Supplementary Information for: Boosting media literacy using lateral reading and online search interventions

February 12, 2026

1 Treatment Instructions

1.1 Lateral Reading

Instruction Bitte sehen Sie sich dieses kurze Video an. Es ist wichtig, dass Sie sich das gesamte Video ansehen. Das Video erklärt, wie man „lateral“ liest - eine Strategie, die Menschen, die professionell Fakten prüfen, anwenden, um die Vertrauenswürdigkeit von Online-Quellen zu beurteilen. Nach dem Video werden wir Ihnen eine Frage dazu stellen. Das Video dauert 5 Minuten, danach können Sie mit der Umfrage fortfahren.

Video Script Wir leben in einer Welt des Informationsüberflusses. Täglich begegnen wir zahlreichen neuen Informationen im Internet, aber statt alles für bare Münze zu nehmen, sollten wir uns zunächst immer diese wichtigen Fragen stellen: Wer steckt hinter den Informationen? Was sind die Ziele und werden diese Ziele transparent und offen kommuniziert? Oder wird hier versucht zu manipulieren?
Die Stanford History Education Gruppe führte eine Studie durch. An der Studie nahmen Professorinnen und Professoren, Studierende und Leuten die professionell Fakten checken teil. Gesucht wurde die effektivste Methode zur Bewertung von Informationen im Internet. Dabei entdeckten sie drastische Unterschiede in der Art und Weise, wie Menschen das Internet nutzten. Viele der Professorinnen und Professoren, sowie auch der Studierenden konzentrierten sich lange nur auf die zu beurteilende Webseite, lasen von oben nach unten (also “vertikal”), scrollten hoch und runter, beurteilten die Textqualität und betrachteten Abbildungen und das Gesamterscheinungsbild. Sie ließen sich von einem offiziell wirkenden Logo oder dem Namen der Organisation täuschen. So maßen sie zum Beispiel auch der offiziell klingenden Endung ".org" eine hohe Bedeutung bei, dabei ist .org eine offene Domain die sich jede Person einfach kaufen kann. Das Gleiche gilt übrigends auch für .de oder .eu. Sowohl die Professorinnen und Professoren als auch die Studierenden lasen die vermeintlich seriösen Referenzen und „Forschungsberichte“, ohne sich darüber im klaren zu sein, dass diese keine wissenschaftlichen Standards erfüllten. Die Studie zeigt also, dass selbst gut ausgebildete Menschen mit kritischem Denkvermögen sich bei dieser Art Webseiten zu lesen, in die Irre führen lassen.
Leute die professionell Fakten checken gingen dagegen anders an die Sache heran. Sie sind sich bewusst, dass das Internet ein tückisches Terrain ist und man sich leicht vom Erscheinungsbild einer Webseite täuschen lassen kann. Sie wissen, dass man im Netz oft nicht sieht, wer oder was hinter einer Webseite steht. Die Faktenchckerinnen und Faktenchecker hielten sich daher nicht lange auf einer unbekannten Webseite auf, sondern verließen sie wieder, um sich anderweitig über die Webseite zu informieren. Sie suchten nach Informationen, wer hinter der Webseite steckt.

So konnten die Profis viel schneller die richtigen Schlüsse ziehen. Diese Strategie nennt man „laterales Lesen“, also “seitwärts” lesen.
Hier ein kurzes Beispiel: Das hier ist EIKE, das europäische Institut für Klima und Energie. Der Name der Institution klingt schonmal professionell. Ein schickes Logo? Ebenfalls vorhanden. Kein schlechter Eindruck, wenn man die Seite vertikal betrachtet und ein wenig hoch- und runterscrollt. Faktencheckerinnen und Faktenchecker würden sofort einen neuen Tab öffnen und schauen wer hinter den Informationen steckt. Sie praktizieren laterales Lesen. Also gebe ich „EIKE Institut“ in die Suchleiste ein und überfliege die Suchergebnisse. Ich nehme mir ein bisschen Zeit um die Suchergebnisse zu scannen. Auf dieser Seite kann ich mir schon prima einen Überblick verschaffen und gucken welche Seiten ich dann öffnen möchte. Es ist ratsam mehrere Quellen zu Rate zu ziehen und natürlich nicht auf die eigentliche Webseite zu klicken. Ich will ja schließlich schauen, was andere Quellen über das Eike Institut zu sagen haben. Ich finde hier einen Wikipedia-Artikel. Möglicherweise habt ihr gehört, dass Wikipedia nicht vertrauenswürdig ist. Tatsächlich folgt Wikipedia aber Richtlinien, die die Qualität der bereitgestellten Informationen gewährleisten. Wikipedia kann also als hilfreicher Ausgangspunkt für eine Recherche dienen. Besonders hilfreich sind die Quellenangaben am Ende eines Wikipedia-Eintrags. Ich öffne den Wikipedia-Artikel also mit einem Rechtsklick in einem neuen Tab. In meinen Suchresultaten finde ich außerdem einen Eintrag über das Institut auf Lobbypedia. Lobbypedia ist ein unabhängiges, lobbykritisches Online-Lexikon. Ich öffne auch diese Seite in einem neuen Tab. Zuletzt öffne ich noch einen Artikel von der Süddeutschen Zeitung. Beim Überfliegen des Wikipedia-Artikels springt mir direkt folgendes ins Auge: Hier steht, dass das Europäische Institut für Klima und Energie den wissenschaftlichen Konsens über die menschengemachte globale Erwärmung leugnet. Entgegen seines Namens, handelt es sich nicht um ein wissenschaftliches Institut, sondern um eine Lobbyorganisation. Belegt wird das zum Beispiel mit einer wissenschaftlichen Publikation aus dem Jahr 2020. Auch der Beitrag auf Lobbypedia beschreibt EIKE als Verein, dessen Ziel es ist, den menschengemachten Klimawandel zu leugnen und weitere Verschwörungstheorien zu verbreiten. Die wenigen Publikationen, die EIKE auf ihrer Webseite nennt, sind nicht von wissenschaftlich anerkannten Klimaforscherinnen oder forschern verfasst worden. Auch im Artikel der Süddeutschen Zeitung finde ich Informationen, die genau das bestätigen: Es handelt sich bei Eike um gar kein richtiges Institut, sondern einer Gruppe aus Menschen die den Klimawandel anzweifeln. Bei EIKE handelt es sich also um keine vertrauenswürdige und unabhängige Quelle, um etwas über Klima und Energie zu lernen. Dies haben wir durch laterales Lesen herausgefunden.
Die Forschungsergebnisse der Stanford Gruppe und weitere wissenschaftliche Studien haben zeigen können, dass laterales Lesen leicht erlernbar ist. Menschen, die in lateralem Lesen geschult wurden, zeigten im Vergleich zu einer Kontrollgruppe eine signifikante Verbesserung ihrer Fähigkeiten zur adäquaten Bewertung von Webseiten. Laterales Lesen hilft Menschen dabei, Informationen im Netz besser (und schneller) zu prüfen.

1.2 Online Search Instructions

Online search treatment instructions were translated from Aslett *et al.* [1].

Instruction In diesem Abschnitt geht es darum, Belege aus einer anderen Quelle für die zentrale Aussage des Artikels zu finden, den Sie bewerten sollen. Diese Belege sollten es Ihnen ermöglichen, zu beurteilen, ob die zentrale Aussage wahr, falsch oder irgendetwas dazwischen ist.
Anleitung für die Suche nach Belegen für oder gegen die zentrale Aussage, die Sie identifiziert haben:

1. Unter einem Beleg verstehen wir einen Artikel, eine Aussage, ein Foto, ein Video, eine Audioaufnahme oder eine Statistik, die für die zentrale Aussage des Artikels relevant sind.

Dieser Beleg sollte von einer anderen Quelle als dem Autor des Artikels, den Sie untersuchen, stammen. Diese Belege können entweder die ursprüngliche Behauptung unterstützen oder ihr widersprechen.

2. Um Belege für die Behauptung zu finden, sollten Sie eine Stichwortsuche in einer Suchmaschine Ihrer Wahl oder auf der Website einer bestimmten Quelle durchführen, der Sie bezüglich des jeweiligen Themas vertrauen.

