

Supplemental Materials

Table of Contents

A Close Replication of Experiment 1: The Neutral Condition of Experiment 2.....	2
Perceptions of Other's Preference.....	3
Results Experiment 1.....	3
Results Experiment 2.....	4
Results Experiment 3.....	5
Display Order and Content of Candidate Descriptions.....	7
Results Experiment 1.....	7
Candidate preference.....	7
Query order.....	7
Results Experiment 2.....	8
Candidate preference.....	8
Query order.....	8
Results Experiment 3.....	8
Additional Analyses.....	10
Experiment 1.....	10
Experiment 3.....	11
Incumbency effect.....	11
Query order.....	11
Query content.....	13
Power Analyses.....	15

A Close Replication of Experiment 1: The Neutral Condition of Experiment 2

The neutral aspect listing condition of Experiment 2 provided a close replication of Experiment 1. Thus, we conducted the same analyses as in Experiment 1 to verify our conclusions. The following analysis only include participants who were randomly assigned to the natural query order condition.

As in Experiment 1, the SMDR score was used as an indicator of query order: $2(MR_{McGinn} - MR_{Nickels})/n$. However, as this experiment included the response category “other”, the calculation of the SMDR score was slightly different. Specifically, if a participant listed an “other” aspect within their sequence (with the exception of at the end of the sequence) this artificially inflated the mean rank scores for pro-Nickels aspects and/or pro-McGinn aspects. The total number of aspects in the participants’ sequence included any “other” responses in the count. Consequently, the SMDR score no longer only ranged from -1 to 1. However, positive scores continued to indicate that the participants listed aspects in support of Nickels earlier in their query sequence than those in support of McGinn. The reverse is true for negative scores. Participants who only listed “other” aspects received an SMDR score of 0.

Consistent with the results of Experiment 1, an analysis of just the neutral condition revealed that the SMDR score was significantly higher in the Nickels incumbent condition ($M = 0.15, SD = 0.96$) than in the McGinn incumbent condition ($M = -0.26, SD = 0.94$), $t(154.77) = 2.73, p = .007, d = 0.44$. The SMDR score was also positively correlated with personal candidate preference, $r(167) = .56, p < .001$.

A closer look at participant’s query order (SMDR) revealed that the data did not fulfill the assumptions of normality. In fact, all but 19 participants had a SMDR of -1 or 1, making the data bimodal. Therefore, we decided to repeat our analyses with SMDR as a categorical variable. A Pearson’s chi-square test confirmed that query order scores were

significantly different in each of the incumbency conditions, χ^2 (df = 1) = 6.48, $p = .011$.

Participants who read that Nickels was the incumbent were more likely to list all reasons about Nickels first ($n = 51$) than all reasons about McGinn first ($n = 37$). The reverse was true when McGinn was described as the incumbent. Participants were significantly more likely to list all reasons about McGinn first ($n = 40$) than all reasons about Nickels first ($n = 22$).

Perceptions of Other's Preference

In all three experiments, we also estimated the participants' perceptions of other's preferences. Participants responded to the two questions asking, "Who do you think others would prefer to be elected?", and "Who do you think others would be most likely to vote for?" on a nine-point Likert scale. All questions were coded so that higher scores indicated a preference for Nickels, regardless if he was the incumbent or not, and that lower scores indicated a preference for McGinn, regardless if he was the incumbent or not. The average of were used to construct an other-preference score, (Experiment 1: $r[252] = .89$, $p < .001$; Experiment 2: $r[442] = .89$, $p < .001$; Experiment 3: $r[711] = .87$, $p < .001$).

For exploratory purposes, we also investigated whether participants inferred that others would also have a preference for the incumbent and how their query orders are related to this perception.

