MPhil Politics, Comparative Government

Edward Anders

May 03, 2025

Contents

1 Case selection and data gathering

- See pre-analysis plan for details of what to include in this section
- Case selection (why focus on the UK?)
  - I am to have external validity to my research/case?
  - Can I make inferences to other cases?
- What is the case?
  - What is the unit of analysis?
  - What is the time period?
  - What is the geographical scope?
  - What are the key variables?
- What data is being collected?
- How is the data being collected?
  - What is the sampling strategy?
  - Note the UK weighting
- Plan for using agentic modelling
  - Why would I use agentic modelling?
  - What is the agentic modelling?
  - How will I use agentic modelling?

2 Data analysis‘

The survey experiment was facilitated by YouGov across a representative sample of `2,001` UK adults. The sample of participants was drawn from YouGov's panel and weighted to be representative of the UK population.

2.1 Thermometer Analysis

2.1.1 Moderator Heterogeneity Analysis

2.2 Outcome Variable Analysis

2.2.1 Moderator Heterogeneity Analysis

3 Appendix

3.1 Codebook

Variable	Type	Description	Values
identity_client	Identifier	Unique identifier for the respondent	Alphanumeric string
weight	Continuous	Survey weight to ensure national representativeness	Continuous float (e.g., 0.982, 1.034)
age	Continuous	Age of the respondent	Integer values, typically 18–90
profile_gender	Categorical	Gender of the respondent	Female; Male
profile_GOR	Categorical	Government Office Region (region of residence)	East Midlands; East of England; London; North East; North West; Scotland; South East; South West; Wales; West Midlands; Yorkshire and the Humber
voted_ge_2024	Categorical	Did the respondent vote in the 2024 General Election?	Don't know; No, did not vote; Yes, voted
pastvote_ge_2024	Categorical	How the respondent voted in the 2024 General Election	Conservative; Don't know; Green; Labour; Liberal Democrat; Other; Plaid Cymru; Reform UK; Scottish National Party (SNP); Skipped
pastvote_EURef	Categorical	How the respondent voted in the 2016 EU Referendum	Can't remember; I did not vote; I voted to Leave; I voted to Remain
education_recode	Categorical	Re-coded education level (grouped)	High; Medium; Low
profile_work_stat	Categorical	Employment status	Full time student; Not working; Other; Retired; Unemployed; Working full time (30+ hrs); Working part time (8–29 hrs); Working part time (<8 hrs)
political_attention	Continuous	How much attention the respondent pays to politics	Scale (e.g., 0–10 or continuous values)

(continued)

Variable	Type	Description	Values
`split`	Categorical	Randomly assigned treatment group (1–4)	1 = AI-generated, not labelled as AI-generated; 2 = AI-generated and labelled as AI-generated; 3 = Human-generated but labelled as AI-generated; 4 = Human-generated, not labelled as AI-generated
`xconsent`	Categorical	Consent to participate in the survey	I consent to taking part in this study; I do not wish to continue with this study
`mostlikely`	Categorical	Most likely party to receive vote	Conservative Party; Green Party; Labour Party; Liberal Democrats; Reform UK
`leastlikely`	Categorical	Least likely party to receive vote	Conservative Party; Green Party; Labour Party; Liberal Democrats; Reform UK; None of these; Not Asked
`MLthermo_KB`	Continuous	Thermometer rating for Kemi Badenoch (most likely party)	0–100
`MLthermo_KS`	Continuous	Thermometer rating for Keir Starmer	0–100
`MLthermo_NF`	Continuous	Thermometer rating for Nigel Farage	0–100
`MLthermo_ED`	Continuous	Thermometer rating for Ed Davey	0–100
`MLthermo_CD`	Continuous	Thermometer rating for Carla Denyer	0–100
`MLthermo_AR`	Continuous	Thermometer rating for Adrian Ramsay	0–100
`LLthermo_KB`	Continuous	Thermometer rating for Kemi Badenoch (least likely party)	0–100
`LLthermo_KS`	Continuous	Thermometer rating for Keir Starmer	0–100
`LLthermo_NF`	Continuous	Thermometer rating for Nigel Farage	0–100
`LLthermo_ED`	Continuous	Thermometer rating for Ed Davey	0–100

(continued)

