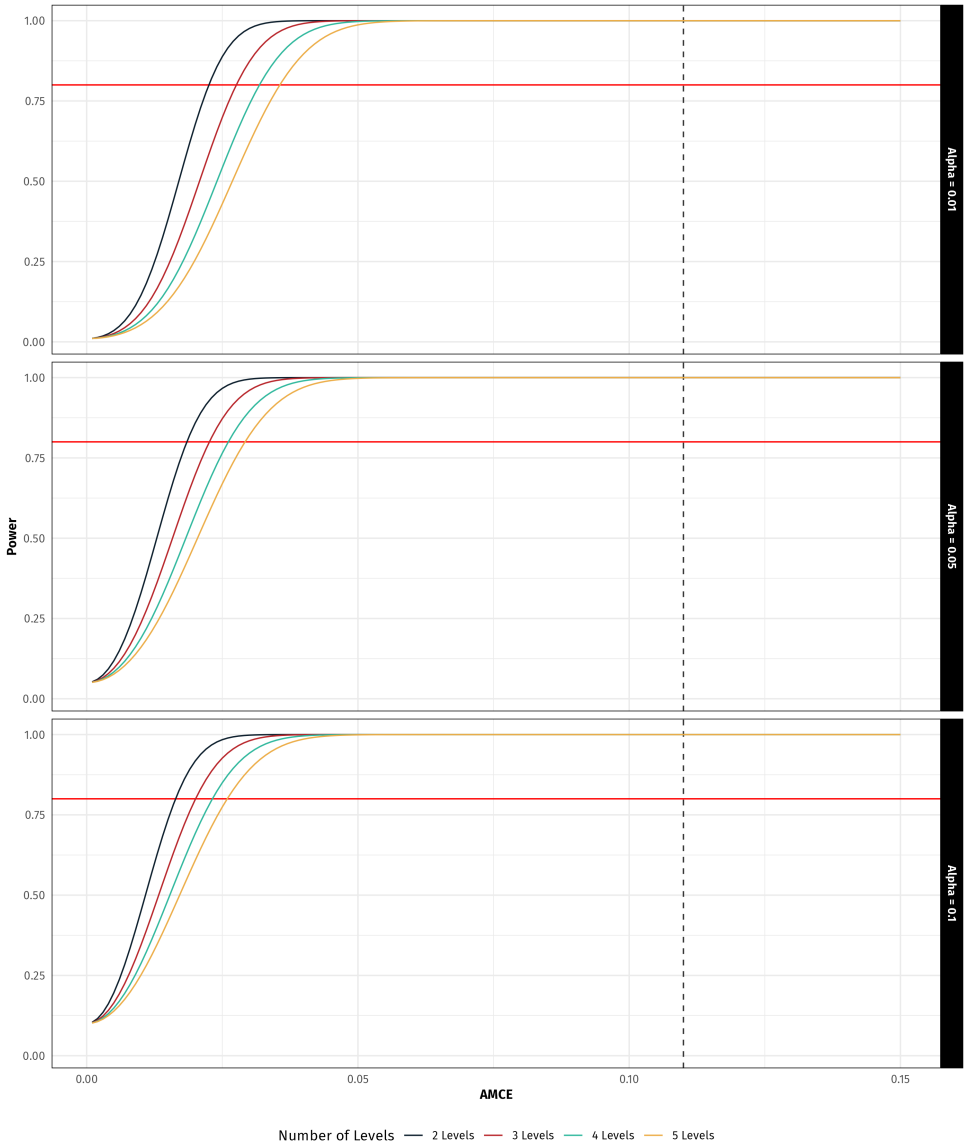


Figure A19: Power Calculations for varying significance levels, numbers of levels, and AMCEs. Red line on the y axis at the .8 level of power. The gray dashed line marks the AMCE of visualizing a poll as a bar plot on the level of trust in the poll in the voter sample at .11.

A.5. Regression Tables

This section includes a tabular presentation of the models reported as coefficient plots in the manuscript.

Baseline models, voters. Average marginal component effects of the randomized poll reporting features on perceived attractiveness, trust, and expectation alignment.

	Attractiveness Rating (PCA)	Trust in Poll Result (PCA)	Expectation Alignment with Forecast Benchmark
Table (Ref: Text)	0.471***	0.044*	0.139
	(0.018)	(0.020)	(0.268)
Lineplot	0.390***	0.062**	-0.005
	(0.020)	(0.020)	(0.270)
Dotplot	0.678***	0.063**	0.412
	(0.020)	(0.020)	(0.265)
Barchart	1.015***	0.112***	0.187
	(0.018)	(0.020)	(0.266)
Other Polls (Ref: No Comp.)	-0.130***	0.042*	0.009
	(0.018)	(0.018)	(0.237)
Previous Election	-0.184***	-0.042*	-0.073
	(0.017)	(0.018)	(0.239)
Previous Poll	0.032	0.034	0.200
	(0.017)	(0.018)	(0.236)
Uncertainty	-0.122***	0.008	-0.306
	(0.012)	(0.013)	(0.168)
Design	-0.017	-0.004	0.012
	(0.012)	(0.013)	(0.168)
Threshold	0.005	-0.010	-0.044
	(0.012)	(0.013)	(0.169)
N	22296	22625	23038
R2	0.131	0.074	0.064
Adj. R2	0.131	0.073	0.063

* p < 0.05, ** p < 0.01, *** p < 0.001

OLS estimates with standard errors clustered by respondent in parentheses.

All models include vote preference and performance fixed effects.

Outcomes: Attract. = attractiveness rating index (PCA of the accessible, helpful, and attractive items); Trust = trust in poll result index (PCA of the two trust items); Alignment = expectation alignment with the forecast benchmark (negative mean absolute error). All outcomes in original index units.

Table A6: Regression models presented in Figure A25.

Data Literacy, subgroup models (voters). Average marginal component effects of the poll reporting features estimated separately within groups defined by data literacy (tercile contrast of the share correct among the answered items of the three item knowledge battery; Low = at most one of three correct, High = all three correct; respondents in between are excluded).

	Attract.: Low	Attract.: High	Trust: Low	Trust: High	Alignment: Low	Alignment: High
Table (Ref: Text)	0.351***	0.515***	0.064	-0.002	-0.710	0.648
	(0.043)	(0.056)	(0.045)	(0.060)	(0.721)	(0.640)
Lineplot	-0.105*	0.639***	0.026	0.011	-0.185	0.821
	(0.048)	(0.062)	(0.046)	(0.058)	(0.738)	(0.669)
Dotplot	0.587***	0.714***	0.103*	0.000	0.948	0.189
	(0.046)	(0.062)	(0.046)	(0.059)	(0.715)	(0.664)
Barchart	0.828***	1.152***	0.180***	0.016	0.439	0.894
	(0.043)	(0.055)	(0.046)	(0.062)	(0.716)	(0.682)
Other Polls (Ref: No Comp.)	-0.107*	-0.171**	0.011	0.008	0.907	-0.122
	(0.042)	(0.053)	(0.042)	(0.054)	(0.650)	(0.580)
Previous Election	-0.083*	-0.269***	-0.065	-0.086	-0.072	0.154
	(0.041)	(0.053)	(0.041)	(0.054)	(0.655)	(0.587)
Previous Poll	0.042	-0.014	0.054	-0.048	0.940	1.040
	(0.040)	(0.052)	(0.041)	(0.056)	(0.648)	(0.581)
Uncertainty	-0.128***	-0.149***	-0.017	0.012	-0.323	0.073
	(0.029)	(0.038)	(0.029)	(0.039)	(0.458)	(0.416)
Design	0.045	-0.064	0.010	0.025	0.332	0.239
	(0.029)	(0.039)	(0.029)	(0.039)	(0.458)	(0.416)
Threshold	-0.016	0.033	0.015	-0.013	0.203	0.240
	(0.029)	(0.038)	(0.029)	(0.039)	(0.458)	(0.417)
N	4338	2201	4397	2188	4468	2220
R2	0.131	0.170	0.065	0.050	0.042	0.077
Adj. R2	0.127	0.163	0.061	0.042	0.037	0.069

* p < 0.05, ** p < 0.01, *** p < 0.001

OLS estimates with standard errors clustered by respondent in parentheses.

All models include vote preference and performance fixed effects.

Outcomes: Attract. = attractiveness rating index (PCA of the accessible, helpful, and attractive items); Trust = trust in poll result index (PCA of the two trust items); Alignment = expectation alignment with the forecast benchmark (negative mean absolute error). All outcomes in original index units.

Table A7: Regression models presented in Figure 3, panel a (data literacy).