Results Experiment 1

Using an independent samples t-test, we found that participants thought others would be more likely to vote for the political incumbent (Nickels: $M = 6.27$, $SD = 1.84$; McGinn: $M = 3.85$, $SD = 1.51$), $t(247.14) = 11.56$, $p < .001$, $d = 1.47$. In fact, this effect proved to be larger than for their own candidate preferences ($d = 0.74$). This incumbency effect was also validated when looking at the difference from the midpoint of the scale. Participants perceived others to have a significant preference for the incumbent ($M = 6.21$, $SD = 1.68$), $t(225) = 11.56$, $p < .001$, $d = 1.45$. However, a correlation analysis revealed that the order in

which participants queried information from memory was correlated with their perceptions about others' preferences, $r(254) = .34, p < .001$, to a much lower extent than to with their own preferences, $r(254) = .66, p < .001$. To test whether the significant correlation between query order and others' preference, we conducted a partial correlation analysis, which controlled for participants' personal candidate preferences. The analysis revealed that only a weak negative correlation (i.e. the opposite direction of the original correlation) remained when controlling for personal candidate preference, $r(253) = -0.16, p = .011$, suggesting that query order primarily affects personal preferences.

Results Experiment 2

We conducted a two-way factorial ANOVA to test the effect of query order and incumbency on participants' perceptions of others¹. No significant interaction effect was found, $F(2, 440) = 1.77, p = .172$, nor was there a main effect of order, $F(2,440) = 0.88, p = .415$. There was, however, a significant main effect of incumbency, $F(1,440) = 157.94, p < .001, \eta^2 = .26$, indicating that participants thought others would always exhibit an incumbency advantage, regardless of the order in which they queried information from memory. This also suggests that query order primarily affects personal preferences.

¹ Levene's Test of homogeneity of variance (median centered) revealed that there was as significant violation of the homogeneity assumption, $F = 4.66, p < .001$.

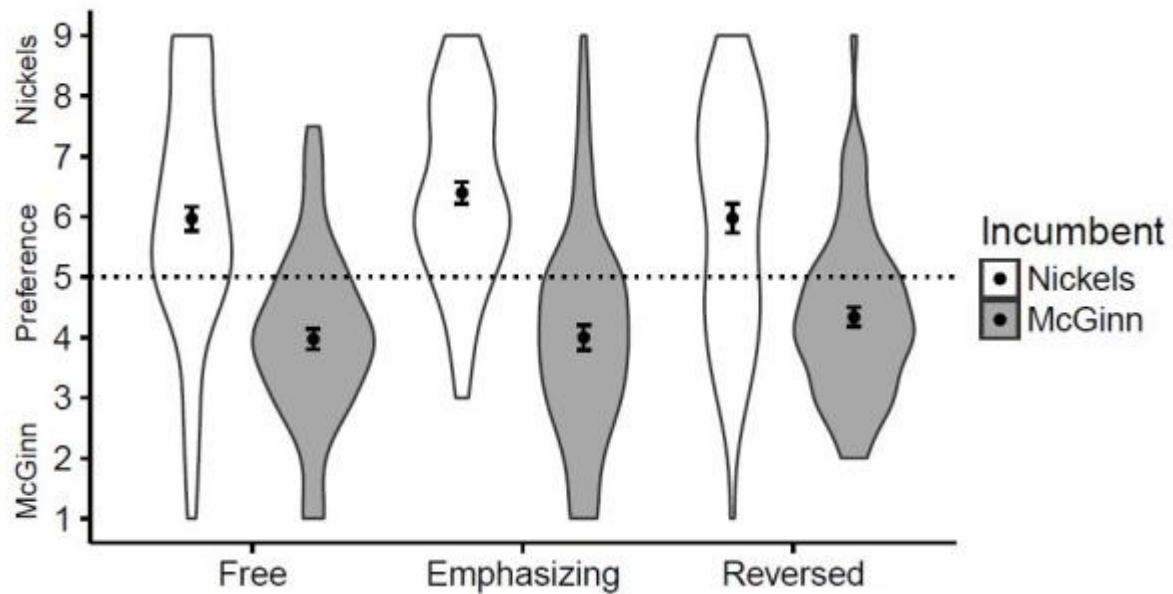


Figure S5. Experiment 2: Violin plots of others' candidate preferences. Error bars represent standard errors. The dotted line represents the neutral midpoint of the scale.

