Supplemental Online Materials (SOM) for

The Head, Heart, and Soul: Lay Theories of Decision Conflict and the Role of the True Self

Table of Contents

Note. This document contains supplemental materials for all studies in our paper, and reports studies not presented in the main text for reasons outlined in each study's write-up below. In this document, we focus on key additional analyses not in the main text, but note that additional study materials and data are available via OSF (https://osf.io/aw439/?view_only=6e56a66f3d6a464898f5a14b6431bb52). Please also see the OSF page for a living version of this document.

Study 1

Controlling for option counterbalancing. As noted in footnote 3 in our main manuscript, we did observe a small but significant interaction between assignment and option counterbalancing, $F(1, 1183) = 4.07$, $p = .044$, $\eta^2 = .003$, such that intuition ($M = 3.56$, $SE = 0.09$), as opposed to deliberation ($M = 3.13$, $SE = 0.09$), was perceived to be more indicative of a target's authentic preference when deliberation was ordered first, $t(1183) = 3.59$, $p < .001$, $d = 0.29$, 95% CI [0.13, 0.45], but not when intuition was ordered first ($M_{Intuition} = 3.31$, $SE_{Intuition} = 0.09$ vs. $M_{Deliberation} = 3.22$, $SE_{Deliberation} = 0.09$), $t(1183) = 0.74$, $p = .462$, $d = 0.06$, 95 % CI [-0.10, 0.22]. Importantly, however, our main results were nearly identical to those reported in the main text. That is, supporting H1, we observed a main effect of assignment (Deliberation Favors Action vs. Intuition Favors Action), $F(1, 1183) = 9.42$, $p = .002$, with a small effect size, $\eta^2 = .010$, such that participants were more likely to say that the target authentically preferred to engage in the behavior when engaging in the behavior was favored by one's "gut" ($M = 3.44$, $SE = 0.06$) versus one's "head" ($M = 3.18$, $SE = 0.06$), $d = 0.18$, 95% CI [0.06, 0.29]. Supporting H2, we also observed a significant main effect of domain (Normative vs. Neutral), $F(1, 1183) = 249.25$, $p < .001$, here with a large effect size, $\eta^2 = .17$, such that participants reported that the target authentically preferred to engage in the behavior significantly more so when the behavior was neutral ($M = 3.98$, $SE = 0.06$), than when the behavior was normatively "bad" ($M = 2.63$, $SE = 0.06$), $d = 0.91$, 95% CI [0.79, 1.03].

We also observed a significant main effect of target, $F(1, 1183) = 10.27$, $p = .001$, $\eta^2 = .01$, which was qualified by a significant domain X target interaction, $F(1, 1183) = 17.56$, $p < .001$, $\eta^2 = .01$. An analysis of the simple effects indicated that although participants were significantly less likely to believe that others authentically preferred to engage in the behavior

when the behavior was normatively "bad" ($M_{Normative}$ = 2.95, $SE_{Normative}$ = 0.09 vs. $M_{Neutral}$ = 3.94, $SE_{Neutral}$ = 0.09), $t(1183) = -8.21$, $p < .001$, $d = -0.67$, 95 % CI [-0.83, -0.51], this effect was stronger when the target was oneself ($M_{Normative}$ = 2.31, $SE_{Normative}$ = 0.09 vs. $M_{Neutral}$ = 4.03, $SE_{Neutral}$ = 0.09), $t(1183) = -14.12$, $p < .001$, $d = -1.15$, 95% CI [-1.32, -0.99]. No other main effects or interactions were observed (see "truepreferences_study1_supplementary_additional.R").

Justifications (Chi-Square Tests of Independence). As originally stated in our preregistration (https://aspredicted.org/3RQ_FB6), we also conducted a series of chi-square tests of independence to compare choice frequency of the "source" of the target's authentic preference. Partially supporting H1, across five of the eight scenarios (coffee, buy sandwich, drug, steal sandwich, and child labor), participants were more likely to select the "gut" as the source of one's authentic preferences as opposed to one's "head" (see Table S1A). Supporting H2, in the normative scenarios we see a crossover pattern that is consistent with the idea that when doing the "bad" thing was favored by intuition (deliberation), participants were more likely to indicate the head (gut) as the source of one's authentic preferences (see Table S1B). Interestingly, despite the normative valence of the domain shifting participants' perceptions of the source of one's authentic preferences, participants were more likely to indicate that authentic preferences were revealed by the gut or head as opposed to norms in the drug, stealing, and harmful product (but not child labor) scenarios (see Table S1C).