3. Wir bitten Sie, für die Bewertung der zentralen Aussage in Ihrer Suche die hochwertigsten Belege zu verwenden, die Sie finden können. Wenn Sie bei keiner Quelle der Sie vertrauen Belege für die Behauptung finden können, sollten Sie versuchen, die relevantesten Belege für die Behauptung aus einer beliebigen Quelle zu finden, auch wenn Sie ihr nicht vertrauen.

2 Methodological Details

2.1 Preregistered decisions on random-effects structure of the primary models

Initially, the following full random-effects structure was specified in the preregistration:

```
rating ~ item_type * phase * treatment +
    (1 + item_type + phase | id) +
    (1 + phase + treatment | item)
```

- `rating` refers to the source-trustworthiness rating assigned to an item (model 1s, H1s) or the claim credibility assigned to an item (model 1c, H1c);
- `treatment` is a factor variable representing the three conditions;
- `item_type` is a factor variable indicating whether the sources in the item are trustworthy or untrustworthy, thus quantifying discernment ability (i.e., the difference between mean trust judgments in trustworthy sources and mean trust judgments in untrustworthy sources);
- `phase` is a factor variable distinguishing between pre- vs. post-intervention judgments;
- `id` serves as an anonymous unique identifier for each participant; and
- `item` is an unique identifier for each item.

Implementing this structure led to issues of singular fit.
As preregistered and following Bates *et al.* [2], we simplified the initial structure to the following model specification:

```
rating ~ item_type * phase * treatment
    + (1 + item_type + phase | id)
    + (1 | item)
```

that is, we dropped the item-random slope for treatment.
For consistency, we kept this model specification for any downstream analyses.
Initially, we considered estimating our models with `rlmer()` instead of `lmer()` but given our sample size, `rlmer()` could not compute effects within a reasonable timeframe. As preregistered, we therefore defaulted to the widely used and well supported `lmer()`.

2.2 Exploration of strategy adoption using web tracking data

A subset of participants ($N = 436$) were members of YouGov's Pulse Panel that collects browser histories using passive metering technology implemented in a browser plug-in that participants install actively under informed consent. Among these participants $n = 202$ has visible traces of the experiment (the specific URLs used as experimental stimuli within the survey taking time recorded in their survey data). For the remaining participants the wrong device was tracked — for example, participants may have installed the tracking software on their desktop PC while they took the survey and participated in the website evaluation task of the experiment on their mobile phones.

Among this subset of $n = 202$, we explored traces of lateral reading and online search strategy adoption. We focused on participants' browsing behavior following visits to experimental websites. We extract sequences after experimental website visits up to 50 following clicks. If a subsequent experimental click exists in the candidate window, we cut off the sequence just before it.

As proxy for online search strategy adoption, we identify instances of search as visits to google.com/search and extract the search query using regular expression. We focus on Google as most widely used search engine, that consistently holds search engine market shares of over 90%. We inspect search queries manually, confirming that these were searches related to experimental stimuli. We then enumerate the website visits after experimental website visits according to their relative click position (the position in the sequence) and compute the proportion of searches for each relative click position. Results are reported in main Figure 3D, broken down by experimental condition.

3 Supporting Tables and Figures

4 Stimulus Websites Selection

The stimulus websites for this study, to assess participant's trustworthiness discernment and information veracity discernment, were selected in a bottom-up approach using an externally fact-checked database of German online news sources (NewsGuard), a systematic coding protocol for article selection and a separate empirical pretest study to select the most informative stimulus items for the main study. The following sections provide an overview over this pretest study as well as a summary of the results on the basis of which stimuli for the main study were selected.

4.1 Protocol

The starting point for the curation of a stimulus set was NewsGuard, as largest database for credibility ratings for German news sources [6]. All available news sources with locale = "Germany" and rating level = "proceed with maximum caution" ($N = 43$) were selected. Duplicate websites under different domains, websites that no longer existed, websites with insecure browser connection and websites that host videos only were excluded.

The remaining set of websites was split among 3 researchers and browsed for up to three articles, covering different political topics relevant for the German public discourse that provide concrete examples of misinformation. More specifically, to ensure external fact-checking, a reverse strategy was used: researchers read NewsGuard's profiles about the sources that included fact-checked examples of articles containing misinformation.

The selected set of example articles ($N = 48$) covered misinformation regarding six different political issues: Covid-19, climate change, migration, conspiracies about international organizations, elections and the war in Ukraine. Corresponding to each political issue, three credible articles

Table S1: Sample Characteristics Compared to Benchmarks

Variable		**Sample**[1]	**Benchmark**[2,3]
Age	(excluding under 18)	51 (15)	52 (20)
Gender	Male	1,385 (52%)	49%
	Female	1,281 (48%)	51%
Education	Still in school	17 (0.6%)	3.6%
	Haupt-/Volksschule	565 (21%)	24.4%
	Realschule or equivalent	981 (37%)	29.5%
	Abitur/Fachhochschulreife	1,077 (40%)	36.6%
	No degree	23 (0.9%)	5.9%
State (Bundesland)	Schleswig-Holstein	112 (4.2%)	3.5%
	Hamburg	66 (2.5%)	2.2%
	Niedersachsen	257 (9.6%)	9.6%
	Bremen	23 (0.9%)	0.8%
	Nordrhein-Westfalen	576 (22%)	21.6%
	Hessen	214 (8.0%)	7.5%
	Rheinland-Pfalz	126 (4.7%)	4.9%
	Baden-Württemberg	287 (11%)	13.5%
	Bayern	390 (15%)	15.8%
	Saarland	34 (1.3%)	1.2%
	Berlin	128 (4.8%)	4.4%
	Brandenburg	77 (2.9%)	3.0%
	Mecklenburg-Vorpommern	60 (2.3%)	1.9%
	Sachsen	163 (6.1%)	4.9%
	Sachsen-Anhalt	72 (2.7%)	2.6%
	Thüringen	81 (3.0%)	2.5%
Second Vote (Intention)	CDU/CSU	677 (25%)	28.6%
	SPD	394 (15%)	16.4%
	Die Grünen	326 (12%)	11.6%
	FDP	73 (2.7%)	4.3%
	Die Linke	83 (3.1%)	8.8%
	BSW	152 (5.7%)	4.9%
	AfD	418 (16%)	20.8%
	Andere	103 (3.9%)	4.4%
	Weiß nicht	440 (17%)	—

[1] Mean (SD); n (%).
[2] Benchmarks from Destatis Mikrozensus 2024 (sheets: 12411-0005 (age), 12411-0003 (gender), 12211-0205 (education), 12411-0010 (Bundesland/state, Bundesamt [3]); and [3] Bundeswahlleiterin [4].

Figure S1: Supplementary descriptive reporting of main experiment. Top: average ratings with bootstrapped 95% confidence intervals (displayed on raw scale, see bottom for full credibility scale); center: website difficulty (percentage correct ratings) and website familiarity (having encountered the website before); bottom: raw claim credibility ratings.