Polling Literacy, subgroup models (voters). Average marginal component effects of the poll reporting features estimated separately within groups defined by polling literacy (tercile contrast of the five item knowledge battery; Low = at most one of five correct, High = at least three of five correct; respondents with two correct are excluded).

	Attract.: Low	Attract.: High	Trust: Low	Trust: High	Alignment: Low	Alignment: High
Table (Ref: Text)	0.434***	0.487***	0.052	0.034	0.777	-0.025
	(0.036)	(0.026)	(0.039)	(0.029)	(0.577)	(0.371)
Lineplot	0.187***	0.468***	0.071	0.071*	0.517	0.053
	(0.040)	(0.029)	(0.038)	(0.029)	(0.570)	(0.375)
Dotplot	0.619***	0.679***	0.075	0.044	0.482	0.200
	(0.038)	(0.029)	(0.039)	(0.030)	(0.575)	(0.368)
Barchart	0.902***	1.035***	0.185***	0.075*	1.060	-0.027
	(0.035)	(0.026)	(0.038)	(0.029)	(0.563)	(0.369)
Other Polls (Ref: No Comp.)	-0.196***	-0.118***	0.023	0.054*	-0.574	0.261
	(0.034)	(0.025)	(0.035)	(0.026)	(0.506)	(0.327)
Previous Election	-0.171***	-0.197***	-0.095**	-0.014	-0.490	-0.110
	(0.034)	(0.025)	(0.035)	(0.026)	(0.509)	(0.333)
Previous Poll	-0.012	0.044	-0.005	0.051*	-0.455	0.417
	(0.033)	(0.024)	(0.034)	(0.026)	(0.502)	(0.326)
Uncertainty	-0.118***	-0.110***	-0.006	0.015	-0.384	-0.439
	(0.024)	(0.018)	(0.024)	(0.018)	(0.361)	(0.233)
Design	-0.036	-0.003	-0.010	-0.020	0.095	0.027
	(0.024)	(0.018)	(0.025)	(0.018)	(0.362)	(0.233)
Threshold	0.066**	-0.013	0.009	0.006	0.100	0.047
	(0.024)	(0.018)	(0.024)	(0.018)	(0.359)	(0.234)
N	5940	10747	6083	10904	6183	11105
R2	0.120	0.135	0.057	0.091	0.049	0.078
Adj. R2	0.117	0.134	0.054	0.089	0.046	0.077

* p < 0.05, ** p < 0.01, *** p < 0.001

OLS estimates with standard errors clustered by respondent in parentheses.

All models include vote preference and performance fixed effects.

Outcomes: Attract. = attractiveness rating index (PCA of the accessible, helpful, and attractive items); Trust = trust in poll result index (PCA of the two trust items); Alignment = expectation alignment with the forecast benchmark (negative mean absolute error). All outcomes in original index units.

Table A8: Regression models presented in Figure 3, panel b (polling literacy).

Polling Belief Accuracy, subgroup models (voters). Average marginal component effects of the poll reporting features estimated separately within groups defined by polling belief accuracy (tercile contrast of the mean absolute deviation between prior beliefs and the displayed poll; High = bottom tercile of the deviation, i.e., the most accurate beliefs; Low = top tercile; middle tercile excluded).

	Attract.: Low	Attract.: High	Trust: Low	Trust: High	Alignment: Low	Alignment: High
Table (Ref: Text)	0.370***	0.564***	0.083*	-0.021	-0.066	0.180
	(0.032)	(0.032)	(0.034)	(0.031)	(0.548)	(0.276)
Lineplot	0.107**	0.628***	0.092**	0.001	0.079	-0.171
	(0.035)	(0.035)	(0.034)	(0.032)	(0.554)	(0.293)
Dotplot	0.643***	0.700***	0.072*	0.015	0.562	0.108
	(0.034)	(0.035)	(0.035)	(0.033)	(0.549)	(0.281)
Barchart	0.860***	1.124***	0.145***	0.047	0.510	-0.125
	(0.032)	(0.031)	(0.035)	(0.032)	(0.549)	(0.276)
Other Polls (Ref: No Comp.)	-0.126***	-0.156***	0.015	0.024	-0.250	0.030
	(0.031)	(0.030)	(0.031)	(0.029)	(0.489)	(0.250)
Previous Election	-0.134***	-0.226***	-0.045	-0.015	-0.099	0.156
	(0.030)	(0.031)	(0.031)	(0.029)	(0.492)	(0.252)
Previous Poll	0.025	0.032	0.061	0.020	0.096	0.374
	(0.030)	(0.029)	(0.031)	(0.028)	(0.491)	(0.248)
Uncertainty	-0.123***	-0.128***	-0.018	0.029	-0.512	-0.092
	(0.022)	(0.022)	(0.022)	(0.020)	(0.347)	(0.176)
Design	0.001	-0.035	-0.025	0.006	-0.375	0.070
	(0.022)	(0.022)	(0.022)	(0.020)	(0.347)	(0.175)
Threshold	0.001	0.011	0.000	0.009	0.048	0.247
	(0.022)	(0.022)	(0.022)	(0.020)	(0.349)	(0.176)
N	7612	7054	7760	7134	7954	7220
R2	0.121	0.154	0.061	0.041	0.023	0.084
Adj. R2	0.119	0.152	0.059	0.038	0.021	0.082

* p < 0.05, ** p < 0.01, *** p < 0.001

OLS estimates with standard errors clustered by respondent in parentheses.

All models include vote preference and performance fixed effects.

Outcomes: Attract. = attractiveness rating index (PCA of the accessible, helpful, and attractive items); Trust = trust in poll result index (PCA of the two trust items); Alignment = expectation alignment with the forecast benchmark (negative mean absolute error). All outcomes in original index units.

Table A9: Regression models presented in Figure 3, panel c (prior-belief accuracy).

A.6. Heterogeneous effects using continuous moderators

In the main text, we illustrate heterogeneity by contrasting respondents in the bottom and top third of each moderator. This section reports the formal tests, which use the full variation in each moderator. For each outcome, we interact the visualization format with the continuous moderator, retaining the fixed effects of the main models:

$$y_i = \beta\,\text{Format}_i + \gamma\left(\text{Format}_i \times M_i\right) + \delta\,M_i + \alpha_{g(i)} + \varepsilon_i, \tag{3}$$

where Format_i contrasts each format against the text baseline, M_i is the continuous moderator, and $\alpha_{g(i)}$ are voting-intention and performance fixed effects. For belief inaccuracy, the model also includes the square of the mean absolute deviation. Moderation is tested with a joint Wald test of the interaction terms for each format.

To make the moderation interpretable, Figures A20 and A22 plot the marginal effect of each format (relative to text) across the full range of the moderator, with shaded 90% and 95% confidence bands. A flat line indicates no moderation; a sloped line indicates that the format's effect strengthens or weakens as the moderator increases. To show where the data actually lie, and to connect the continuous view back to the group contrasts in the main text, we overlay the empirical distribution of the moderator as a histogram in the background of each panel, colored by the low and high groups used in the manuscript, with the excluded middle group in grey.

The continuous estimates show the same direction and ordering of the moderated format effects as the group contrasts in the main text, so the conclusions do not depend on splitting the moderators into groups (high versus low).