Table S1A

Participants' Justifications by Scenario

Source	Proportion	χ^2	*p*
Coffee (N = 571)			
Gut	59.85%	20.48	< .001
Head	40.15%		
Buy Sandwich (N = 546)			
Gut	57.49%	10.94	< .001
Head	42.50%		
Recommend Product (N = 544)			
Gut	52.11%	0.76	.383
Head	47.89%		
Change Supplier (N = 527)			
Gut	48.18%	0.58	.446
Head	51.82%		
Drug (N = 570)			
Gut	57.99%	11.66	< .001
Head	42.01%		
Steal Sandwich (N = 562)			
Gut	55.52%	3.98	.046
Head	44.48%		
Recommend Harmful Product (N = 547)			
Gut	53.85%	19.15	< 001
Head	6.15%		
Change to Unethical Supplier (N = 538)			
Gut	57.34%	19.15	< .001
Head	42.65%		

Table S1B

Participants' Justifications by Assignment for Normative Scenarios

Source	Deliberation Favors Action	Intuition Favors Action	χ^2	*p*
Drug (N = 570)				
Gut	58.01%	35.29%	41.75	< .001
Head	21.35%	45.67%		
Norms	16.37%	16.26%		
Other	4.27%	2.77%		
Steal Sandwich (N = 562)				
Gut	39.64%	24.82%	39.01	< .001
Head	14.64%	36.88%		
Norms	41.43%	35.82%		
Other	4.29%	2.48%		
Recommend Harmful Product (N = 547)				
Gut	37.97%	26.33%	26.39	< .001
Head	18.04%	36.30%		
Norms	40.60%	36.29%		
Other	3.38%	1.07%		
Change to Unethical Supplier (N = 538)				
Gut	37.40%	23.91%	19.15	< .001
Head	15.65%	29.35%		
Norms	45.80%	45.65%		
Other	1.14%	1.09%		

Table S1C

Dual Process vs. Normative Justifications for Authentic Preferences

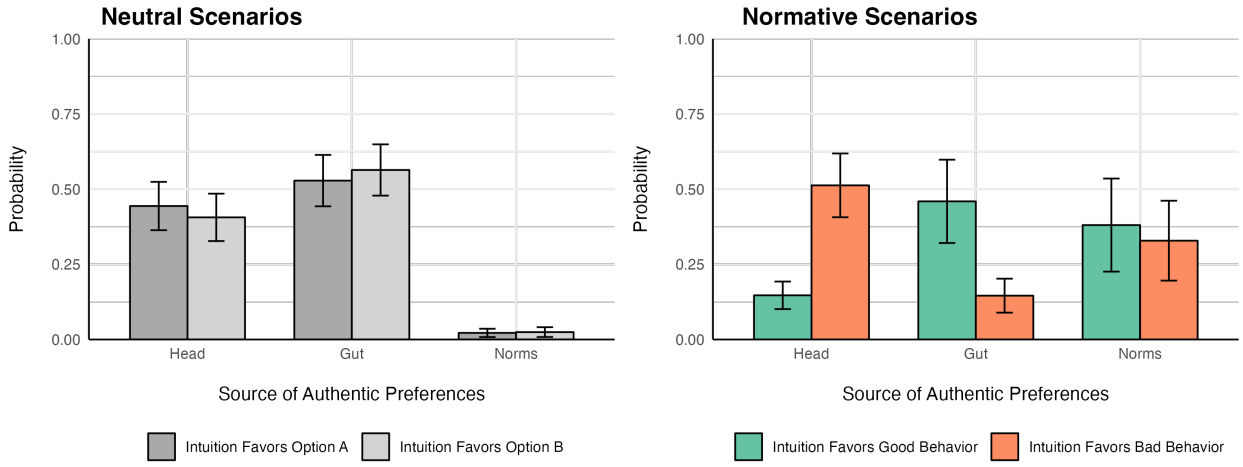
Source	Proportion	χ^2	*p*
Drug (N = 570)			
Dual-Process	67.80%%	85.46	< .001
Normative	32.20%%		
Steal Sandwich (N = 562)			
Dual-Process	60.04%	21.88	< .001
Normative	39.96%		
Recommend Harmful Product (N = 547)			
Dual-Process	60.75%	24.72	< .001
Normative	39.25%		
Change to Unethical Supplier (N = 538)			
Dual-Process	53.76%	3.01	.083
Normative	46.24%		