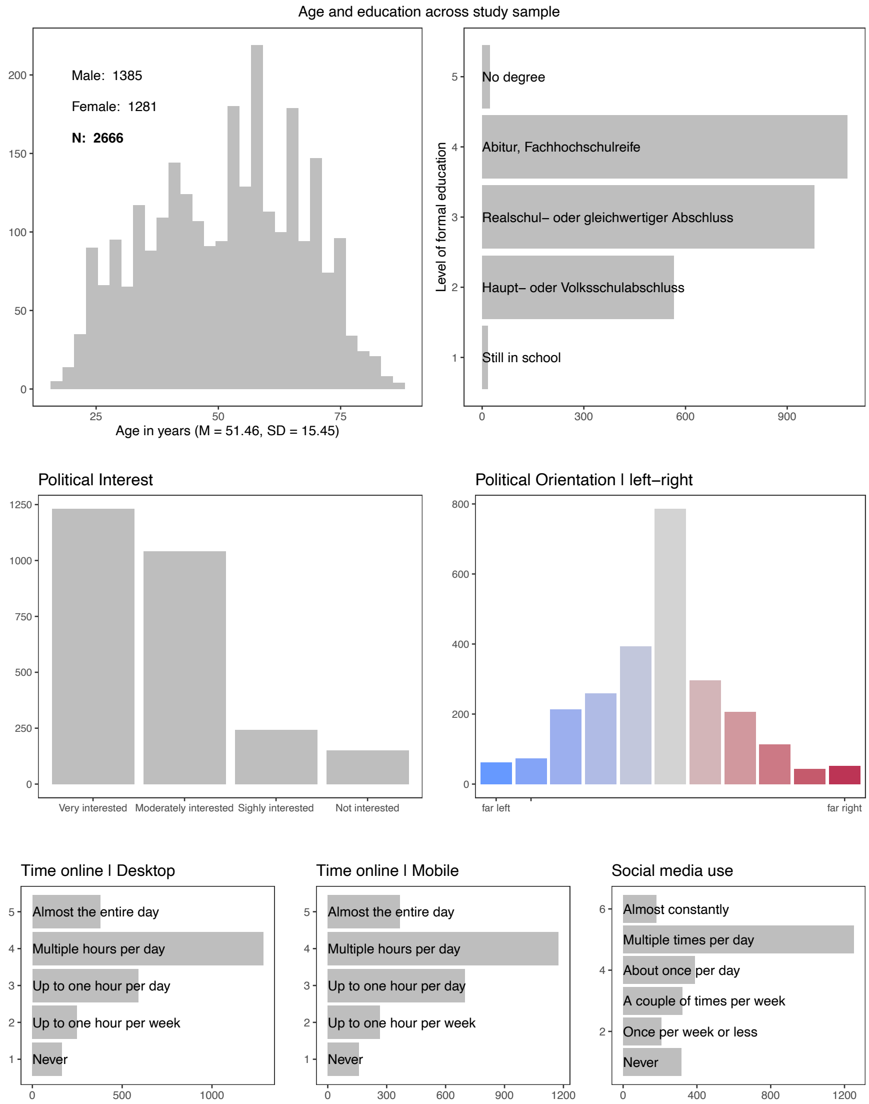


Figure S2: Demographic description of sample. Top: age and level of formal education; center: political interest and self-placement of left-right scale of political orientation; bottom: self-report indicators of online activity.

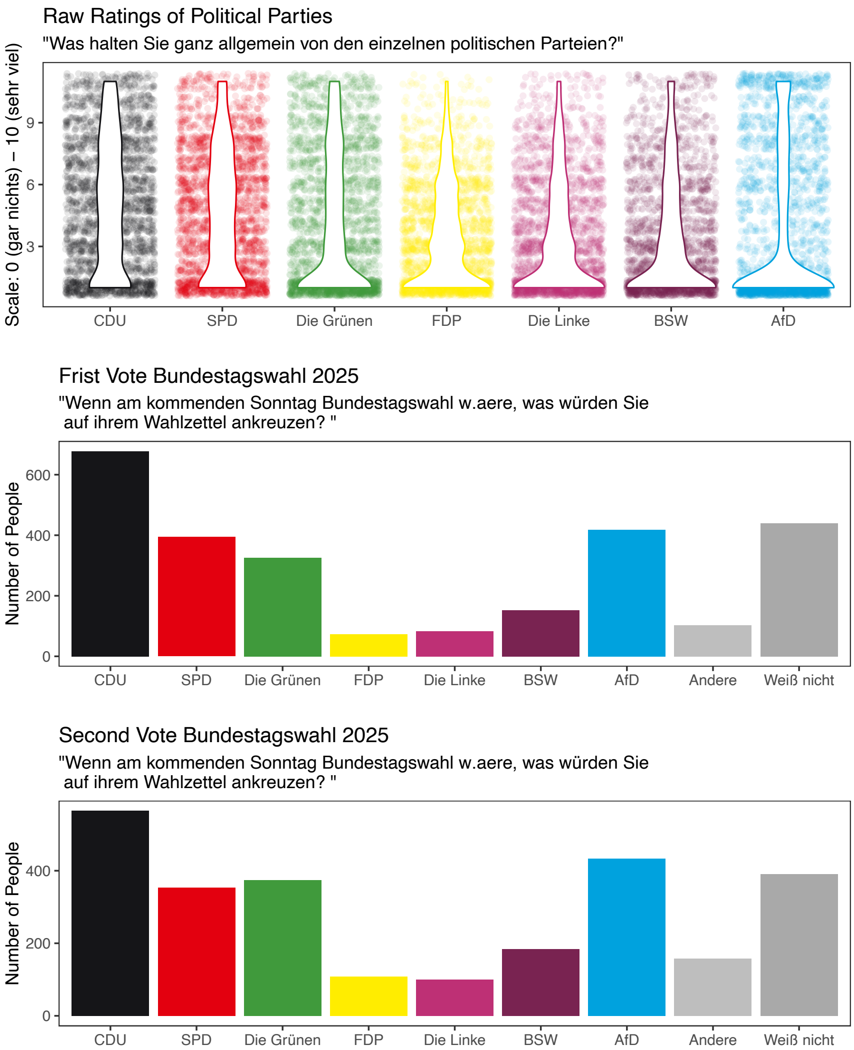


Figure S3: Party preferences among sample. Top: general rating of party; center: voting intention first vote (candidate) federal election; bottom: voting intention second vote (seats in parliament) federal election.

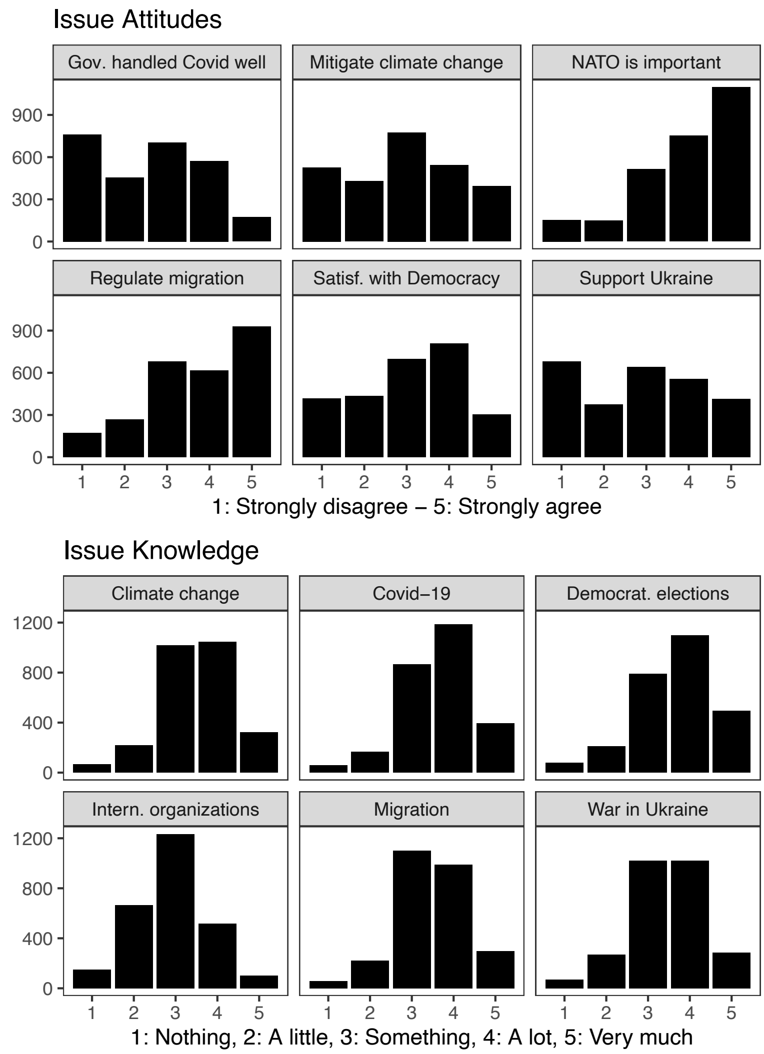
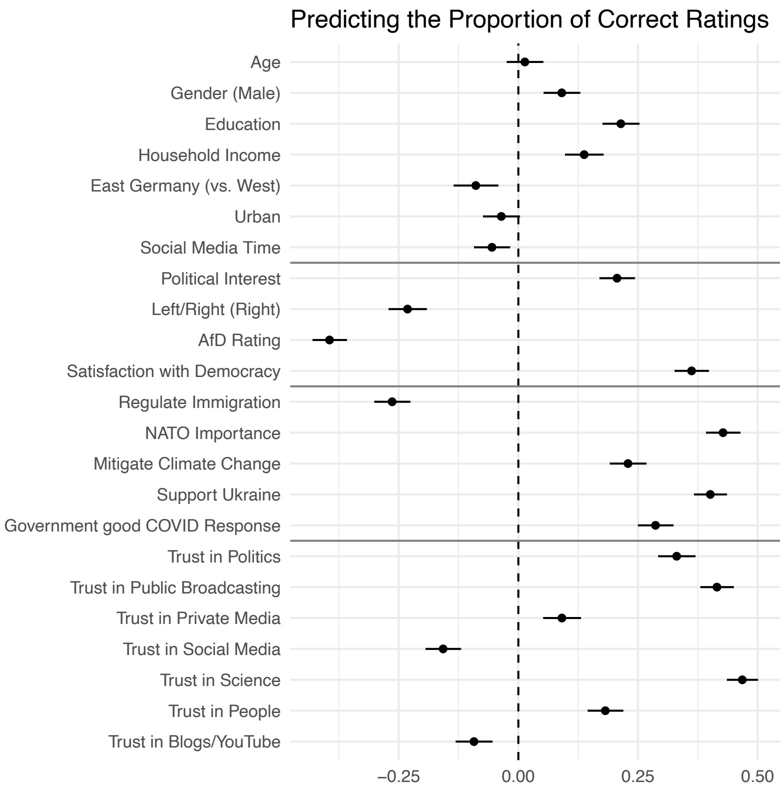