Variable	Type	Description	Values
`LLthermo_CD`	Continuous	Thermometer rating for Carla Denyer	0–100
`LLthermo_AR`	Continuous	Thermometer rating for Adrian Ramsay	0–100
`agreedisagree`	Ordinal	Agreement with statement shown in the survey	Strongly disagree; Tend to disagree; Neither agree nor disagree; Tend to agree; Strongly agree
`xtrust`	Ordinal	Level of trust in the content shown	Almost never; Once in a while; About half of the time; Most of the time; Always
`child`	Ordinal	Respondent's emotional reaction to child-focused content	Extremely upset; Somewhat upset; Neither happy nor upset; Somewhat happy; Extremely happy
`MLthermoMean`	Continuous	Average thermometer score for most likely party	0–100 (row mean of MLthermo scores)
`LLthermoMean`	Continuous	Average thermometer score for least likely party	0–100 (row mean of LLthermo scores)
`ai_treatment`	Binary	Treatment status for AI-generated content	1 = Treated (shown AI-generated); 0 = Control (shown human-generated)
`label_treatment`	Binary	Treatment status for AI-labelled content	1 = Treated (labelled as AI-generated); 0 = Control (labelled as human-generated)

3.2 Data Cleaning

`2,001` respondents were provided with the survey experiment. Respondents who did not give consent to participate in the survey were removed. Respondents were given the option to skip questions. When skipped, a value of `997` was assigned to the question, which was then recoded to `NA`, as were `Not asked` values.

The survey was interested in understanding respondents' views towards their most and least preferred party. When asked who the `mostlikely` and `leastlikely` party was, respondents were given the option to select `None of these`. Respondents who selected `None of these` were removed from the sample as they were

unable to answer the follow-up questions.

Categorical variables were recoded to be `factors` in R, these were `profile_gender`, `profile_GOR`, `voted_ge_2024`, `pastvote_ge_2024`, `pastvote_EURef`, `profile_education_level`, `education_recode`, `profile_work_stat`, `xconsent`, `mostlikely`, `leastlikely`, `agreedisagree`, `xtrust`, and `child`.

Each of the thermometer variables were recoded to be `numeric` variables: `MLthermo_KB`, `MLthermo_KS`, `MLthermo_NF`, `MLthermo_ED`, `MLthermo_CD`, `MLthermo_AR`, `LLthermo_KB`, `LLthermo_KS`, `LLthermo_NF`, `LLthermo_ED`, `LLthermo_CD`, and `LLthermo_AR`. As the Green Party has two co-leaders, a mean thermometer score is calculated and used for most and least likely party thermometer scores, coded as `MLthermoMean` and `LLthermoMean`.

For treatment effect analysis, respondents were classified into two treatment groups: those shown AI-generated content (`ai_treatment`), identified where the split variable equalled `1` or `2`; and those shown AI-labelled content (`label_treatment`), identified where the split variable equalled `2` or `3`. Participants in the other split groups were coded as receiving human-generated or unlabelled content. These variables were coded as binary variables, where `1` indicated the treatment group and `0` indicated the control group.

3.3 Balance Check

To ensure that the randomisation process of the treatment allocation was successful, a balance check is conducted to ensure that the treatment and control groups are comparable in every way other than their treatment assignment status. The tables below report the balance of the covariates across the treatment groups. The continuous variables of `age` and `political_attention` are reported as means with the standard deviations in parentheses. The remaining categorical variables are reported as a count from the sample, with the proportions in parentheses. If there was a significant difference between the treatment and control groups, this is indicated with a `*` for $p < 0.05$, `**` for $p < 0.01$, and `***` for $p < 0.001$. The balance check shows that randomisation was successful across all covariates for both treatment groups as no covariates were significantly different between the treatment and control groups.

Table 2: Balance Table of Covariates by AI Treatment Group

Variable	Control	Treatment	p-value	Signif.
Age	52.12 (16.74)	51.56 (16.75)	0.521	-
Political attention	6.69 (1.92)	6.61 (1.98)	0.452	-
Gender (male)	374 (50.1)	412 (54.8)	0.080	-
Female	372 (49.9)	340 (45.2)	NA	
Education level (High)	308 (41.3)	308 (41.0)	0.882	-
Low	151 (20.2)	160 (21.3)	NA	
Medium	287 (38.5)	284 (37.8)	NA	
Employment status (Full time student)	31 (4.2)	35 (4.7)	0.759	-
Not working	32 (4.3)	37 (4.9)	NA	
Other	16 (2.1)	13 (1.7)	NA	
Retired	222 (29.8)	210 (27.9)	NA	
Unemployed	12 (1.6)	21 (2.8)	NA	
Working full time (30 or more hours per week)	327 (43.8)	338 (44.9)	NA	
Working part time (8-29 hours a week)	94 (12.6)	87 (11.6)	NA	
Working part time (Less than 8 hours a week)	12 (1.6)	11 (1.5)	NA	
Voted in 2024 General Election (Don't know)	3 (0.4)	1 (0.1)	0.574	-
No, did not vote	97 (13.0)	102 (13.6)	NA	
Yes, voted	646 (86.6)	649 (86.3)	NA	
Vote in 2024 General Election (Conservative)	162 (25.1)	143 (22.0)	0.587	-
Don't know	2 (0.3)	6 (0.9)	NA	
Green	58 (9.0)	51 (7.9)	NA	
Labour	211 (32.7)	245 (37.8)	NA	
Liberal Democrat	90 (13.9)	84 (12.9)	NA	
Other	13 (2.0)	12 (1.8)	NA	
Plaid Cymru	2 (0.3)	2 (0.3)	NA	
Reform UK	98 (15.2)	96 (14.8)	NA	
Scottish National Party (SNP)	9 (1.4)	10 (1.5)	NA	
Skipped	1 (0.2)	0 (0.0)	NA	
Vote in EU Referendum (Can't remember)	11 (1.5)	10 (1.3)	0.835	-
I did not vote	125 (16.8)	132 (17.6)	NA	
I voted to Leave	287 (38.5)	273 (36.3)	NA	
I voted to Remain	323 (43.3)	337 (44.8)	NA	
Region (East Midlands)	49 (6.6)	61 (8.1)	0.376	-
East of England	89 (11.9)	79 (10.5)	NA	
London	94 (12.6)	73 (9.7)	NA	
North East	34 (4.6)	26 (3.5)	NA	
North West	83 (11.1)	84 (11.2)	NA	
Scotland	44 (5.9)	64 (8.5)	NA	
South East	109 (14.6)	120 (16.0)	NA	
South West	79 (10.6)	70 (9.3)	NA	
Wales	31 (4.2)	35 (4.7)	NA	
West Midlands	62 (8.3)	66 (8.8)	NA	
Yorkshire and the Humber	72 (9.7)	74 (9.8)	NA	