Results Experiment 3

In Experiment 3, we investigated the relationship between ideological compatibility participants' perceptions of others' candidate preferences. A two-way factorial ANOVA revealed significant interaction effect of ideological compatibility and incumbency, $F(2, 707) = 5.31, p = .005, \eta^2 = .01$. A simple slope analysis revealed that the pattern of results was different from that of personal preferences. In all three conditions of ideological compatibility, participants indicated that they thought others would be significantly more likely to prefer the incumbent than the challenger (neutral: $F(1, 707) = 64.05, p < .001, d = 0.99$; compatible: $F(1, 707) = 73.44, p < .001, d = 1.17$; incompatible: $F(1, 707) = 16.74, p < .001, d = 0.53$). In fact, the compatibility of political ideology seemed to have only had a minor effect on participants' perception of others' candidate preferences. Instead, they thought others would exhibit an incumbency advantage remained regardless of the compatibility condition.

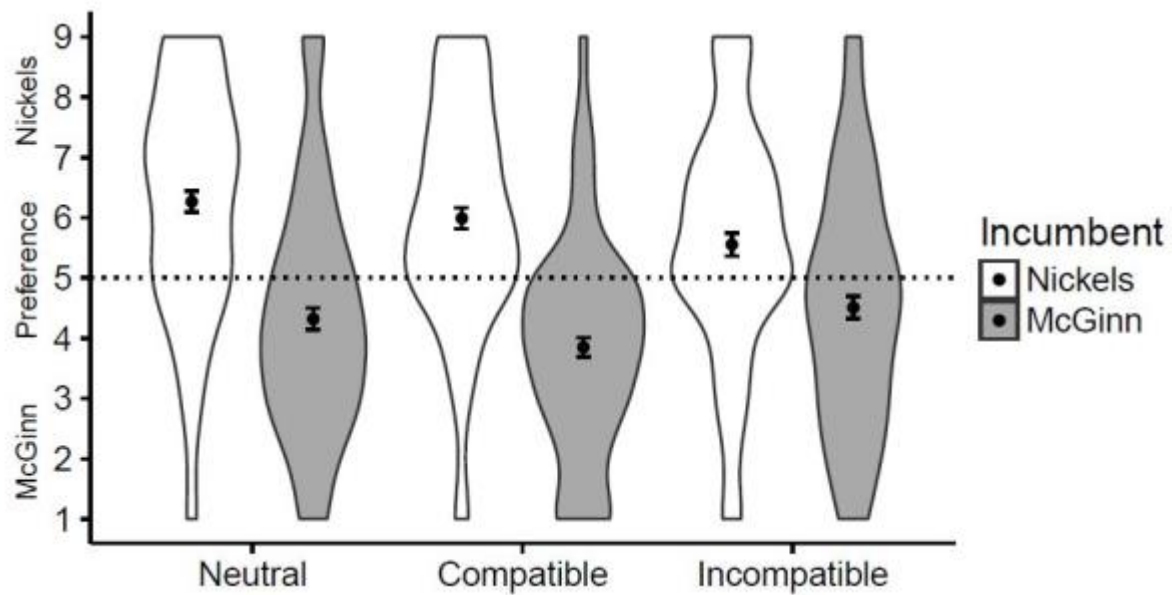


Figure S5. Experiment 2: Violin plots of others' candidate preferences. Error bars represent standard errors. The dotted line represents the neutral midpoint of the scale.

To test the relationship between query order and participants' impression about others, we tested the correlation between these two variables. We found they correlated, $r(711) = .47, p < .001$, to a lower extent than query order and personal candidate preference, $r(711) = .80, p < .001$. To test whether the significant correlation between query order and others' preference, we conducted a partial correlation analysis, which controlled for participants' personal candidate preferences. The analysis revealed that, when controlling for personal candidate preference, the correlation between query order and other's candidate preferences disappeared, $r(710) = -0.03, p = .352$.