Figure S4: Survey indicators regarding six different political issues that were subject of experimental stimulus websites. Top: issue attitudes; bottom: self-reported issue knowledge.



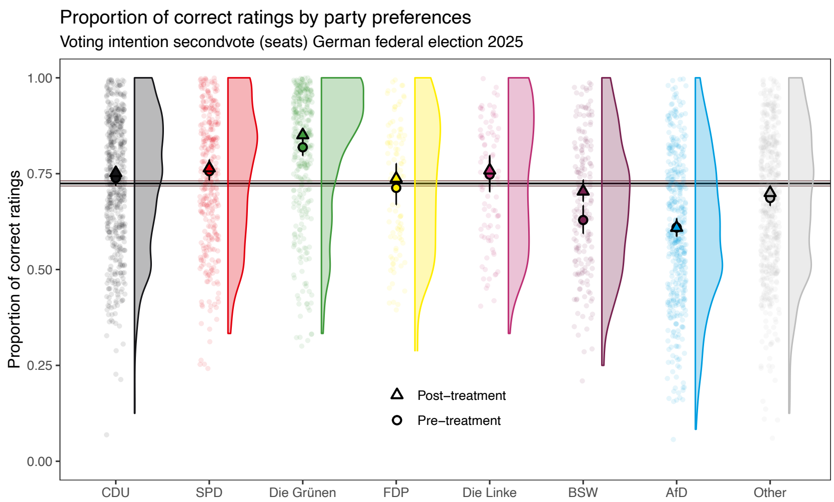


Figure S5: Differences in rating accuracy. Top: coefficient plot of predictors for proportion of correct ratings, variables scaled by dividing by two standard deviations to allow direct comparison of binary and continuous variables [5]; bottom: pre-treatment and post-treatment average proportions of correct ratings by secondvote party preference. Raw data displays pre-treatment and post-treatment ratings.

Table S2: Preregistered treatment effects (pre vs. post; pre vs. follow-up) with 90% CIs

Treatment	Estimate	SE	Outcome	90% CI Lower	90% CI Upper	Phase
Lateral reading	0.011	0.007	Source trustworthiness	-0.001	0.023	1. Pre-post
Online search	0.015	0.007	Source trustworthiness	0.003	0.027	1. Pre-post
Lateral reading	0.016	0.007	Claim credibility	0.003	0.028	1. Pre-post
Online search	0.014	0.008	Claim credibility	0.002	0.027	1. Pre-post
Lateral reading	0.007	0.008	Source trustworthiness	-0.006	0.020	2. Pre-follow-up
Online search	-0.004	0.008	Source trustworthiness	-0.018	0.009	2. Pre-follow-up
Lateral reading	0.005	0.008	Claim credibility	-0.008	0.019	2. Pre-follow-up
Online search	-0.002	0.008	Claim credibility	-0.016	0.012	2. Pre-follow-up

Table S3: Treatment effects - controlling for source familiarity

Treatment	Estimate	SE	Outcome	90% CI Lower	90% CI Upper	Phase
Lateral Reading	0.012	0.007	Source Trustworthiness	0.000	0.024	1. Pre-Post
Online Search	0.015	0.007	Source Trustworthiness	0.003	0.027	1. Pre-Post
Lateral Reading	0.017	0.007	Claim Credibility	0.004	0.029	1. Pre-Post
Online Search	0.015	0.007	Claim Credibility	0.002	0.027	1. Pre-Post
Lateral Reading	0.009	0.008	Source Trustworthiness	NA	NA	2. Pre-Followup
Online Search	-0.003	0.008	Source Trustworthiness	NA	NA	2. Pre-Followup
Lateral Reading	0.007	0.008	Claim Credibility	NA	NA	2. Pre-Followup
Online Search	-0.001	0.008	Claim Credibility	NA	NA	2. Pre-Followup

Table S4: Treatment effects | ordinal outcome (CLMM)

Treatment	Estimate	SE	p-value	Outcome	90% CI Lower	90% CI Upper	Phase
Lateral Reading	0.112	0.068	0.101	Source Trustworthiness	0.000	0.225	1. Pre-Post
Online Search	0.145	0.069	0.035	Source Trustworthiness	0.032	0.259	1. Pre-Post
Lateral Reading	0.150	0.067	0.026	Claim Credibility	0.039	0.261	1. Pre-Post
Online Search	0.138	0.068	0.043	Claim Credibility	0.026	0.250	1. Pre-Post
Lateral Reading	0.062	0.075	0.406	Source Trustworthiness	-0.061	0.185	2. Pre-Followup
Online Search	-0.022	0.075	0.768	Source Trustworthiness	-0.146	0.102	2. Pre-Followup
Lateral Reading	0.049	0.074	0.506	Claim Credibility	-0.072	0.170	2. Pre-Followup
Online Search	0.001	0.074	0.989	Claim Credibility	-0.121	0.123	2. Pre-Followup

Table S5: Treatment effects | bayesian models (BRMS)

Treatment	Estimate	SE	Outcome	90% CI Lower	90% CI Upper	Phase
Lateral Reading	0.113	0.068	Source Trustworthiness	0.001	0.226	1. Pre-Post
Online Search	0.147	0.069	Source Trustworthiness	0.033	0.262	1. Pre-Post
Lateral Reading	0.152	0.069	Claim Credibility	0.039	0.265	1. Pre-Post
Online Search	0.140	0.069	Claim Credibility	0.026	0.254	1. Pre-Post
Lateral Reading	0.063	0.076	Source Trustworthiness	-0.062	0.189	2. Pre-Followup
Online Search	-0.021	0.076	Source Trustworthiness	-0.147	0.104	2. Pre-Followup
Lateral Reading	0.048	0.073	Claim Credibility	-0.073	0.168	2. Pre-Followup
Online Search	0.002	0.076	Claim Credibility	-0.123	0.126	2. Pre-Followup

were selected from different sources that were labelled as trustworthy by NewsGuard, resulting in a total set of 66 online news articles (48 false, 18 true). The procedure is summarized by the flowchart in Fig. S12.