Authentic preferences (multilevel model). To generalize our findings to non-sampled stimuli (Yarkoni, 2022), we also replicated our predictions of the target's authentic preferences with a 2 (Domain: Normative vs. Neutral) x 2 (Target: Self vs. Other) x 2 (Assignment: Deliberation Favors Action vs. Intuition Favors Action) mixed OLS regression model, with participants (N = 1199) and scenarios (N = 8) treated as random effects. Results were nearly identical to our main results. Supporting H1, we observed a significant main effect of assignment, $\chi^2(1) = 9.33$, $p = .002$, such that participants were more likely to believe that the target authentically preferred to engage in the behavior when the behavior was favored by intuition ($M = 3.44$, $SE = 0.13$) as opposed to deliberation ($M = 3.17$, $SE = 0.13$), $d = 0.20$, 95% CI [0.05, 0.34]. Supporting H2, we also observed a significant main effect of domain, $\chi^2(1) = 29.84$, $p < .001$, whereby participants reported that the target authentically preferred to engage in the behavior significantly more so when the behavior was neutral ($M = 3.98$, $SE = 0.17$) than when the behavior was normatively "bad" ($M = 2.63$, $SE = 0.17$), $d = 1.03$, 95% CI [0.59, 1.47]. We also observed a significant main effect of target, $\chi^2(1) = 10.33$, $p = .001$, which was qualified by a significant domain X target interaction, $\chi^2(1) = 17.48$, $p < .001$. An analysis of the simple effects indicated that although participants were significantly less likely to believe that others authentically preferred to engage in the behavior when the behavior was "normatively bad" ($M_{Normative} = 2.95$, $SE_{Normative} = 0.18$ vs. $M_{Neutral} = 3.94$, $SE_{Neutral} = 0.18$), $t(9.20) = -3.80$, $p = .004$, $d = -0.75$, 95 % CI [-1.20, -0.31], this effect was stronger when the target was oneself ($M_{Normative} = 2.31$, $SE_{Normative} = 0.18$ vs. $M_{Neutral} = 4.02$, $SE_{Neutral} = 0.19$), $t(9.22) = -6.53$, $p < .001$, $d = -1.30$, 95% CI [-1.75, -0.85]. No other main effects or interactions were observed (see "truepreferences_study1_supplementary_multilevel.R").

Justifications (multilevel model). For this analysis, we omitted observations that indicated that the target did not have an authentic preference in each of the scenarios (i.e., by selecting "4 - Neither/no true preference") and for those who selected "Other" for the source of the target's authentic preference, leaving N = 4238 observations. We then conducted a 2 (Decision Type: Normative vs. Neutral) x 2 (Assignment: Deliberation Favors Action vs. Intuition Favors Action) mixed multinomial regression analysis predicting the source of one's authentic preference (i.e., the probability of selecting the gut, head, or norms), with participants (N = 922) and Scenarios (N = 8) treated as random effects. As shown on the left hand side of Figure S1A, for the neutral scenarios we observed that most participants believed that the authentic preferences were revealed by the gut (P = 54.5%, SE = 8.5%) as opposed to the head (P = 42.5%, SE = 7.9%), although this difference was not statistically significant, Z = 1.03, OR = 1.13, 95% CI [0.90, 1.42], p = .300. In contrast, as shown on the right hand side of Figure S1A, in the normative scenarios we see a crossover pattern that is consistent with the idea that when doing the "bad" thing was favored by intuition, a significantly greater proportion of participants thought that one's authentic preference was revealed by the head (P = 51.2%, SE = 10.6%) as opposed to the gut (P = 14.6%, SE = 5.6%), Z = 3.05, OR = 1.44, 95% CI [1.14, 1.83], p = .002. However, when doing the "bad" thing was favored by deliberation, participants were significantly more likely to report that one's authentic preference was revealed by the gut (P = 45.9%, SE = 13.9%) as opposed to the head (P = 14.7%, SE = 4.6%), Z = 2.14, OR = 1.37, 95% CI [1.03, 1.82], p = .032. Interestingly, in these normative cases only a third of participants (P = 35.4%, SE = 14.4%) cited "because it is the right thing to do" (normative justification) as the reason for their decision (see "truepreferences_study1_supplementary_multilevel.R").