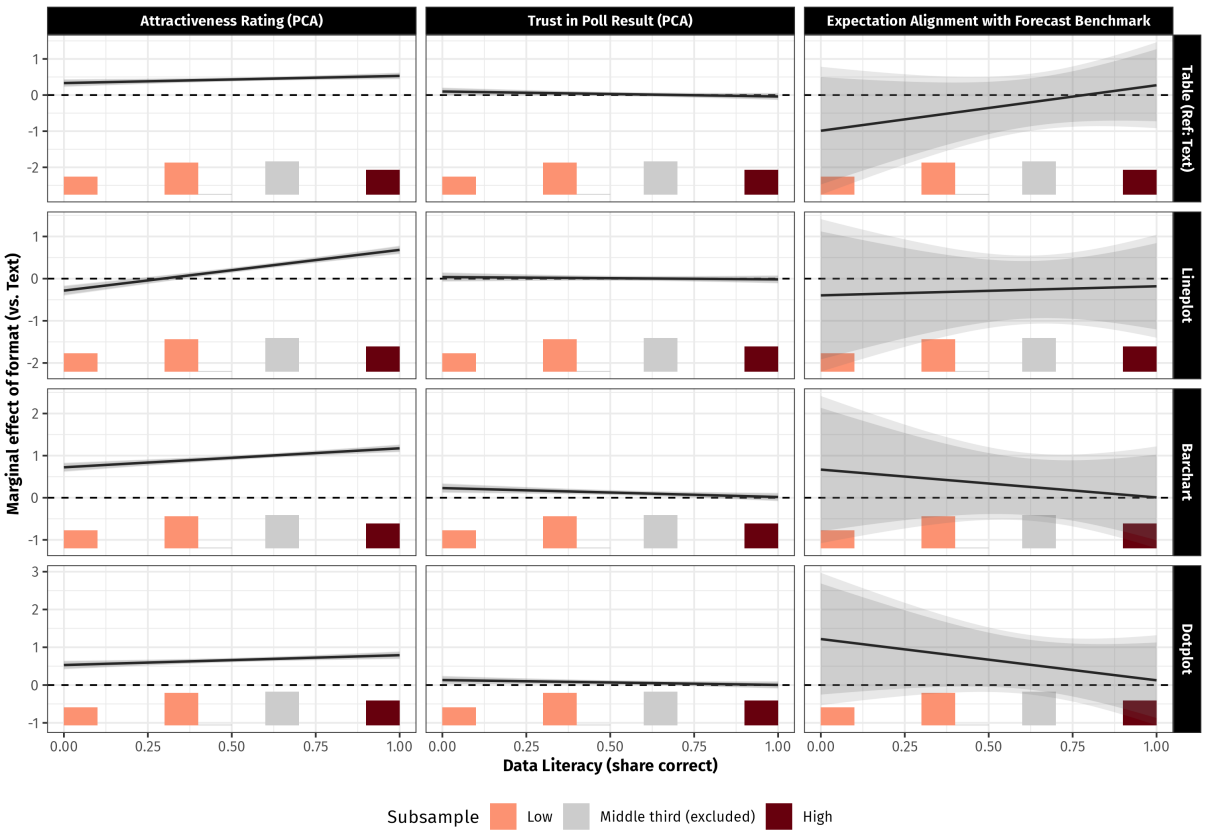


Figure A20: Marginal effect of each visualization format (vs. text) across the distribution of *data literacy* (share of answered items answered correctly). Lines are point estimates; shaded bands are 90% and 95% confidence intervals. The background histograms show the low and high groups, with the middle group excluded from the group contrasts in grey.

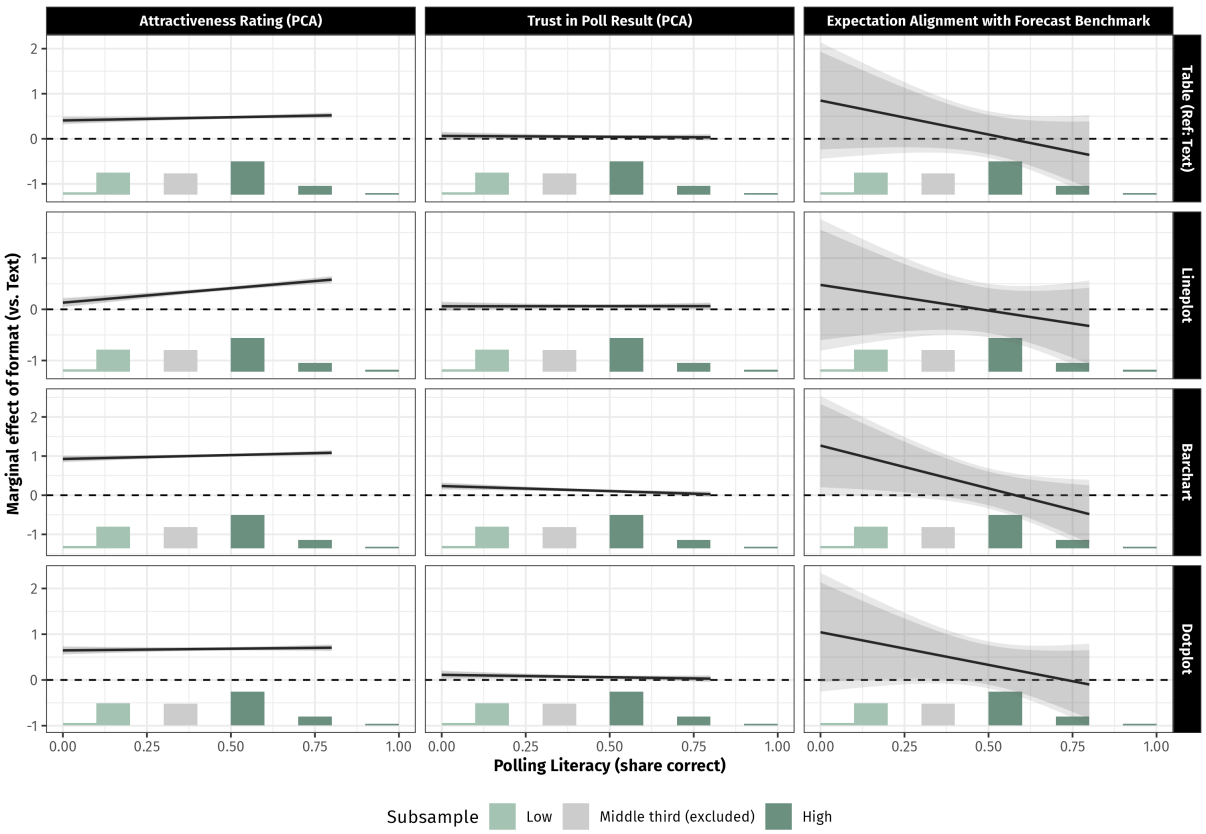


Figure A21: Marginal effect of each visualization format (vs. text) across the distribution of *polling literacy* (share of five statements classified correctly). Lines are point estimates; shaded bands are 90% and 95% confidence intervals. The background histograms show the low and high groups, with the middle group excluded from the group contrasts in grey.

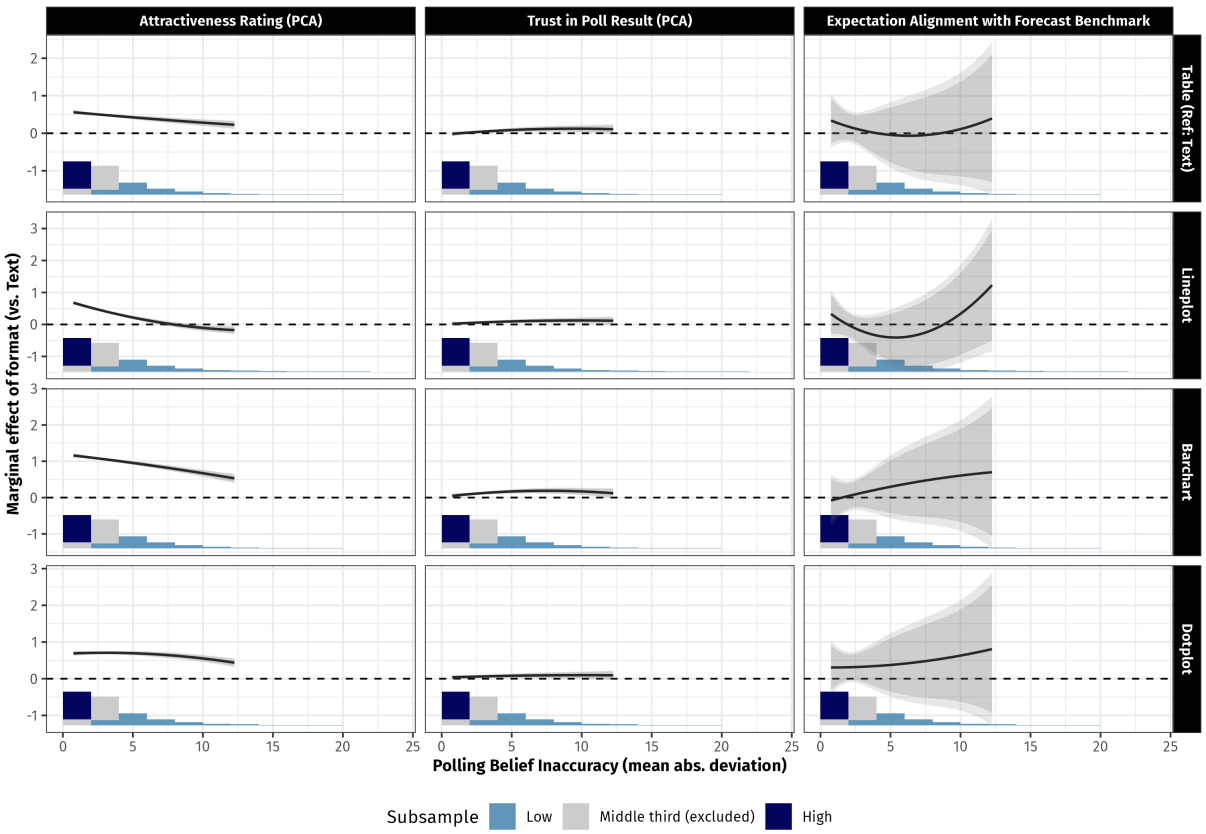


Figure A22: Marginal effect of each visualization format (vs. text) across the distribution of *belief accuracy* (absolute deviation between a respondent's prior expectation and the displayed poll). Bands are 90% and 95% confidence intervals; the background histograms show the low and high groups, with the excluded middle third in grey.