Table S6: Fixed effect coefficients for all preregistered models

Model	Term	β	SE	t	95% CI
Model 1s: Source Trustworthiness (Pre-Post)	(Intercept)	-0.034	0.017	-1.96	[-0.067, 0.000]
	item_type	0.142	0.017	8.33	[0.109, 0.175]
	phasepost	0.022	0.006	3.76	[0.010, 0.033]
	treatmentlateral_reading	0.009	0.009	1.07	[-0.008, 0.026]
	treatmentonline_search	-0.007	0.009	-0.81	[-0.024, 0.010]
	item_type:phasepost	0.000	0.005	0.02	[-0.010, 0.010]
	item_type:treatmentlateral_reading	-0.001	0.008	-0.08	[-0.016, 0.015]
	item_type:treatmentonline_search	-0.001	0.008	-0.16	[-0.017, 0.014]
	phasepost:treatmentlateral_reading	-0.027	0.008	-3.34	[-0.043, -0.011]
	phasepost:treatmentonline_search	0.011	0.008	1.34	[-0.005, 0.027]
	item_type:phasepost:treatmentlateral_reading	**0.011**	0.007	1.53	[-0.003, 0.025]
	item_type:phasepost:treatmentonline_search	**0.015**	0.007	2.00	[0.000, 0.029]
Model 2s: Source Trustworthiness (Pre-Follow-up)	(Intercept)	-0.034	0.018	-1.89	[-0.069, 0.001]
	item_type	0.142	0.018	8.00	[0.107, 0.177]
	phasefollow-up	0.029	0.007	4.15	[0.015, 0.042]
	treatmentlateral_reading	0.009	0.009	1.10	[-0.007, 0.026]
	treatmentonline_search	-0.007	0.009	-0.81	[-0.024, 0.010]
	item_type:phasefollow-up	0.024	0.006	4.18	[0.013, 0.035]
	item_type:treatmentlateral_reading	-0.001	0.008	-0.09	[-0.016, 0.015]
	item_type:treatmentonline_search	-0.001	0.008	-0.16	[-0.017, 0.014]
	phasefollow-up:treatmentlateral_reading	-0.011	0.010	-1.18	[-0.031, 0.008]
	phasefollow-up:treatmentonline_search	-0.005	0.010	-0.48	[-0.024, 0.014]
	item_type:phasefollow-up:treatmentlateral_reading	**0.007**	0.008	0.87	[-0.009, 0.023]
	item_type:phasefollow-up:treatmentonline_search	**-0.004**	0.008	-0.55	[-0.020, 0.011]
Model 1c: Claim Credibility (Pre-Post)	(Intercept)	-0.008	0.017	-0.47	[-0.041, 0.025]
	item_type	0.139	0.017	8.39	[0.107, 0.172]
	phasepost	0.014	0.006	2.43	[0.003, 0.026]
	treatmentlateral_reading	0.010	0.009	1.12	[-0.007, 0.026]
	treatmentonline_search	-0.013	0.009	-1.55	[-0.030, 0.004]
	item_type:phasepost	0.004	0.005	0.78	[-0.006, 0.015]
	item_type:treatmentlateral_reading	-0.004	0.008	-0.54	[-0.020, 0.011]
	item_type:treatmentonline_search	-0.002	0.008	-0.28	[-0.018, 0.013]
	phasepost:treatmentlateral_reading	-0.026	0.008	-3.10	[-0.042, -0.010]
	phasepost:treatmentonline_search	0.019	0.008	2.21	[0.002, 0.035]
	item_type:phasepost:treatmentlateral_reading	**0.016**	0.007	2.09	[0.001, 0.030]
	item_type:phasepost:treatmentonline_search	**0.014**	0.008	1.90	[-0.000, 0.029]
Model 2c: Claim Credibility (Pre-Follow-up)	(Intercept)	-0.008	0.017	-0.46	[-0.042, 0.026]
	item_type	0.139	0.017	8.20	[0.106, 0.173]
	phasefollow-up	0.020	0.007	2.82	[0.006, 0.033]
	treatmentlateral_reading	0.010	0.009	1.13	[-0.007, 0.027]
	treatmentonline_search	-0.013	0.009	-1.54	[-0.030, 0.004]
	item_type:phasefollow-up	0.027	0.006	4.68	[0.016, 0.039]
	item_type:treatmentlateral_reading	-0.004	0.008	-0.54	[-0.020, 0.011]
	item_type:treatmentonline_search	-0.002	0.008	-0.29	[-0.018, 0.013]
	phasefollow-up:treatmentlateral_reading	-0.013	0.010	-1.34	[-0.033, 0.006]
	phasefollow-up:treatmentonline_search	0.003	0.010	0.33	[-0.016, 0.023]
	item_type:phasefollow-up:treatmentlateral_reading	**0.005**	0.008	0.66	[-0.011, 0.022]
	item_type:phasefollow-up:treatmentonline_search	**-0.002**	0.008	-0.23	[-0.018, 0.014]

Note. All models include random intercepts for participant ID and item (link_number). Random slopes for item_type and phase are included by participant. Discernment effects are captured by the item_type × phase × treatment interaction. Reference categories: phase = pre, treatment = control; item_type is coded numerically with -1 representing untrustworthy/false items, 1 representing trustworthy/true items.

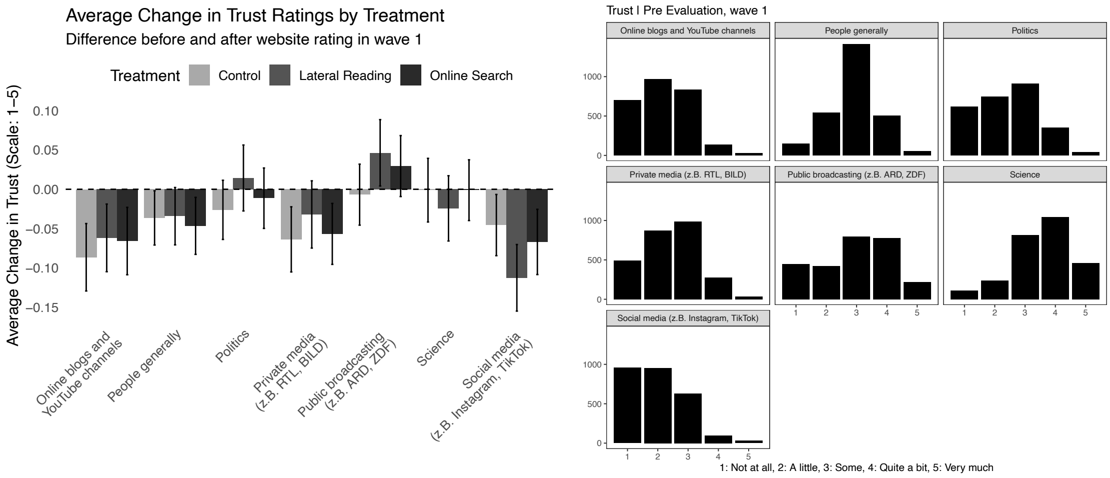


Figure S6: Trust. Right: baseline levels of trust in different media and institutions, ratings before any website evaluation; left: average change in trust ratings by treatment.

Table S7: Results of treatment contrasts (post vs. follow-up) with 90% CIs

Treatment	Estimate	SE	90% CI Lower	90% CI Upper	t-value	p-value
Control	0.023	0.007	0.012	0.034	3.491	0.000
Lateral reading	0.019	0.007	0.008	0.030	2.883	0.004
Online search	0.004	0.007	-0.008	0.015	0.519	0.604

4.2 Pretest

In total 66 online articles from 48 different sources were pretested with 301 participants on Prolific (quota sampled for gender) to select the most informative cases for the main study (see Fig. S13 for the pretest study design). Each participant was asked to assess in total 8 randomly selected articles, from the entire pool of untrustworthy and trustworthy articles. Participants assessed the trustworthiness of the source, the veracity of the claim, the political leaning of the source, as well as the familiarity with the source. Outcome distributions as well as IRT parameters, namely item difficulty and item discrimination, were calculated for each website to determine the informativeness of each stimulus for the main study.

Participants took a survey hosted on Qualtrics, that took, on average, 15 minutes to complete. Participants with response times below three minutes were excluded from the sample. After providing informed consent that included a comprehension check and responding to a short set of questions regarding demographic and political variables, participants were subsequently presented with 8 randomly selected stimulus websites. Participants were asked to follow the respective links to the websites, which opened in new browser tabs, and to take up to 2 minutes to rate the trustworthiness of the respective source, the credibility of the claim, the website's intent, as well as the familiarity and the political leaning of the website.