Note: P-values are from t-tests (continuous) or chi-squared tests (categorical) comparing groups. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 3: Balance Table of Covariates by Label Treatment Group

Variable	Control	Treatment	p-value	Signif.
Age	51.84 (16.62)	51.84 (16.88)	0.996	-
Political attention	6.58 (1.94)	6.71 (1.96)	0.200	-
Gender (male)	408 (54.0)	378 (50.9)	0.240	-
Female	347 (46.0)	365 (49.1)	NA	
Education level (High)	321 (42.5)	295 (39.7)	0.542	-
Low	153 (20.3)	158 (21.3)	NA	
Medium	281 (37.2)	290 (39.0)	NA	
Employment status (Full time student)	31 (4.1)	35 (4.7)	0.966	-
Not working	37 (4.9)	32 (4.3)	NA	
Other	16 (2.1)	13 (1.7)	NA	
Retired	213 (28.2)	219 (29.5)	NA	
Unemployed	19 (2.5)	14 (1.9)	NA	
Working full time (30 or more hours per week)	338 (44.8)	327 (44.0)	NA	
Working part time (8-29 hours a week)	90 (11.9)	91 (12.2)	NA	
Working part time (Less than 8 hours a week)	11 (1.5)	12 (1.6)	NA	
Voted in 2024 General Election (Don't know)	2 (0.3)	2 (0.3)	0.154	-
No, did not vote	113 (15.0)	86 (11.6)	NA	
Yes, voted	640 (84.8)	655 (88.2)	NA	
Vote in 2024 General Election (Conservative)	148 (23.1)	157 (24.0)	0.927	-
Don't know	4 (0.6)	4 (0.6)	NA	
Green	55 (8.6)	54 (8.2)	NA	
Labour	233 (36.4)	223 (34.0)	NA	
Liberal Democrat	85 (13.3)	89 (13.6)	NA	
Other	10 (1.6)	15 (2.3)	NA	
Plaid Cymru	2 (0.3)	2 (0.3)	NA	
Reform UK	96 (15.0)	98 (15.0)	NA	
Scottish National Party (SNP)	7 (1.1)	12 (1.8)	NA	
Skipped	0 (0.0)	1 (0.2)	NA	
Vote in EU Referendum (Can't remember)	9 (1.2)	12 (1.6)	0.591	-
I did not vote	131 (17.4)	126 (17.0)	NA	
I voted to Leave	272 (36.0)	288 (38.8)	NA	
I voted to Remain	343 (45.4)	317 (42.7)	NA	
Region (East Midlands)	56 (7.4)	54 (7.3)	0.700	-
East of England	78 (10.3)	90 (12.1)	NA	
London	84 (11.1)	83 (11.2)	NA	
North East	32 (4.2)	28 (3.8)	NA	
North West	86 (11.4)	81 (10.9)	NA	
Scotland	57 (7.5)	51 (6.9)	NA	
South East	116 (15.4)	113 (15.2)	NA	
South West	80 (10.6)	69 (9.3)	NA	
Wales	28 (3.7)	38 (5.1)	NA	
West Midlands	72 (9.5)	56 (7.5)	NA	
Yorkshire and the Humber	66 (8.7)	80 (10.8)	NA	

Note: P-values are from t-tests (continuous) or chi-squared tests (categorical) comparing groups. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.