From the results of all three studies, we can draw the conclusion that voters most likely rely on cognitive mechanism different from information retrieval order, when forming impressions about others. Additionally, we noted that the voters perceive different cues to have the most informative validity when forming preferences for themselves compared to forming impressions about others. In the third experiment, we find that participants seem to disregard the ideological compatibility cue, which is so important to their own preference formation, when thinking about what others prefer. Participants are likely aware of the

variability in the population in political ideology, thus they focus on the less variable label of incumbency. Addition, incumbency may hold some normative cues, as incumbency status indicates that others have voted for this candidate in the past, therefore voters assume to assume that others may do so again in the future.

Display Order and Content of Candidate Descriptions

In the candidate descriptions, we systematically varied both the display order and content of the candidate descriptions in order to rule out other variables that may produce a preference for the incumbent. To ensure that the systematic variation worked as planned, we tested whether the content or the order of the candidate descriptions significantly predicted query order.

Results Experiment 1

Candidate preference. We tested the effects of incumbency on candidate preferences, while controlling for display order and content of the candidate descriptions. A 2 (incumbency) \times 2 (display order) \times 2 (content) factorial ANOVA revealed that there were no significant interaction effects, $ps > .300$, nor was there main effects of display order or content, $ps > .100$. However, there was a significant main effect of incumbency, $F(1,248) = 34.33$, $p < .001$, $\eta^2 = 0.11$, thus confirming that only incumbency drove candidate preferences in our experiment.

Query order. We also tested the effects of incumbency on query order (SMDR), while controlling for display order and content of the candidate descriptions. A 2 (incumbency) \times 2 (display order) \times 2 (content) factorial ANOVA revealed that there were no significant interaction effects, $ps > .400$, nor was there main effects of content, $p = .536$. However, there was a significant main effect of incumbency, $F(1,248) = 7.75$, $p = .006$, $\eta^2 = 0.03$, as well as display order, $F(1,248) = 5.58$, $p = .019$, $\eta^2 = 0.02$. As the effect of

incumbency persists and display order does not predict candidate preferences we can assume that incumbency is driving our main results as predicted.

Results Experiment 2

Candidate preference. In order to control for display order and content of the candidate descriptions on candidate preferences, we included these of these variables to 2 (incumbent) \times 3 (query order) \times 2 (display order) \times 2 (content) factorial ANOVA. All four, three, and two-way interactions were non-significant, $ps > .09$. The two exceptions were the predicted two-way interaction of incumbency and query order, $F(2, 422) = 7.48, p < .001, \eta^2 = 0.03$, and an interaction between display order and content, $F(1, 422) = 10.17, p = .001, \eta^2 < 0.01$.

Query order. In the neutral condition, which is a close replication of Experiment 1, we also assessed whether display order or content of the candidate descriptions predicted participants' query orders (SMDR). A 2 (incumbency) \times 2 (display order) \times 2 (content) factorial ANOVA revealed that there was no three-way interaction between these variables, $p = .651$. The two-way interactions were not significant, $p > .400$, with the exception of a marginally significant interaction between display order and content, $F(1,161) = 3.90, p = .050, \eta^2 = 0.02$. More importantly the predicted main effect of incumbency on query order persisted, $F(1,161) = 7.89, p = .006, \eta^2 = 0.04$. There was also a significant main effect of display order on query order, $F(1,161) = 11.24, p = .001, \eta^2 = 0.06$. As the effect of incumbency persists and display order does not predict candidate preferences we can assume that incumbency is driving our main results as predicted. Description content was not statistically significant, $p = .435$.