A.7. Anchoring effect sizes in raw outcome units

Because the main analysis reports trust and attractiveness as standardized PCA scores, whose magnitudes are hard to interpret, we re-estimate the baseline AMCEs on the original outcome scales. The specification, fixed effects, and samples are identical to the main models; only the dependent variable changes—trust on the raw 0–100 slider (both items and their mean) and attractiveness on the original 1–5 rating items (each of the four fielded items, and the mean of the three items that form the attractiveness index). AMCEs can thus be read directly as changes in slider or rating points. Relative to text-only reporting, a bar chart raises mean trust by about 2.4 points on the 100-point scale (95% CI [1.5, 3.2]), with line and dot plots gaining about 1.2 points. Reporting the four attractiveness items separately shows the composite is not driven by any single dimension. Overall, the raw-unit estimates reproduce the ordering and significance of the standardized results. Full results are in Tables A10–A12.

Table A10: Trust AMCEs in raw slider points (0–100), voters. Standard errors in parentheses.

Trust AMCEs in raw slider points (0-100), Voters

	Trust: Current (0-100)	Trust: Election (0-100)	Trust: Mean (0-100)
Table (Ref: Text)	0.793	1.109*	0.983*
	(0.435)	(0.451)	(0.415)
Lineplot	0.918*	1.614***	1.241**
	(0.443)	(0.454)	(0.420)
Dotplot	1.383**	1.164*	1.243**
	(0.444)	(0.460)	(0.424)
Barchart	2.699***	1.996***	2.367***
	(0.439)	(0.461)	(0.422)
Other Polls (Ref: No Comp.)	0.948*	0.860*	0.942*
	(0.393)	(0.407)	(0.374)
Previous Election	-1.240**	-0.617	-0.978**
	(0.397)	(0.408)	(0.377)
Previous Poll	0.582	0.779	0.655
	(0.393)	(0.405)	(0.374)
Uncertainty	0.271	0.078	0.193
	(0.278)	(0.288)	(0.265)
Design	-0.019	-0.157	-0.096
	(0.278)	(0.288)	(0.265)
Threshold	-0.110	-0.229	-0.131
	(0.278)	(0.288)	(0.265)
N	22802	22810	22987
R2	0.072	0.061	0.076
Adj. R2	0.071	0.060	0.075

* p < 0.05, ** p < 0.01, *** p < 0.001
Standard errors clustered by respondent in parentheses.
Fixed effects: vote preference and performance.

Table A11: Attractiveness AMCEs in raw item points (1–5), voters. Standard errors in parentheses.

Attractiveness AMCEs in raw item points (1-5), Voters					
	Accessible (1-5)	Informative (1-5, amount of information)	Helpful (1-5)	Attractive (1-5)	Attractiveness: Mean (1-5)
Table (Ref: Text)	0.370***	0.145***	0.253***	0.781***	0.471***
	(0.025)	(0.019)	(0.023)	(0.026)	(0.018)
Lineplot	0.111***	0.006	0.224***	0.831***	0.382***
	(0.027)	(0.020)	(0.024)	(0.028)	(0.020)
Dotplot	0.458***	0.148***	0.420***	1.134***	0.669***
	(0.026)	(0.018)	(0.023)	(0.027)	(0.020)
Barchart	0.753***	0.104***	0.688***	1.567***	1.004***
	(0.024)	(0.018)	(0.022)	(0.026)	(0.018)
Other Polls (Ref: No Comp.)	-0.282***	-0.230***	-0.057**	-0.019	-0.118***
	(0.022)	(0.017)	(0.021)	(0.024)	(0.017)
Previous Election	-0.323***	-0.182***	-0.075***	-0.130***	-0.177***
	(0.022)	(0.016)	(0.021)	(0.024)	(0.017)
Previous Poll	-0.072***	-0.072***	0.062**	0.111***	0.031
	(0.022)	(0.016)	(0.020)	(0.024)	(0.017)
Uncertainty	-0.191***	-0.114***	-0.066***	-0.090***	-0.116***
	(0.016)	(0.012)	(0.015)	(0.017)	(0.012)
Design	-0.016	-0.029*	-0.031*	-0.006	-0.021
	(0.016)	(0.012)	(0.015)	(0.017)	(0.012)
Threshold	-0.028	-0.038**	0.000	0.044**	0.005
	(0.016)	(0.012)	(0.015)	(0.017)	(0.012)
N	22672	22402	22468	22504	22895
R2	0.068	0.023	0.055	0.152	0.129
Adj. R2	0.067	0.023	0.055	0.152	0.129

* p < 0.05, ** p < 0.01, *** p < 0.001

Standard errors clustered by respondent in parentheses.

Fixed effects: vote preference and performance.

The attractiveness composite excludes the informative item (different scale endpoints).

Table A12: Composite AMCEs in raw units (mean trust 0–100, mean attractiveness 1–5), by sample. Standard errors in parentheses.

Composite AMCEs in raw units, by sample

	Voters		Journalists		Politicians	
	Trust (0-100)	Attract. (1-5)	Trust (0-100)	Attract. (1-5)	Trust (0-100)	Attract. (1-5)
Table (Ref: Text)	0.983*	0.471***	-10.823	-0.291	-6.952	0.519**
	(0.415)	(0.018)	(6.335)	(0.280)	(4.622)	(0.184)
Lineplot	1.241**	0.382***	-10.297	0.622*	0.532	0.845***
	(0.420)	(0.020)	(5.261)	(0.285)	(4.324)	(0.231)
Dotplot	1.243**	0.669***	-13.806*	0.444	-9.080	0.285
	(0.424)	(0.020)	(6.075)	(0.290)	(5.132)	(0.252)
Barchart	2.367***	1.004***	-13.884*	0.707*	1.451	1.050***
	(0.422)	(0.018)	(5.896)	(0.296)	(4.229)	(0.166)
Other Polls (Ref: No Comp.)	0.942*	-0.118***	0.390	0.038	3.918	-0.273
	(0.374)	(0.017)	(6.508)	(0.313)	(4.590)	(0.218)
Previous Election	-0.978**	-0.177***	-1.458	-0.345	4.160	-0.303
	(0.377)	(0.017)	(5.810)	(0.262)	(4.154)	(0.191)
Previous Poll	0.655	0.031	4.748	-0.161	-3.050	0.253
	(0.374)	(0.017)	(6.067)	(0.255)	(3.641)	(0.170)
Uncertainty	0.193	-0.116***	1.837	-0.334	4.734	-0.001
	(0.265)	(0.012)	(4.325)	(0.190)	(3.115)	(0.133)
Design	-0.096	-0.021	0.394	0.115	-2.333	0.065
	(0.265)	(0.012)	(4.300)	(0.199)	(3.185)	(0.140)
Threshold	-0.131	0.005	3.550	0.384	1.767	0.285*
	(0.265)	(0.012)	(4.329)	(0.197)	(3.026)	(0.134)
N	22987	22895	108	108	143	143
R2	0.076	0.129	0.165	0.215	0.293	0.296
Adj. R2	0.075	0.129	0.059	0.116	0.228	0.231

* p < 0.05, ** p < 0.01, *** p < 0.001

Standard errors clustered by respondent in parentheses.

Fixed effects as in baseline models.

The attractiveness composite excludes the informative item (different scale endpoints).

A.8. False discovery rate correction

Because we estimate many effects at once, we check whether the results hold under a correction for multiple testing. This correction was not part of the pre-analysis plan, which specifies conventional significance levels; we add it as a conservative robustness check. We apply the Benjamini–Hochberg procedure to the AMCEs of each sample, treating all 30 effects shown in a sample's coefficient plot (10 feature levels $\times$ 3 outcomes) as one family. The same families determine the solid and hollow dots in the coefficient plots: an effect is shown as a solid dot if its adjusted p-value is below 0.05. In the heterogeneity figures, the family is all effects within a panel.

Tables A13–A15 report the estimates with unadjusted and adjusted p-values; estimates are in the original units of each outcome. Among voters, 10 of the 13 effects that are significant at $p < 0.05$ remain significant after the correction. These include all format effects on attractiveness, the effects of uncertainty information and of the other-polls and previous-election references on attractiveness, and the trust gains from bar charts, dot plots, and line graphs. The trust effects of tables, of the polling average, and of the previous election result do not survive the correction (adjusted p between 0.051 and 0.063). In the elite samples, none of the six nominally significant effects among journalists survives the correction; among candidates, the attractiveness gains from bar charts and line graphs do.