Figure S1A

Source of the Target's Authentic Preference for Neutral and Normative Scenarios (Study 1)



Note. Error bars represent the standard error of the predicted probabilities.

Supplemental Study S1A

The main goal of Supplemental Study S1A was to replicate our findings for people's justifications without first having participants make decisions about a target's authentic preference. One limitation of how we measured justifications in the main text is that these items came after participants had already made their decisions about a target's authentic preferences, and thus, these measures may capture post hoc justifications rather than truly capturing the lay theories that participants hold in general. Thus, we conducted Study S1A to assess lay theories about authentic preferences in normative domains without prior prompting. As in Study 1, we included a target condition (self vs. other) to examine whether the results would hold across these target types. Our preregistered design and analysis plan are available here: https://aspredicted.org/BSX_THR.

Participants. We posted for 200 North American participants collected on Prolific Academic, and 199 provided complete data. As outlined in our preregistration, we did not exclude any participants.

Procedure. Participants were randomly assigned to a self versus other condition in a between-subjects design. Participants in the self condition were asked to "imagine you are contemplating what to do in a decision that involves morals or issues of "right" and "wrong" (e.g., whether or not to take an illegal drug, whether or not to steal, or whether or not to cheat on your spouse)." Participants in the other condition were given the same prompt, but instead asked to imagine "someone" contemplating. They were then asked, "What would indicate your [the person's] true preference in the decision? That is, the option you [they] truly prefer deep down?" The options were: "In general, a person's true preference is revealed by what their automatic, "gut" instinct is telling them to do"; "In general, a person's true preference is best revealed by what their careful, deliberate thoughts are telling them to do"; "In general, a person's true preference is best revealed by whatever most people would agree is "good" or "right"; and "Other (please specify)."

Results and Discussion. First, a chi-square goodness of fit test revealed that participants were significantly more likely to select the dual process options (intuition and deliberation coded as 1) versus the normative option (coded as 0), $\chi^2(N = 197) = 144.98$, $p < .001$. In terms of the specific options, 65.3% of participants selected intuition, 28.6% selected deliberation, and 6.1% selected the normative option. Two participants selected 'other.' Examining their qualitative responses, one participant indicated that there are no authentic preferences, and another indicated that authentic preferences must be figured out from developing one's "self-awareness." A chi-square test revealed that there was no effect of condition on option selection, $\chi^2(1, N = 197) =$

0.29, $p = .593$, Cramer's $V = .04$ (see Table S1C for results).

Table S1C

Percentage of Participants Selecting Each Option by Condition

Condition	Intuition	Deliberation	Normative	Other
Self	52.5%	36.4%	6.1%	1.0%
Other	55.0%	36.0%	8.0%	1.0%

The results of Supplemental Study S1A again suggest that participants tend to not have normative lay theories (i.e., that the normative valence of an action determines whether it is a n preference), but rather ground their beliefs in cognitive theories (usually that authentic preferences are revealed by intuition).

Supplemental Study S1B

We note that, for the normative item, we chose the phrasing "whatever most people would agree is "good" or "right"" in order to best match our paradigm whereby observers' own moral judgments are shaping beliefs about authentic preferences. However, to ensure that the results are not driven by invoking "most people," we also ran an identical version of this study ($N = 200$ Prolific participants, same paradigm) in which the normative item was phrased "In general, a person's authentic preference is best revealed by whatever is "good" or "right." The results replicated with this wording tweak (see Table S1D). A chi-square goodness of fit test revealed that participants were significantly more likely to select the dual process options (intuition and deliberation coded as 1) versus the normative option (coded as 0), $\chi^2(N = 195) = 113.85$, $p < .001$. In terms of the specific options, 56.0% of participants selected intuition, 30.0%

selected deliberation, and 11.5% selected the normative option. Five participants selected 'other.' The results were again not affected by condition (p = .595).