Participants were informed not to react to any pop-up messages when interacting with the live websites and this instruction was followed again by a comprehension check. Participants were debriefed for every stimulus they encounter during the study (see Fig. S11). They received a colour coded and easy to understand label (trustworthy vs. untrustworthy) and more detailed

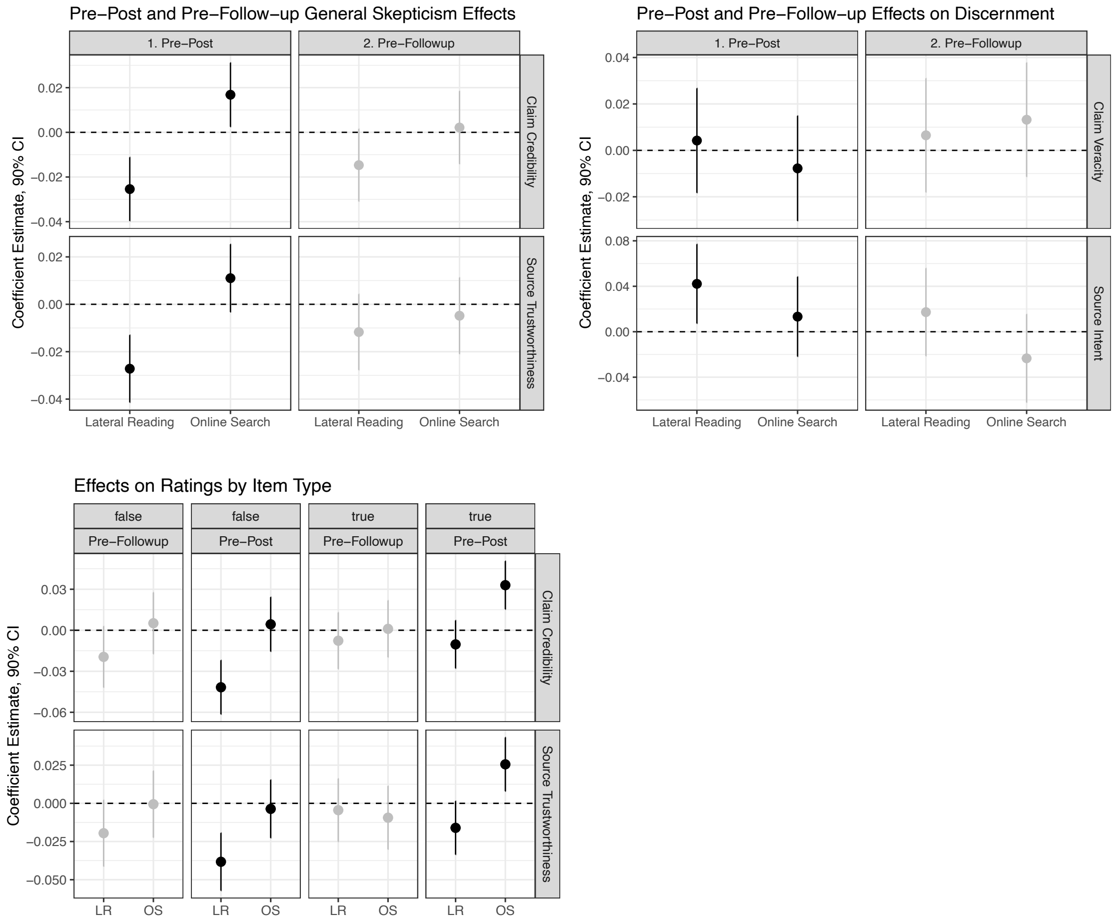


Figure S7: Alternative outcomes. Top left: treatment effects on ratings pooled over item types (removing interaction term `rating*item` from model) to consider general skepticism effect; bottom left: treatment effects on ratings, computed separately for different item types; top right: treatment effects on claim veracity discernment (binary rating: true vs. false) and treatment effects on source intent discernment (inform, convince, deceive).

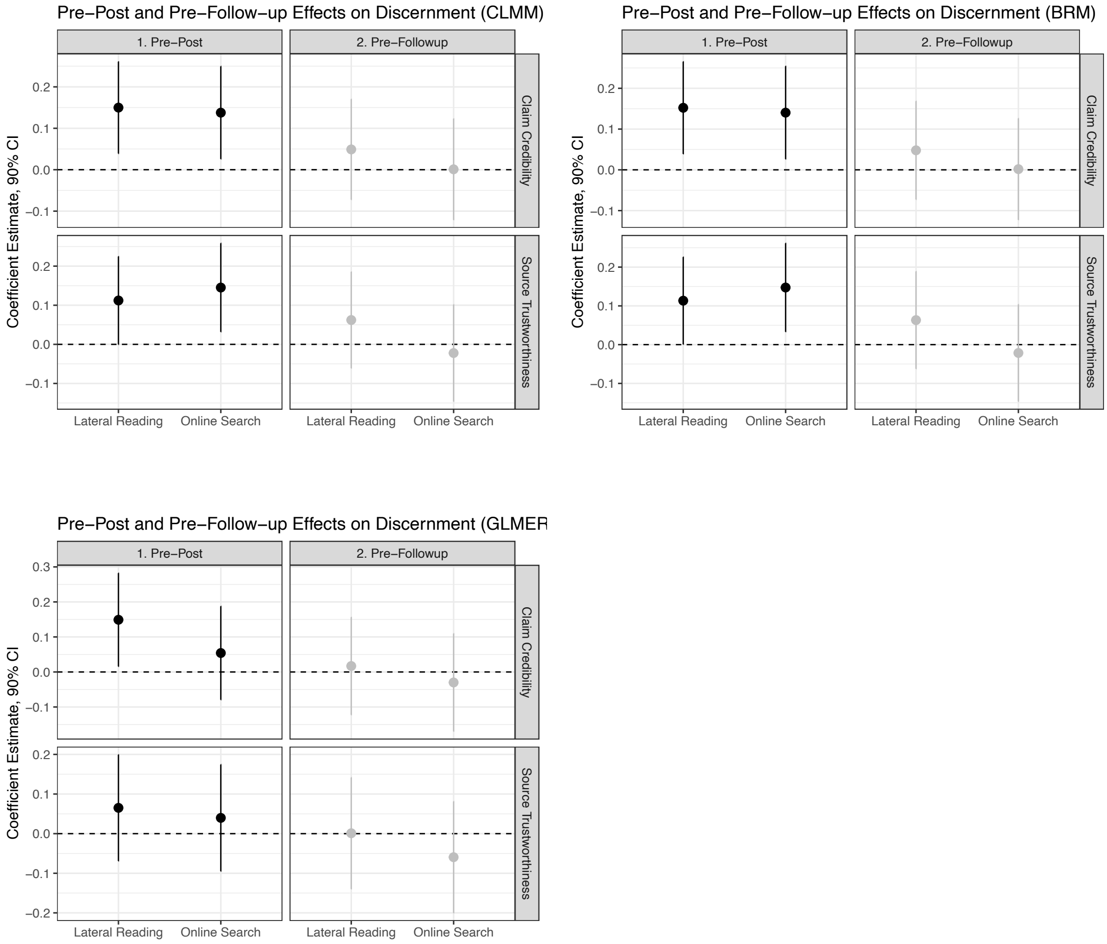


Figure S8: Alternative model specifications. Top left: ordinal handling of outcome variable, estimation with `ordinal::clmm()`; top right: bayesian model, estimation with `brms::brm()`; bottom: binarized outcome, estimated as `lme4::glmer()`.