Table A13: AMCEs with Benjamini–Hochberg false discovery rate correction, voters.

Main AMCEs with BH-FDR control (Voters)									
	Attractiveness Rating (PCA)			Trust in Poll Result (PCA)			Expectation Alignment with Forecast Benchmark		
	Est.	p	p (BH)	Est.	p	p (BH)	Est.	p	p (BH)
Table (Ref: Text)	0.471	0.000	**0.000**	0.044	0.027	0.063	0.139	0.604	0.787
Lineplot	0.390	0.000	**0.000**	0.062	0.002	**0.006**	−0.005	0.986	0.986
Dotplot	0.678	0.000	**0.000**	0.063	0.002	**0.006**	0.412	0.120	0.211
Barchart	1.015	0.000	**0.000**	0.112	0.000	**0.000**	0.187	0.482	0.689
Other Polls (Ref: No Comp.)	−0.130	0.000	**0.000**	0.042	0.021	0.051	0.009	0.971	0.986
Previous Election	−0.184	0.000	**0.000**	−0.042	0.020	0.051	−0.073	0.759	0.881
Previous Poll	0.032	0.058	0.116	0.034	0.058	0.116	0.200	0.397	0.627
Uncertainty	−0.122	0.000	**0.000**	0.008	0.554	0.756	−0.306	0.069	0.130
Design	−0.017	0.178	0.296	−0.004	0.767	0.881	0.012	0.943	0.986
Threshold	0.005	0.709	0.881	−0.010	0.426	0.639	−0.044	0.793	0.881

Benjamini-Hochberg FDR correction across all 30 effects of the sample (10 features x 3 outcomes), the same family as in the coefficient plots. Adjusted p < .05 in bold.

Table A14: AMCEs with Benjamini–Hochberg false discovery rate correction, journalists.

Main AMCEs with BH-FDR control (Journalists)									
	Attractiveness Rating (PCA)			Trust in Poll Result (PCA)			Expectation Alignment with Forecast Benchmark		
	Est.	p	p (BH)	Est.	p	p (BH)	Est.	p	p (BH)
Table (Ref: Text)	−0.324	0.272	0.679	−0.523	0.090	0.319	0.935	0.651	0.855
Lineplot	0.614	0.039	0.234	−0.487	0.060	0.258	−0.989	0.496	0.810
Dotplot	0.441	0.144	0.433	−0.661	0.026	0.197	−1.196	0.472	0.810
Barchart	0.737	0.020	0.197	−0.668	0.020	0.197	−1.448	0.608	0.855
Other Polls (Ref: No Comp.)	0.009	0.977	0.977	0.011	0.972	0.977	0.950	0.698	0.872
Previous Election	−0.389	0.161	0.439	−0.076	0.787	0.924	−0.314	0.894	0.977
Previous Poll	−0.192	0.474	0.810	0.219	0.461	0.810	0.577	0.801	0.924
Uncertainty	−0.338	0.096	0.319	0.095	0.655	0.855	−2.843	0.026	0.197
Design	0.109	0.599	0.855	0.013	0.950	0.977	−0.944	0.498	0.810
Threshold	0.410	0.050	0.249	0.164	0.438	0.810	0.882	0.513	0.810

Benjamini-Hochberg FDR correction across all 30 effects of the sample (10 features x 3 outcomes), the same family as in the coefficient plots. Adjusted p < .05 in bold.

Table A15: AMCEs with Benjamini–Hochberg false discovery rate correction, political candidates.

Main AMCEs with BH-FDR control (Politicians)									
	Attractiveness Rating (PCA)			Trust in Poll Result (PCA)			Expectation Alignment with Forecast Benchmark		
	Est.	p	p (BH)	Est.	p	p (BH)	Est.	p	p (BH)
Table (Ref: Text)	0.511	0.008	0.080	−0.334	0.136	0.399	2.068	0.256	0.529
Lineplot	0.854	0.000	**0.007**	0.025	0.904	0.904	−0.621	0.760	0.844
Dotplot	0.347	0.176	0.441	−0.438	0.078	0.366	1.700	0.433	0.619
Barchart	1.060	0.000	**0.000**	0.068	0.737	0.844	−0.679	0.725	0.844
Other Polls (Ref: No Comp.)	−0.227	0.304	0.532	0.188	0.396	0.603	−0.353	0.844	0.891
Previous Election	−0.317	0.111	0.399	0.200	0.319	0.532	−2.306	0.194	0.448
Previous Poll	0.254	0.146	0.399	−0.147	0.402	0.603	−3.022	0.085	0.366
Uncertainty	−0.024	0.862	0.891	0.228	0.130	0.399	2.376	0.081	0.366
Design	0.073	0.611	0.763	−0.112	0.466	0.636	−1.388	0.288	0.532
Threshold	0.265	0.057	0.366	0.084	0.566	0.739	−1.521	0.265	0.529

Benjamini-Hochberg FDR correction across all 30 effects of the sample (10 features x 3 outcomes), the same family as in the coefficient plots. Adjusted p < .05 in bold.

A.9. Marginal means

In addition to the AMCEs, we report marginal means for all levels of each feature (8). A marginal mean is the average outcome among all vignettes that show a given feature level, averaged over the other features. Marginal means do not depend on the choice of a reference category, and they show the level of each outcome and not only the differences between levels. Standard errors are clustered by respondent. Since the features were randomized independently, differences between marginal means are almost identical to the AMCEs in the main text. For example, bar charts score 1.02 points higher than text on the attractiveness index, the same as the AMCE. Figure A23 shows the outcomes in their original units: the attractiveness and trust indices are PCA scores, and expectation alignment is the negative mean absolute deviation from the forecast benchmark in percentage points. The dashed line marks the sample mean.

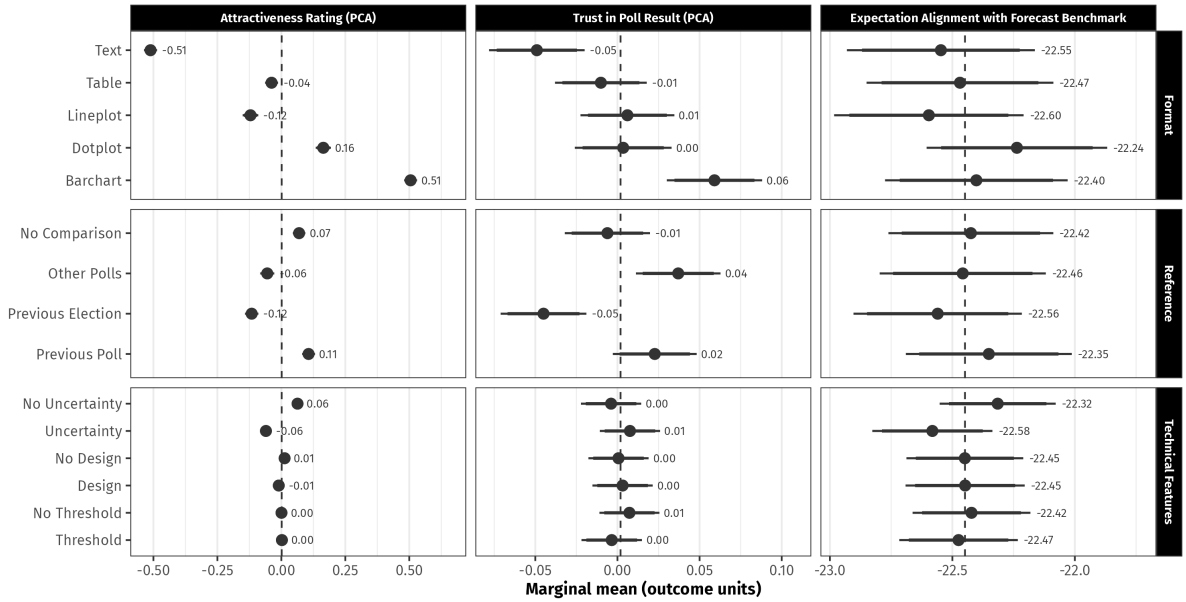


Figure A23: Marginal means by feature level, voters. Estimates with 95% and 90% confidence intervals. The dashed line marks the sample mean of each outcome.

A.10. Replication Materials

All data and code needed to reproduce the analyses, figures, and tables in the manuscript and in these Supplementary Materials will be made available in a replication archive. The archive also contains the code that generated the 160 poll vignettes.