Results Experiment 3

In order to control for display order and content of the candidate descriptions on candidate preferences and query, we included these of these variables to 2 (incumbent) \times 3

(query order) \times 2 (display order) \times 2 (content) factorial ANOVAs. For both candidate preference and query order, we found that the predicted interaction effect of incumbency and ideological compatibility persisted and was revealed to have the strongest effect size (Table S1).

Although the predicted interaction of incumbency and ideological compatibility remained consistent across conditions, we took a closer look at what was driving the significant four-way interaction on SMDR. In the condition in which McGinn (from Long Island) was described as the incumbent but Nickels (from Seattle) was displayed first, participants were significantly more likely to first query information supporting the challenger Nickels. This is contrary to our prediction that information supporting the incumbent will be queried from memory earlier. However, we are cautious to draw conclusions about this abnormality due to the small effect size of the interaction and low number of participants in each individual cell after taking order and content into account (e.g., this cell $n = 28$). All other cells showed a pattern in line with our original conclusion.

Table S1

Summary of factorial ANOVAs testing the effect of incumbency and ideological compatibility on candidate preference and query order (SMDR) while controlling for display order and content of the candidate descriptions

	<i>df</i>	Candidate Preference			SMDR		
		<i>F</i>	<i>p</i>	η^2	<i>F</i>	<i>p</i>	η^2
Incumbent (A)	1	54.51	< .001	0.06	25.33	< .001	0.03
Compatibility (B)	2	2.57	.077	0.01	1.221	.296	< 0.01
Display Order (C)	1	3.07	.080	< 0.01	17.66	< .001	0.02
Content (D)	1	0.03	.863	< 0.01	0.89	.345	< 0.01
A \times B	2	70.86	< .001	0.15	59.93	< .001	0.13

A × C	1	7.36	.007	0.01	7.31	.007	0.01
B × C	2	0.25	.781	< 0.01	0.28	.753	< 0.01
A × D	1	0.05	.830	< 0.01	0.38	.537	< 0.01
B × D	2	0.01	.992	< 0.01	0.13	.882	< 0.01
C × D	1	23.421	< .001	0.02	12.29	< .001	0.01
A × B × C	2	0.094	.910	< 0.01	0.45	.636	< 0.01
A × B × D	2	0.53	.588	< 0.01	0.42	.655	< 0.01
A × C × D	1	1.97	.161	< 0.01	0.24	.623	< 0.01
B × C × D	2	7.02	.001	0.01	5.91	.003	0.01
A × B × C × D	2	2.69	.069	< 0.01	3.53	.030	0.01

Note: The bolded row highlights the predicted interaction of incumbency and ideological compatibility on candidate preference and query order (SMDR).

Additional Analyses

Experiment 1

A closer look at participant's query order (SMDR) revealed that the data did not fulfill the assumptions of normality. In fact, all but 18 participants had a SMDR of -1 or 1, making the data extremely bimodal. Therefore, we decided to repeat our analyses with SMDR as a categorical variable. After eliminating the 18 participants who could not be easily dichotomized, we ran a Pearson's chi-square test. The results confirmed that query order scores were significantly different in each of the incumbency conditions, χ^2 (df = 1) = 6.73, p = .009. Participants who read that Nickels was the incumbent were more likely to list all reasons about Nickels first ($n = 72$) than all reasons about McGinn first ($n = 47$). The reverse was true when McGinn was described as the incumbent. Participants were significantly more likely to list all reasons about McGinn first ($n = 68$) than all reasons about Nickels first ($n = 51$).

Experiment 3

Incumbency effect. When running our confirmatory analyses we found that there was a significant violation of the homogeneity assumption, $F = 3.55, p = .004$. Hartley's $F_{\max} = 1.54$ (ratio of variances) suggests that the violation of the homogeneity test is not serious (critical $F_{\max} > 2.00$). Nonetheless, we decided to run an additional analysis to confirm our conclusions.