Table S1D

Percentage of Participants Selecting Each Option by Condition (Wording Tweak)

Condition	Intuition	Deliberation	Normative	Other
Self	56.6%	31.3%	11.1%	4.0%
Other	55.4%	28.7%	11.9%	1.0%

Study 2

Pre-test

For Study 2, we selected scenarios that met four criteria: (1) normative scenarios in which preferences were correlated with political orientation, (2) neutral scenarios in which preferences were bipartisan but exhibited bimodal distributions, (3) scenarios rated at or above the midpoint on a 7-point scale of preference strength (1 = *Not at all strong*, 7 = *Very strong*, anchored at 4 = *Somewhat strong*), and (4) normative scenarios that were rated as significantly more moral than neutral scenarios. To do so, we randomly assigned participants to evaluate four out of twelve scenarios (six normative and six neutral scenarios) and asked them to indicate their preference on the issue, the strength of their preference, and to what extent the scenario was moral in nature.

Method

Participants

We recruited 300 Prolific participants (58.5% female, 2.0% other/non-binary; M_{age} = 37.85 *SD* = 11.60), with the goal of having 100 participants rate each scenario.

Procedure

Participants were randomly assigned to evaluate four out twelve scenarios. Half of the scenarios were normative in nature and have previously been shown to have partisan differences in endorsement (pro-life vs. pro-choice, climate change as top vs. low priority; strict vs. lenient gun regulations, pro-life imprisonment vs. pro-death penalty, and pro-increase vs. pro-decrease military spending; see Pew Research Center, 2021, 2022, 2024). For instance, for the abortion scenario, participants were told, "Suppose that someone is deciding what their attitudes on abortion are. They are deciding between being pro-life (supporting a fetus' right to be born) or being pro-choice (supporting a woman's right to choose)".

The other half of the scenarios were neutral in nature but were hypothesized to be differentially endorsed (Mac vs. PC, horror vs. romance film, aisle vs. window airplane seat, relaxing vs. exciting vacation, adopting a cat vs. dog, and living in the city vs. the country). For instance, in the computer scenario, participants were asked to "Suppose that someone is deciding to buy their first computer. They are deciding between a Mac or a PC."

After reading the scenario, participants were asked, "How strong is your preference on this issue?" and indicated their response on a scale from 1 (*Not at all strong*) to 7 (*Very strong*) anchored at 4 (*Somewhat strong*). Then they were asked, "What is your preference on this issue" to which they indicated their response on a scale from 1 (*Strongly prefer [Mac/Pro-Life]*) to 7 (*Strongly prefer [PC/Pro-Choice]*) anchored at 4 (*Indifferent/No preference*). Finally, they were asked, "In your opinion, to what extent is this a moral issue?" and responded on a scale from 1 (*Not at all*) to 7 (*Very much*) anchored at 4 (*Somewhat*). Participants completed these three items

for each of the four scenarios they were randomly assigned to. Finally, participants completed a few demographic items (i.e., age and gender), and their political orientation (1 = *Extremely liberal*, 4 = *Moderate: Middle of the Road*, 7 = *Extremely conservative*).

Results

To examine our first criterion to select normative scenarios in which preferences were correlated with political orientation, we examined the bivariate correlations between participants' self-reported political orientation (1 = *Extremely liberal*, 7 = *Extremely conservative*) with their preferences across the six normative scenarios (e.g., 1= *Strongly prefer pro-life*, 7 = *Strongly prefer pro-choice*; see Table S2A). Results suggested that preferences regarding abortion, immigration, climate change, and gun control were most strongly correlated with political orientation, such that conservatives were less likely to strongly prefer pro-choice, more likely to strongly prefer closed borders, more likely to strongly prefer viewing climate change as a low priority, and more likely to strongly prefer lenient gun regulations. Political orientation was much more weakly correlated with preferences regarding the death penalty as well as military spending. Therefore, preferences regarding abortion, immigration, climate change, and gun control met our first criteria.