How to reproduce the results Install R 4.4 or later, open the project file `visualisation.Rproj`, and run `_code/99_master_run.R`. This script runs all preparation, description, and analysis steps in order and writes every figure and table to `_figures/`. The script `_code/packages.R` loads all required R packages and installs missing ones. Tables saved as images additionally require the R package `webshot2` and Google Chrome.

Folder structure `_code/` contains all scripts, numbered in the order in which they run; `_data/in/` the input data; `_data/out/` the analysis datasets and estimates produced by the scripts; and `_figures/` all figures and tables.

Scripts

- **01 Vignettes**: run separately, about 30 minutes each
  - `01a_create_vignettes.R`: creates the German vignettes shown to respondents
  - `01b_create_vignettes_ENG.R`: creates the English versions of the vignettes
- **02 Data preparation**
  - `02a_prepare_vignettes.R`: builds the outcomes: attractiveness and trust indices and expectation alignment
  - `02b_prepare_surveydata.R`: cleans the respondent data for voters, journalists, and candidates
  - `02c_prepare_moderators.R`: builds the moderators: data literacy, polling literacy, and belief accuracy
  - `02d_prepare_analysis.R`: merges everything into the analysis datasets
- **03 Descriptives**
  - `03a_describe_benchmark_samples.R`: sample composition and benchmarks
  - `03b_describe_collection_progress.R`: data collection over time
  - `03c_describe_randomization.R`: randomization and balance checks
  - `03d_describe_rating_features.R`: outcome items and PCA results
  - `03e_describe_media_content.R`: media content analysis
  - `03f_describe_pollstability_fieldtime.R`: published polls during the field period
- **04 Analysis**
  - `04a_analyze_main_models.R`: main effects, subgroup models, and FDR correction
  - `04b_analyze_main_plots.R`: coefficient plots
  - `04c_analyze_heterogeneity_continuous.R`: moderation tests with continuous moderators
  - `04d_analyze_regtables.R`: regression tables
  - `04e_analyze_fdr_tables.R`: false discovery rate tables

  - `04f_analyze_amce_rawunits.R`: main effects in raw outcome units
  - `04g_analyze_ranking_all160.R`: ranking of all 160 vignettes
  - `04h_analyze_power.R`: power calculations
  - `04i_analyze_heterogeneity_fieldtime.R`: effects by field week
  - `04j_analyze_marginal_means.R`: marginal means

- **Helpers**
  - `packages.R`: loads all R packages
  - `_functions.R`: shared plotting functions and helpers
  - `99_master_run.R`: runs all 02–04 scripts in order

B. Media Content Analysis

While the findings from our experimental study provide valuable insights into how different design choices in election poll reporting can influence respondents' perceptions, they are limited in their ability to speak to the real-world consequences of these design choices. In particular, the experimental results alone do not reveal how common practices in media reporting might shape public interpretation of poll information.

To bridge this gap, we complemented our experimental study with a systematic media content analysis of newspapers. This analysis allowed us to examine how polling data are typically presented in practice, including the formats, reference points, and contextual information most frequently used by media outlets. By combining experimental evidence with empirical observations from actual media reporting, we can draw more informed inferences about the potential implications of current poll reporting practices for public understanding and political behavior.

This section provides a detailed account of our methodological approach for data retrieval and analysis. We first outline the selection criteria used to identify relevant media reports, followed by a description of the coding procedures and instructions applied to extract key information from the articles. After establishing these methodological foundations, we present the findings from the media content analysis.

B.1. Methodological Approach & Selection Criteria

To define the population of relevant newspaper articles, we relied on the *LexisNexis* database. We retrieved all German articles matching the query "Bundestagswahl" (federal election) and "Umfrag*" (poll), and which were published between January 1 and February 24, 2025. This way, we derived an initial population of 2,219 articles matching these criteria.

To systematically evaluate patterns in news media reporting of election polls within this subset of articles, we trained four human coders to collect an original dataset of election poll reporting. Each of them annotated 555 articles and gathered a wide range of information on how election polls were presented, contextualized and discussed.

B.2. Coding Instructions

When annotating the newspaper articles, the coders followed the coding instructions presented below.

Variable	Description
ID	-
Note	Room for additional comments/notes. Use, e.g., if you're uncertain or want to flag this article as remarkable.
Dup	Duplicate
Neg	Select if the article is irrelevant, i.e., does not report a poll
Main	Poll main subject of the article?
Poll	Is this about an election poll or several polls? Note that we're only interested in election polls on party support, not polls on political issues, candidate popularity, etc.
WK	Is a particular election district mentioned AND related to poll?
Provider	List provider(s) of poll. If multiple, separate by ",". Note that "Auftraggeber" is not a provider.
Wk_names	List name(s) of districts (Wahlkreise) if mentioned; if multiple, separate by ","
Reference_to	Are the reported poll results put into context by comparing them to previous poll, polling average, or the past election result?
Uncert_rep	Is statistical/predictive uncertainty of the polling result addressed in some way (e.g., by providing ranges in addition to or instead of point estimates, or by referring to an "error range" of X%)? ONLY select this if the uncertainty is discussed directly in the reporting of the results, e.g., "13% (11-15%)". If the article provides just a more general statement/disclaimer on the uncertainty of polls, please use the indicator "uncert_gen".

Uncert_gen Does the article provide a general statement/disclaimer on the uncertainty of polls? E.g.: "Wahlumfragen sind generell mit Unsicherheiten behaftet. Unter anderem erschweren nachlassende Parteibindungen und immer kurzfristigere Wahlentscheidungen den Meinungsforschungsinstituten die Gewichtung der erhobenen Daten. Grundsätzlich spiegeln Umfragen nur das Meinungsbild zum Zeitpunkt der Befragung wider und sind keine Prognosen auf den Wahlausgang."

N Is the number of respondents reported?

Timing Is the field time of the survey reported? (Note: we're not interested in the publishing date!) Only tick this box if precise information is given (not just "current poll").

Mode Is the survey mode of the poll reported? (E.g., online, face to face, telephone, mixed, etc.).

System Are the polling results discussed in relation to features of the electoral system? (E.g., 5% threshold, Grundmandatsklausel, Erststimme vs. Zweitstimme)

Causal Does the article provide a causal explanation or at least points to an association of a (non-)change in the polls with, e.g., an external event? E.g.: "Party X gained support after the deadly attack in Aschaffenburg". "Party X gained no support despite the much-discussed vote in parliament"

Quality Flag article if you think that this is an example of high-quality reporting on poll

B.3. Results

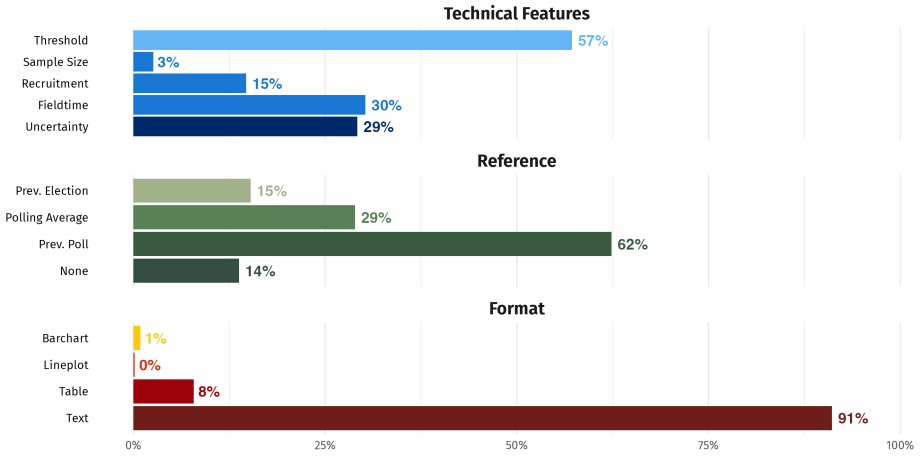


Figure A24: Results from manual coded newspaper articles on election polls (online/offline) during the 2025 German federal election campaign.

Three patterns stand out: First, and most prominently, written text is by far the most common mode of reporting election polls. More than nine out of ten articles in our sample referencing an election poll did so exclusively reporting its outcome in written text (see Figure A24). Second, newspapers differ considerably in the technical information they include when reporting poll results. The presented polling outcome was discussed taking the electoral threshold, a critical element in the competition between smaller parties, into consideration in more than every second article. Relevant methodological features, namely the field time (30%), the recruitment mode (15%), and the sample size (3%) were reported far less frequently. Most importantly, however, since this has been raised as a main concern by scholars in the past, almost every third article referred to the uncertainties associated with the respective poll. Finally, the large majority of articles (86%) provide contextual benchmarks for the reported vote shares. A party's current support is most often compared to its previous result from the same polling firm (62%), followed by its average across pollsters (29%) and its previous election result (15%); articles can use several benchmarks.