Unfortunately, there is no standard non-parametric test for a 2×3 between-subjects design. Thus, we decided to run a one-way ANOVA with six conditions using Welch's F which is a F -test that is more robust to violations of homogeneity. If these results support our original analyses, we will be able to more confidently make conclusions about the effect of incumbency and ideological compatibility on candidate preferences.

Our one-way ANOVA confirmed, that there was indeed a significant difference between the means of all six conditions, Welch's $F(5, 327.99) = 42.443, p < .001$. To investigate the relationship between the six conditions, we ran a Games-Howell post-hoc test. The post hoc analysis revealed that there is indeed a significant incumbency effect in the neutral condition, $t(247) = 4.2, p < .001, d = 0.53$, but also in the compatible condition, $t(235) = 13.2, p < .001, d = 1.72$. Finally, participants significantly prefer the challenger over the incumbent in the incompatible condition, $t(222) = 4.0, p < .001, d = 0.54$. These findings support the conclusions of our original analysis.

Query order. There was also a significant violation of the homogeneity assumption and the normality assumption for participant's query order scores across conditions. As the majority of the participants ($n = 682$) had query order scores of -1 or 1 (recodes as 0 and 1), we decided to run a binary logistic regression to investigate the effects of incumbency (0 = Nickels, 1 = McGinn) and ideological compatibility on query order. The neutral ideological compatibility condition is the reference category for all results reported below.

The binary logistic regression revealed that there is indeed a significant and substantial interaction effect between incumbency and ideological compatibility (Table S2 & Figure S5), thus confirming our original analysis. A closer look at the results, show that in the neutral and compatible conditions, it is significantly more probable that participants will first list reasons in support of the incumbent. In the incompatible condition, however, participants show a higher probability of preferring the challenger who has a similar political ideology (Figure S5).

Table S2

Summary of Binomial Multiple Logistic Regression for the Effects of Incumbency and Ideological Compatibility on Query Order (SMDR)

	<i>B (SE)</i>	<i>OR</i>	<u>95% CI for <i>OR</i></u>		<i>z</i>	<i>p</i>
			Lower	Upper		
Intercept	0.77 (0.19)	2.15	1.49	3.16	4.00	<.001
<i>Incumbency</i>						
McGinn	-1.06 (0.27)	0.35	0.20	0.58	-3.96	<.001
<i>Ideological Compatibility</i>						
Compatible	0.50 (0.30)	1.66	0.93	2.99	1.70	.089
Incompatible	-1.52 (0.29)	0.22	0.12	0.38	-5.325	<.001
<i>Interactions</i>						
McGinn x Compatible	-1.44 (0.41)	0.24	0.10	0.53	-3.48	.001
McGinn x Incompatible	2.49 (0.39)	12.07	5.62	26.31	6.33	<.001

Note: $R^2 = .16$ (Cox & Snell); $R^2 = .21$ (Nagelkerke)