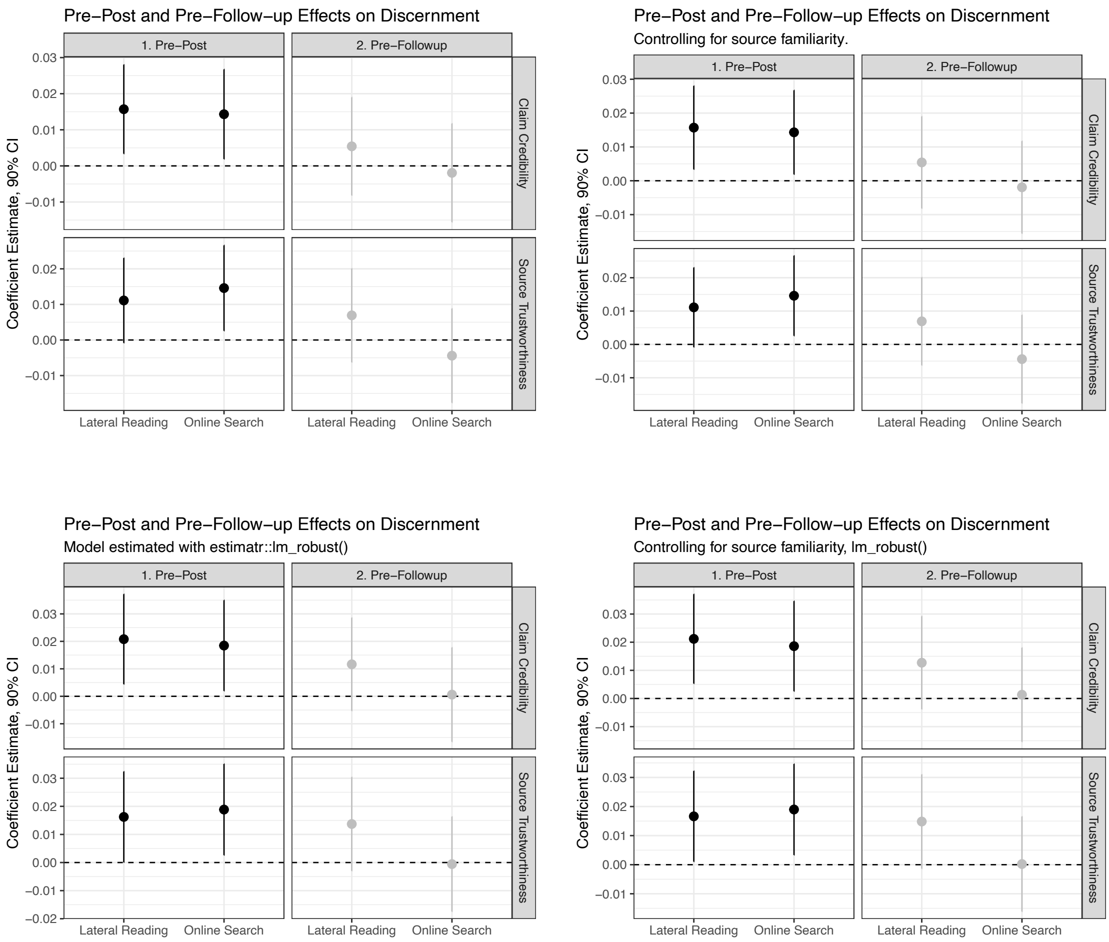


Figure S9: Alternative model specifications. Top left: main model estimated with `lme4::lmer()` for reference; top right: main model including source familiarity as control; bottom left: simplified model without multi-level structure estimated with `estimatr::lm_robust()`; bottom right: simplified model including source familiarity as control variable.

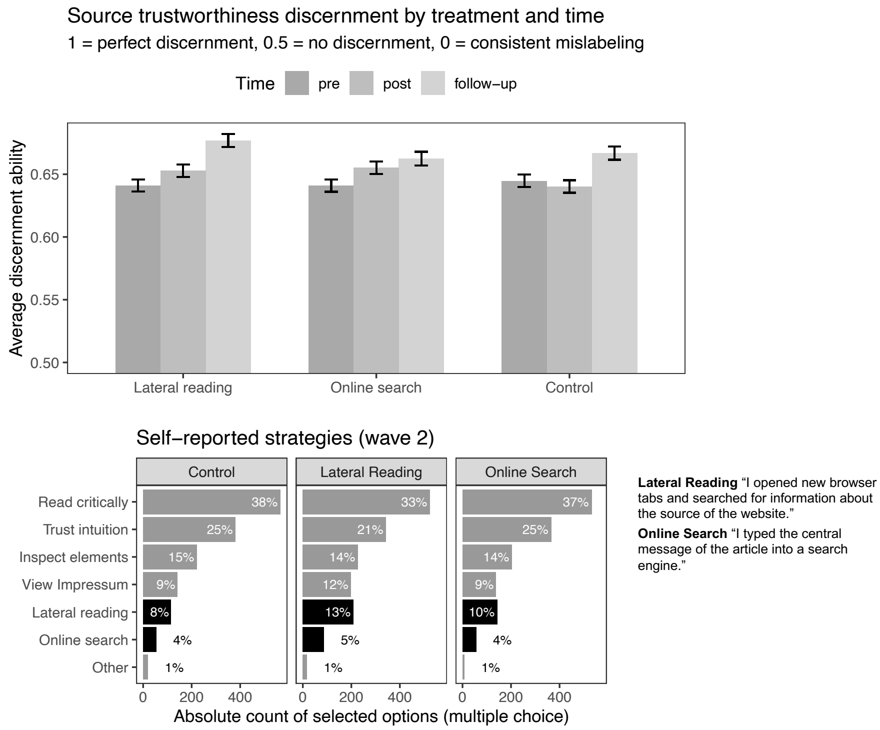


Figure S10: Top: Source trustworthiness discernment (derived from 4-point scale) by treatment group and time. Bottom: self-reported strategy use. This measure was collected at the end of wave 2.

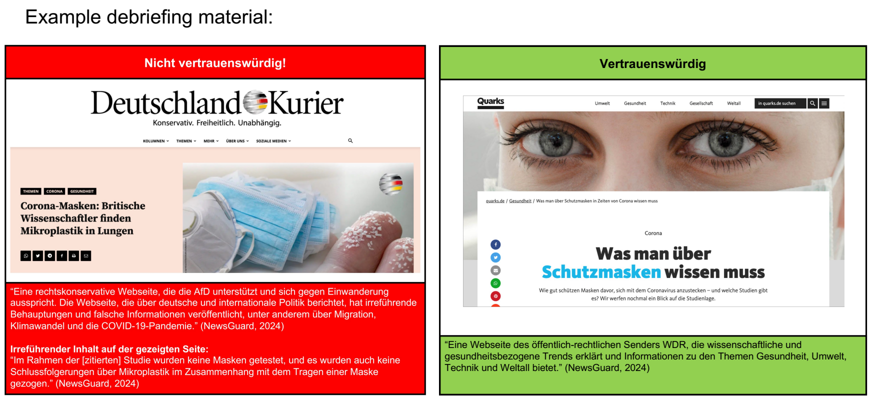


Figure S11: Example of debriefing material. Participants were debriefed at the end of wave 1 as well as at the end of wave 2. Debriefings were specific to the items people saw in the respective waves. The debriefings included a screenshot of the page, together with a colour-coded heading "turstworthy" or "untrustworthy", as well as an assessment of the source (quoted, with permission, from NewsGuard) and—only for the websites making false claims—an explanation for why the claim is false (also quoted from NewsGuard).

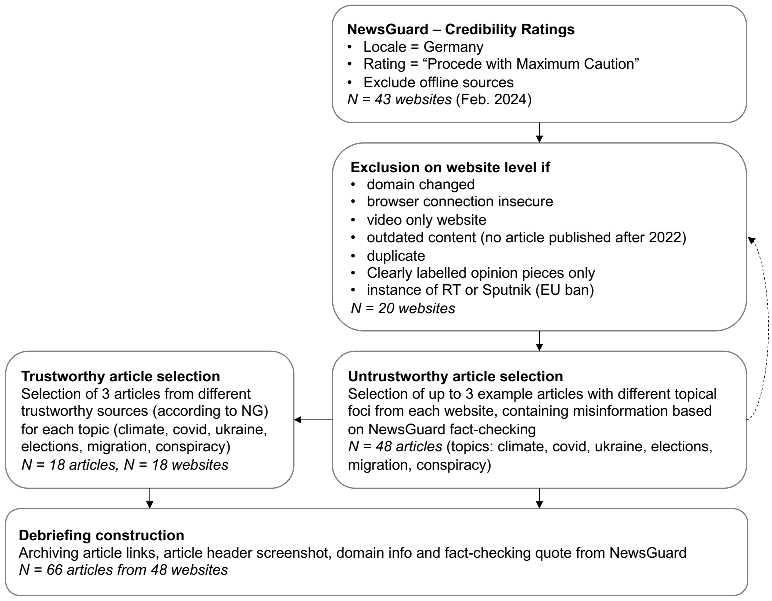


Figure S12: Stimulus Selection Protocol

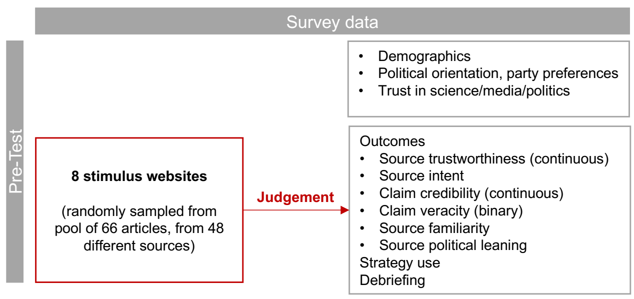


Figure S13: Stimulus Selection Pretest Design

information about the source and the specific claim made below a screenshot of the website. Both information were quotes from the professional fact-checking organization NewsGuard. We obtained permission from NewsGuard to provide the quotes to participants but the data may not be published in full to not violate their terms of service.
The pre-analysis plan of the pretest was registered under the Open Science Framework prior to data collection and the pretest study obtained individual ethical approval by the local ethics board on May 13, 2024.