C. Further Analyses

C.1. Comparing voters to political candidates and journalists

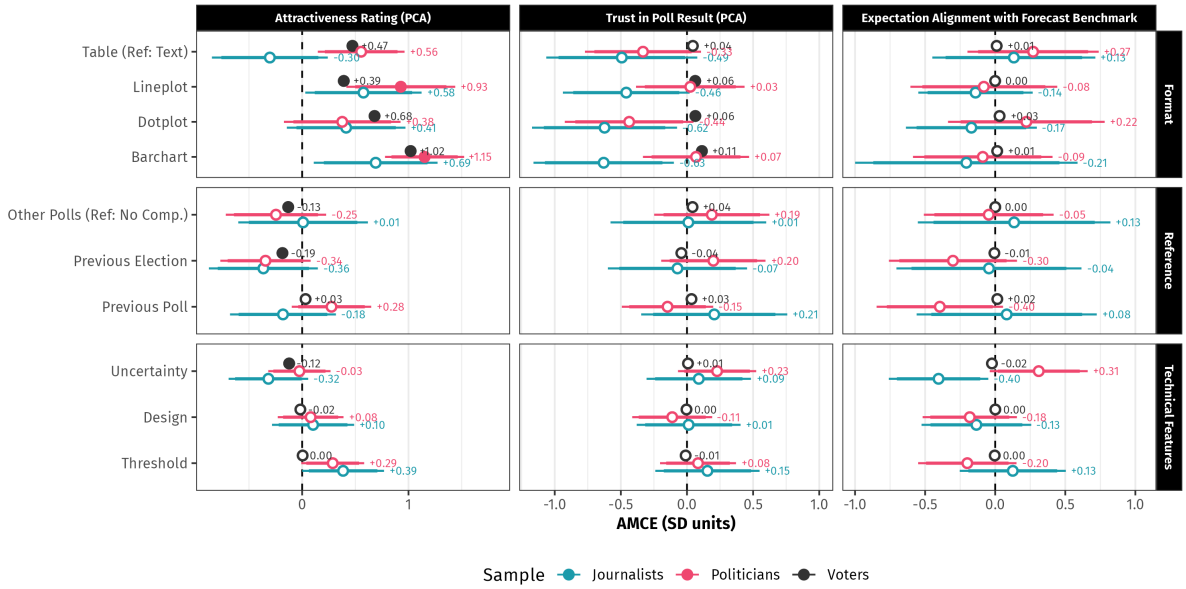


Figure A25: **Baseline effects for all varied features among voters, journalists, and political candidates.** 95% (narrow) and 90% (wide) confidence intervals reported.

Baseline models, elites. Average marginal component effects of the randomized poll reporting features among journalists (J) and political candidates (P), separately for each outcome.

	Attract. (J)	Attract. (P)	Trust (J)	Trust (P)	Alignment (J)	Alignment (P)
Table (Ref: Text)	-0.324	0.511**	-0.523	-0.334	0.935	2.068
	(0.293)	(0.190)	(0.305)	(0.223)	(2.064)	(1.812)
Lineplot	0.614*	0.854***	-0.487	0.025	-0.989	-0.621
	(0.294)	(0.238)	(0.256)	(0.208)	(1.449)	(2.027)
Dotplot	0.441	0.347	-0.661*	-0.438	-1.196	1.700
	(0.300)	(0.255)	(0.293)	(0.247)	(1.658)	(2.162)
Barchart	0.737*	1.060***	-0.668*	0.068	-1.448	-0.679
	(0.313)	(0.173)	(0.284)	(0.203)	(2.816)	(1.926)
Other Polls (Ref: No Comp.)	0.009	-0.227	0.011	0.188	0.950	-0.353
	(0.328)	(0.220)	(0.315)	(0.221)	(2.439)	(1.791)
Previous Election	-0.389	-0.317	-0.076	0.200	-0.314	-2.306
	(0.275)	(0.198)	(0.281)	(0.200)	(2.343)	(1.768)
Previous Poll	-0.192	0.254	0.219	-0.147	0.577	-3.022
	(0.268)	(0.174)	(0.295)	(0.175)	(2.281)	(1.745)
Uncertainty	-0.338	-0.024	0.095	0.228	-2.843*	2.376
	(0.201)	(0.136)	(0.211)	(0.150)	(1.260)	(1.352)
Design	0.109	0.073	0.013	-0.112	-0.944	-1.388
	(0.207)	(0.144)	(0.210)	(0.153)	(1.389)	(1.301)
Threshold	0.410*	0.265	0.164	0.084	0.882	-1.521
	(0.207)	(0.138)	(0.211)	(0.146)	(1.343)	(1.358)
N	107	142	107	143	108	143
R2	0.215	0.279	0.162	0.293	0.074	0.096
Adj. R2	0.115	0.212	0.055	0.228	-0.043	0.013

* p < 0.05, ** p < 0.01, *** p < 0.001

OLS estimates with standard errors clustered by respondent in parentheses.

All models include performance fixed effects.

Outcomes: Attract. = attractiveness rating index (PCA of the accessible, helpful, and attractive items); Trust = trust in poll result index (PCA of the two trust items); Alignment = expectation alignment with the forecast benchmark (negative mean absolute error). All outcomes in original index units.

Figure A26: Regression models presented in Figure 2.

C.2. Ranking all 160 poll presentations

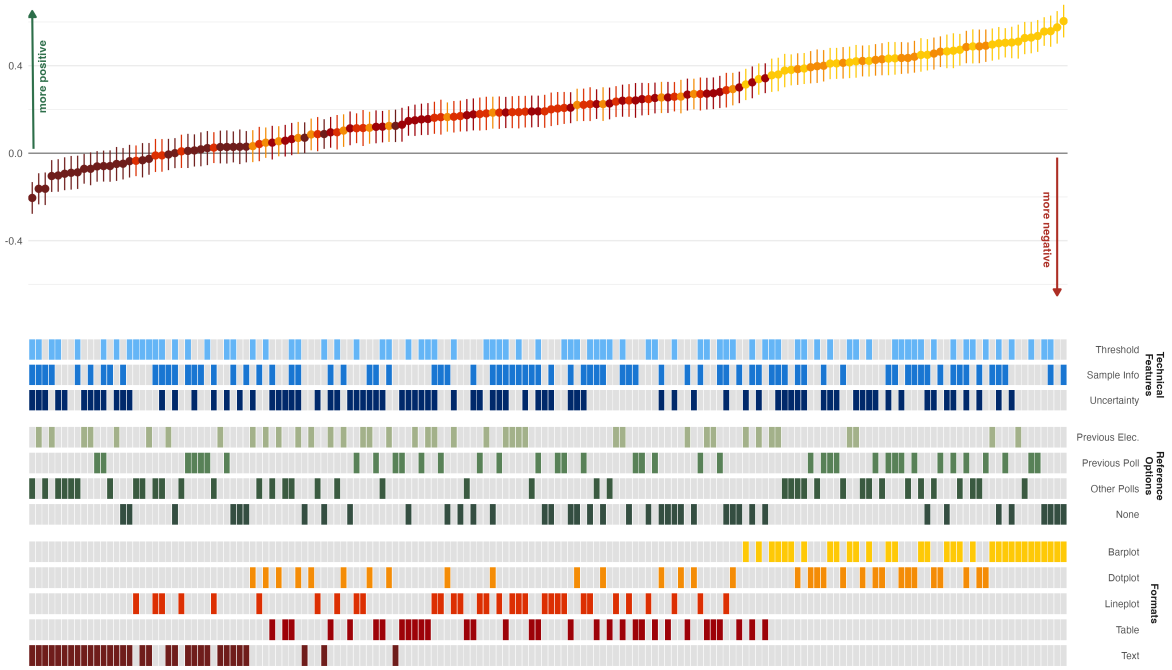


Figure A27: Coefficients for the rating of the 160 poll depictions. The outcome is the attractiveness index, the first principal component of three rating items on how visually attractive, easy to understand, and helpful respondents find the depiction.

Which depictions are perceived as most and least attractive? The most striking pattern is that visual displays of polling results are consistently rated as more attractive than text-based presentations (see Figure A27). Further, formats that are regarded most appealing generally contain fewer supplementary technical elements, while those perceived as least attractive tend to include a greater amount of technical information. The fact that the 12 most popular formats are bar plots may partly reflect a familiarity effect, as bar plots are a common visual convention. However, the consistently strong performance of graphical displays relative to text-based formats more broadly points to a higher baseline attractiveness of visual representations.