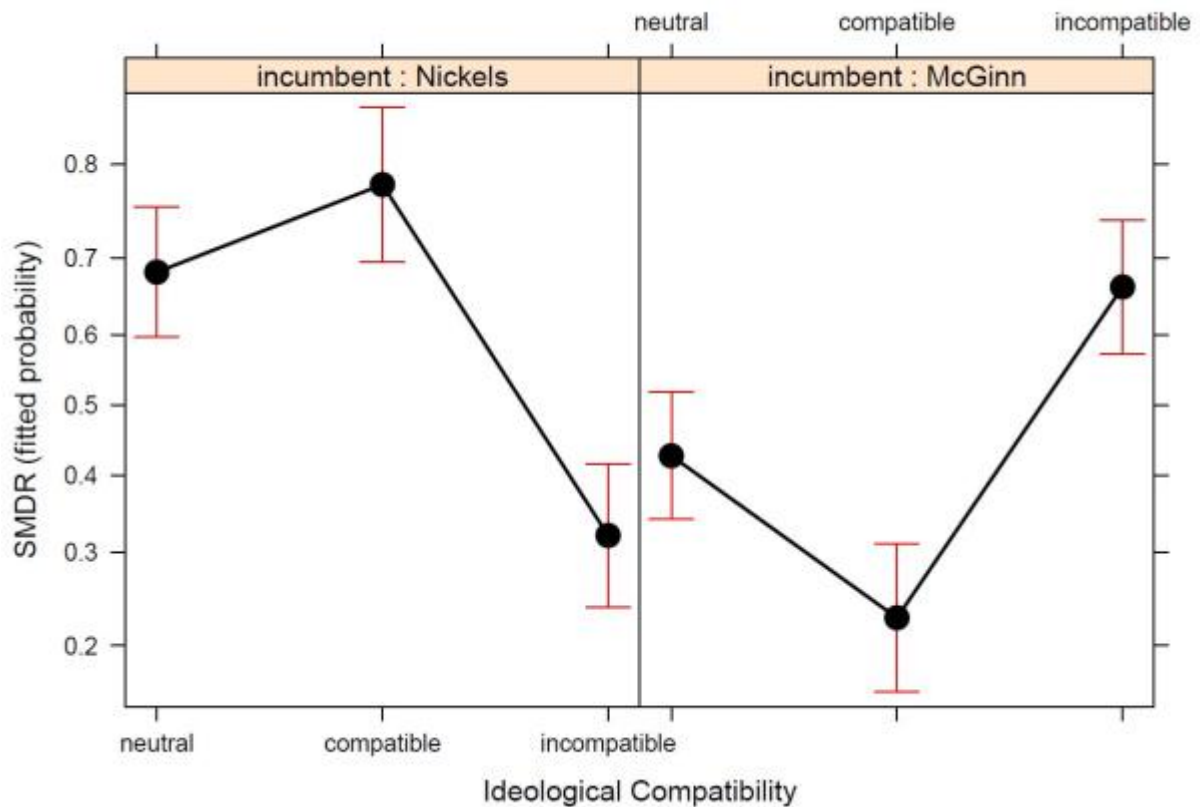


Figure S5. Experiment 3: Effect of incumbency and ideological compatibility on the fitted probability of query order (dichotomized measure). Higher scores indicate a higher probability of listing all supporting Nickels before those supporting McGinn.

Query content. There was also a significant violation of homogeneity for participant's query content scores. As the majority of the participants ($n = 575$) had query content scores of -1 or 1 (recodes as 0 and 1), we decided to run a binary logistic regression to investigate the effects of incumbency (0 = Nickels, 1 = McGinn) and ideological compatibility on query order. The neutral ideological compatibility condition is the reference category for all results reported below.

The binary logistic regression revealed that there is indeed a significant and substantial interaction effect between incumbency and ideological compatibility (Table S3 & Figure S6), thus confirming our original analysis. A closer look at the results, show that in the neutral and compatible conditions, it is significantly more probable that participants will first

list reasons in support of the incumbent. In the incompatible condition, however, participants show a higher probability of preferring the challenger who has a similar political ideology (Figure S6).

Table S3

Summary of Binomial Multiple Logistic Regression for the Effects of Incumbency and Ideological Compatibility on Query Content

	<i>B (SE)</i>	<i>OR</i>	<u>95% CI for <i>OR</i></u>		<i>z</i>	<i>p</i>
			Lower	Upper		
Intercept	0.75 (0.21)	2.11	1.43	3.20	3.650	<.001
<i>Incumbency</i>						
McGinn	-0.95 (0.29)	0.39	0.22	0.69	-3.221	.001
<i>Ideological Compatibility</i>						
Compatible	0.73 (0.33)	2.08	1.09	4.05	2.20	.028
Incompatible	-1.57 (0.32)	0.21	0.11	0.38	-4.98	<.001
<i>Interactions</i>						
McGinn x Compatible	-1.99 (0.47)	0.14	0.05	0.34	-4.26	<.001
McGinn x Incompatible	2.45 (0.44)	11.57	4.96	27.55	5.60	<.001