4.3 Stimulus evaluation

In addition to the main outcome variables, source trustworthiness and claim veracity, three other variables were considered crucial outcomes for the pretest: website familiarity, perceived intent and perceived political leaning (for a summary of pretest outcomes, split by NewsGuard ratings (trustworthy vs. untrustworthy), see Fig. S14).
While the 18 trustworthy outlets showed an almost uniform distribution of familiarity, most untrustworthy outlets were unknown to participants. The dominant perceived intent of trustworthy websites was to 'inform' whereas the dominant perceived intent of untrustworthy websites was to 'convince'. Trustworthy outlets were perceived as 'neither left nor right' on a 5-point political left-right scale, whereas the average perceived political leaning of untrustworthy German websites was 'rather right wing' (see Fig. S14 A).
Fig. S14 B shows the distribution of participants main outcome judgments. While participants discernment ability appears almost symmetrical, a few trustworthy outlets were, on average (including ca. 40 judgments per stimulus), rated as 'probably untrustworthy'. On the flip side, no untrustworthy website was, on average, rated as trustworthy. Considering the ratios of correct judgments between outcome measures, we see an almost perfect overlap between the claim veracity judgment (measured on a 4-point scale: definitely credible, probably credible, probably not credible, definitely credible) and source trustworthiness judgments (measured on a corresponding 4-point scale: definitely trustworthy, probably trustworthy, probably untrustworthy, definitely untrustworthy). Overall, participants' average ability to correctly assess the stimuli is at an already high level of above 80% that drops below 75% if people have to make a binary choice (true vs. false) including a 'don't know' option.
Additionally, using NewsGuard's ratings as benchmark, item response theory (IRT) indicators of item difficulty as well as item discrimination were computed for every stimulus article using bayesian 2PL-models in the brms() R-package [7] (see Fig. S14 left). In addition to varying item-difficulty parameters, 2PL-models allow each item to have its own discrimination, which is a more realistic assumption in our case compared to the more restrictive 1PL model. The 2PL model was formally selected over the 1PL model in a leave-one-out-cross-validation. We opted for a Bayesian instead of a frequentist model due to relatively sparse data (every individual rated only 8 out of 66 stimuli) and several instances of perfect correctness (everyone rated an item correctly) which led to extreme values in the item parameters in the frequentist 2PL models using the mirt() R-package [8]. The Bayesian 2PL model was specified as partially pooled model with the difficulty parameter forced to be positive and weakly informative priors (following the recommendations of Bürkner [7]).
Based on the results of this pretest, the existing set of 66 stimulus articles is condensed to a set of 12 stimuli for the main study using the following rules: We include two articles, one from a trustworthy and one from an untrustworthy outlet, for each political topic: climate change, Covid-19, the war in Ukraine, elections, migration and conspiracies around international organization. We exclude sources with very high familiarity, with very high rating correctness and with extreme political leaning to avoid ceiling effects for the main discernment task of the experiment. After

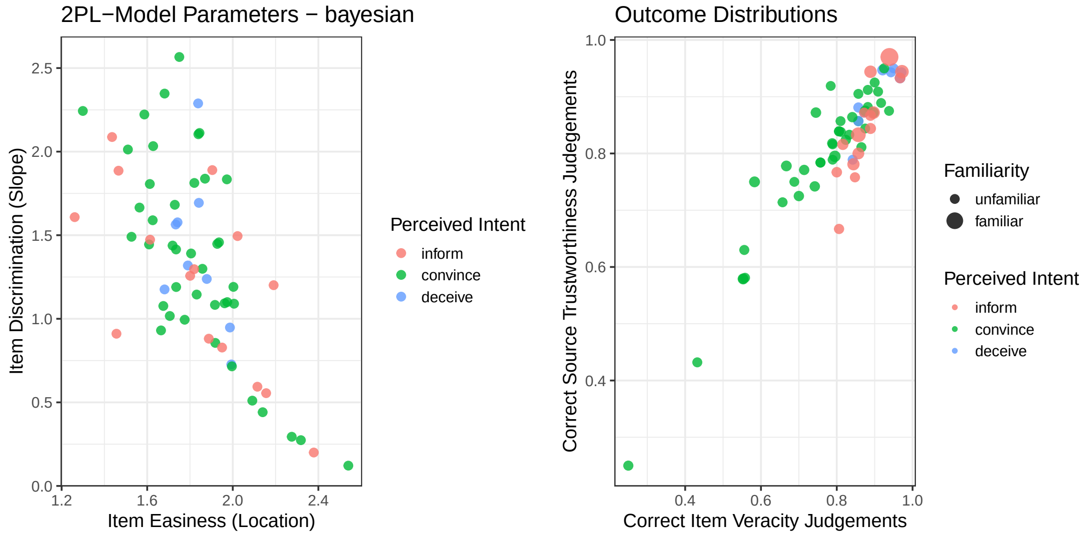


Figure S14: Outcome distributions in pretest. Legend: N = not trustworthy, T = trustworthy, NewsGuard rating. **Left:** secondary outcomes: familiarity, perceived intent and perceived political leaning of websites. **Right:** primary outcome variables: source trustworthiness and claim veracity, measured on two scales.

the described criteria are met, the stimuli with highest discrimination and highest difficulty are based on Bayesian IRT parameters. An example pair of websites can be found in the debriefing example. Data and code for the pretest can be found in the OSF repository of the project.

5 NewsGuard

NewsGuard is a database of credibility ratings for news and information websites, based on nine apolitical journalistic criteria that are rated by trained journalists (see `https://www.newsguardtech.com/` for details). NewsGuard operates in nine language spaces, covering news sources that, according to NewsGuard, account for at least 95% of online engagement in each language space respectively. NewsGuard Technologies Inc., the company behind the tool, also provides services such as misinformation tracking and brand safety for advertisers, search engines, social media platforms, cybersecurity firms, and government agencies (Wikipedia, 2024).
Profiles written by NewsGuard's analysts explain how they reached their conclusions for each website and provide links to specific fact-checked articles. Other researchers have used NewsGuard ratings and associated metadata, for example, to examine the use of untrusted websites during the 2020 US election [9].

References

1. Aslett, K. *et al.* Online searches to evaluate misinformation can increase its perceived veracity. en. *Nature.* Publisher: Nature Publishing Group, 1–9. ISSN: 1476-4687. https://www.nature.com/articles/s41586-023-06883-y (2024) (Dec. 2023).
2. Bates, D., Kliegl, R., Vasishth, S. & Baayen, H. *Parsimonious Mixed Models* arXiv:1506.04967. May 2018. http://arxiv.org/abs/1506.04967 (2024).
3. Bundesamt, S. *Bevölkerung: Deutschland, Stichtag, Altersjahre, Nationalität/Geschlecht/Familienstand* 2024. https://www-genesis.destatis.de/datenbank/online/table/12411-0006/search/s/MTI0MTEtMDAwNg== (2025).
4. Bundeswahlleiterin, D. *Bundestagswahl 2025* 2025. https://www.bundeswahlleiterin.de/bundestagswahlen/2025/ergebnisse/bund-99.html (2025).
5. Gelman, A. Scaling regression inputs by dividing by two standard deviations. *Statistics in medicine* **27,** 2865–2873 (2008).
6. Lin, H. *et al.* High level of correspondence across different news domain quality rating sets. *PNAS Nexus* **2,** pgad286. ISSN: 2752-6542. https://doi.org/10.1093/pnasnexus/pgad286 (2023) (Sept. 2023).
7. Bürkner, P.-C. Bayesian Item Response Modeling in R with brms and Stan. en. *Journal of Statistical Software* **100,** 1–54. ISSN: 1548-7660. https://doi.org/10.18637/jss.v100.i05 (2024) (Nov. 2021).
8. Chalmers, R. P. mirt: A Multidimensional Item Response Theory Package for the R Environment. en. *Journal of Statistical Software* **48,** 1–29. ISSN: 1548-7660. https://doi.org/10.18637/jss.v048.i06 (2024) (May 2012).
9. Moore, R. C., Dahlke, R. & Hancock, J. T. Exposure to untrustworthy websites in the 2020 US election. en. *Nature Human Behaviour* **7.** Publisher: Nature Publishing Group, 1096–1105. ISSN: 2397-3374. https://www.nature.com/articles/s41562-023-01564-2 (2024) (July 2023).