Please find below the disaggregated versions of Figure A27.

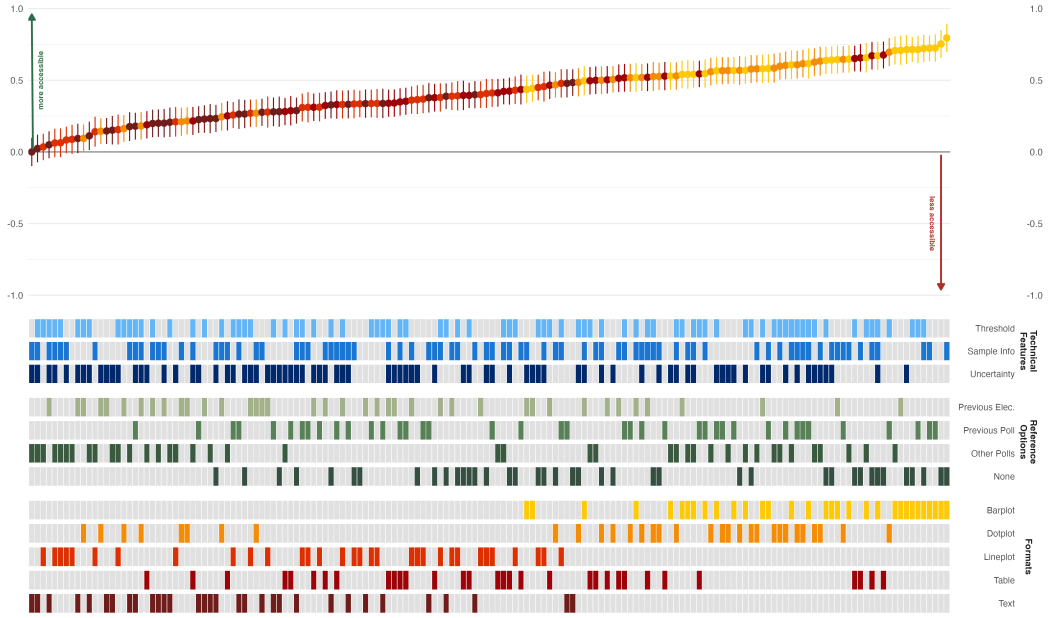


Figure A28: Figure A27 exclusively based on accessibility outcome.

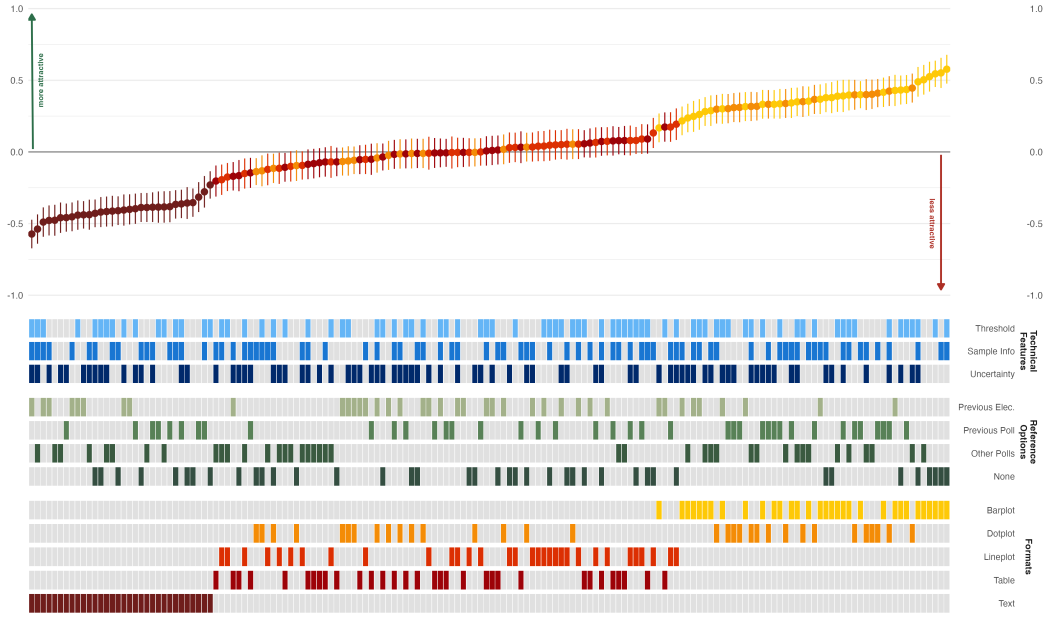


Figure A29: Figure A27 exclusively based on attractiveness outcome.

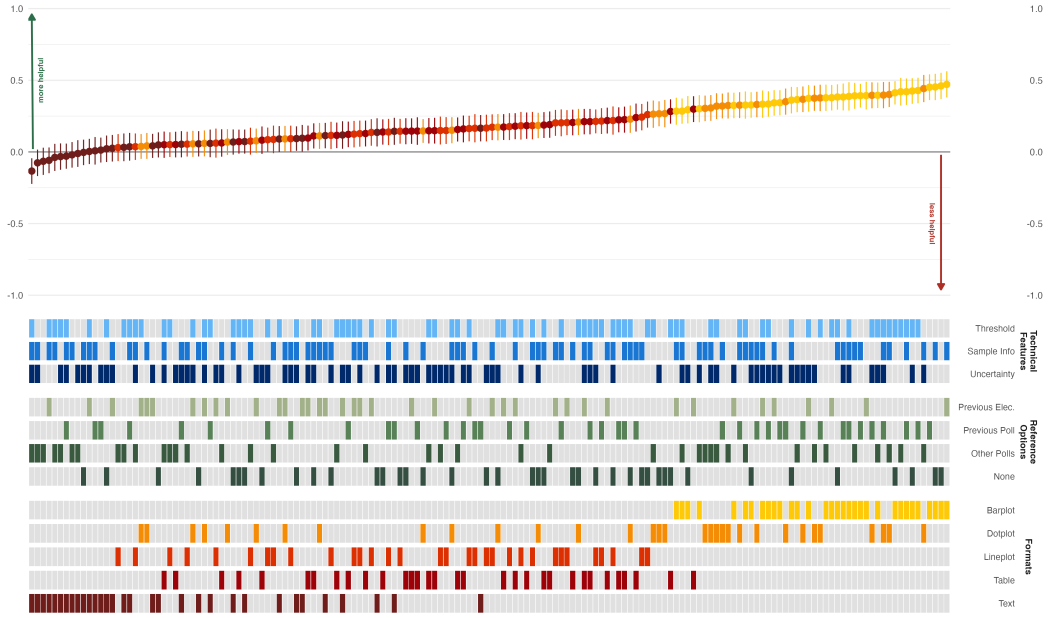


Figure A30: Figure A27 exclusively based on helpfulness outcome.

SI References

1. A. Neundorf & A. Öztürk. *Recruiting Research Participants through Facebook Advertisements: A Handbook*. 2021. DOI: 10.31219/osf.io/87rg3.
2. A. Neundorf & A. Öztürk. "How to improve representativeness and cost-effectiveness in samples recruited through meta: A comparison of advertisement tools". In: *PLOS ONE* 18.2 (2023). DOI: 10.1371/journal.pone.0281243.
3. A. Neundorf & A. Öztürk. "Advertising Online Surveys on Social Media: How Your Advertisements Affect Your Study". In: *Public Opinion Quarterly* 89.2 (2025), pp. 335–360. DOI: 10.1093/poq/nfaf018.
4. H. Wickham, W. Chang, L. Henry, T. L. Pedersen, K. Takahashi, C. Wilke, K. Woo, H. Yutani, D. Dunnington & RStudio. *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. 2022.
5. C. Erfort, L. F. Stoetzer, T. Gschwend, E. Koch, S. Munzert & H. Rajski. "The Zweitstimme Forecast for the German Federal Election 2025: Coalition Majorities and Vacant Districts". In: *PS: Political Science & Politics* (2025), pp. 1–12. DOI: 10.1017/S1049096525000150.
6. S. Ramirez-Ruiz & S. Munzert. *Measuring scientific evidence consumption literacy for public policy: Development and validation of the INSPIRE inventory*. 2026.
7. T. Faas & T. Klingelhöfer. "Same old, same old? The return of the (not so) Grand Coalition after the 2025 German Bundestag election". In: *West European Politics* 49.4 (2026), pp. 1116–1132. DOI: 10.1080/01402382.2025.2600960.
8. T. J. Leeper, S. B. Hobolt & J. Tilley. "Measuring subgroup preferences in conjoint experiments". In: *Political Analysis* 28.2 (2020), pp. 207–221.