Note: $R^2 = .19$ (Cox & Snell); $R^2 = .25$ (Nagelkerke)

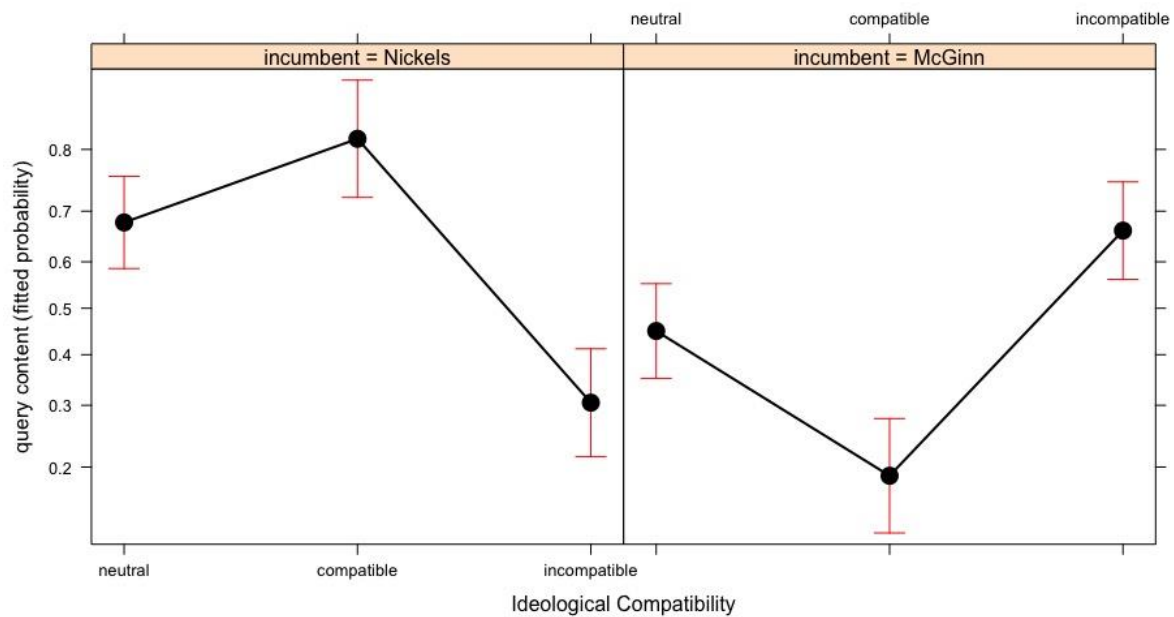


Figure S6. Experiment 3: Effect of incumbency and ideological compatibility on the fitted probability of query content (dichotomized measure). Higher scores indicate a higher probability of listing more reasons supporting Nickels than reasons supporting McGinn.

Power Analyses

In our first experiment, we were interested in the differences between our two experimental conditions of incumbency. However, we also varied display order and content of the candidate descriptions in order to control for these variables. As such, our experimental design included eight cells. We aimed to have 35 to 40 participants per cell (including those of less of interest to our questions), leading us to open a HIT for 300 MTurk workers.

For Experiments 2 and 3, we determined the necessary sample size with a power analysis conducted with the software GPower. We based our power analysis on effect size of query order on the incumbency advantage observed in Experiment 1. We transformed the observed $d = 0.35$ into the equivalent $f = 0.175$. In order to find the predicted 2×3 interactions effect, with 95% power, a sample size of 508 was required. Thus, for Experiment 1 we decided to round up this number to a stopping rule of 100 participants per cell and open a HIT for 600 MTurk workers.

MTurk samples are usually skewed towards the political liberalism. In order to ensure that we would have enough conservative participants in Experiment 3, which tested the effects of ideological compatibility, we decided to increase our sample size. Thus, opted for a stopping rule of about 130 participants per cell and opened a Hit for 600MTurk workers.