To examine our second criteria for selecting neutral scenarios in which preferences were bipartisan but exhibited bimodal distributions, we first examined the bivariate correlations between participants' self-reported political orientation (1 = *Extremely liberal*, 7 = *Extremely conservative*) with their preferences across the six neutral scenarios (e.g., 1= *Strongly prefer Mac*, 7 = *Strongly prefer PC*; see Table S2B). Results suggested that self-reported political orientation was only significantly correlated with job location, such that conservatives were more likely to strongly prefer choosing a job located in the country vs. the city. We then looked at the

distributions of preferences across the neutral scenarios. All of the scenarios demonstrated bimodal distributions except for the vacation scenario (see Figure S2A). Based on these results, the computer, film, seat, and pet scenarios met our criteria of having bipartisan preferences that had bimodal distributions.

All scenarios met our third criteria to be at or above the midpoint on a 7-point scale from 1 (*Not at all strong*) to 7 (*Very strong*) anchored at 4 (*Somewhat strong*) for the response to "How strong is your preference on this issue?" (see Figures S2A and S2B).

Lastly, to examine if the selected four normative scenarios (abortion, immigration, climate change, and gun control) were rated as significantly more moral than the four selected neutral scenarios (computer, film, seat, and pet), we ran a mixed linear OLS regression model predicting moral perceptions by the fixed effect of whether or not the domain was normative or neutral, while accounting for the random effect of the specific domain (N = 8) and the participant (N = 300). Results of the multi-level model suggested that the normative scenarios were rated as significantly and substantially more moral (M = 5.36, SE = 0.22) than neutral scenarios (M = 1.61, SE = 0.22), $t(6.01) = 11.98$, $p < .001$, $d = 2.91$, 95% CI [2.30, 3.52]. Based on these results, we selected the abortion, immigration, climate change, and gun control scenarios for the normative domain, and the computer, film, airplane seat, and pet scenarios for the neutral domain in our main study.

Table S2A

Bivariate Correlations Between Self-Reported Political Orientation and Preferences in Normative Domains (Study 2)

	1.	2.	3.	4.	5.	6.	7.
1. Political orientation (1 = Extremely liberal, 7 = Extremely conservative	--						
2. Abortion (1 = Strongly prefer pro-life, 7 = Strongly prefer pro-choice	-.63***	--					
3. Death penalty (1 = Strongly prefer death penalty, 7 = Strongly prefer life imprisonment)	-.29**	-.07	--				
4. Immigration (1 = Strongly prefer open borders, 7 = Strongly prefer closed borders)	.69***	-.51**	-.06	--			
5. Climate change (1 = Strongly prefer top priority, 7 = Strongly prefer low priority	.68***	-.57***	-.23	.52**	--		
6. Gun control (1 = Strongly prefer strict, 7 = Strongly prefer lenient)	.63***	-.68***	-.06	.21	.65***	--	
7. Military spending (1 = Strongly prefer increased, 7 = Strongly prefer reduced)	-.43***	.38*	.22	-.56**	-.33	-.67***	--

Note. ** $p < .01$, *** $p < .001$

Table S2B

Bivariate Correlations Between Self-Reported Political Orientation and Preferences in Neutral Domains (Study 2)

	1.	2.	3.	4.	5.	6.	7.
1. Political orientation (1 = Extremely liberal, 7 = Extremely conservative	--						
2. Computer (1 = Strongly prefer Mac, 7 = Strongly prefer PC	.06	--					
3. Film (1 = Strongly prefer horror, 7 = Strongly prefer romance)	.05	.03	--				
4. Seat (1 = Strongly prefer aisle, 7 = Strongly prefer window)	-.03	.07	.13	--			
5. Vacation (1 = Strongly prefer exciting, 7 = Strongly prefer relaxing	-.06	.08	.01	-.09	--		
6. Pet (1 = Strongly prefer cat, 7 = Strongly prefer dog)	.16	-.17	-.11	-.11	-.04	--	
7. Job (1 = Strongly prefer city, 7 = Strongly prefer country)	.38***	.16	-.03	-.21	.02	-.09	--

Note. *** $p < .001$

Figure S2A

Preference Distributions and Descriptive Statistics for Neutral Scenarios (Study 2)

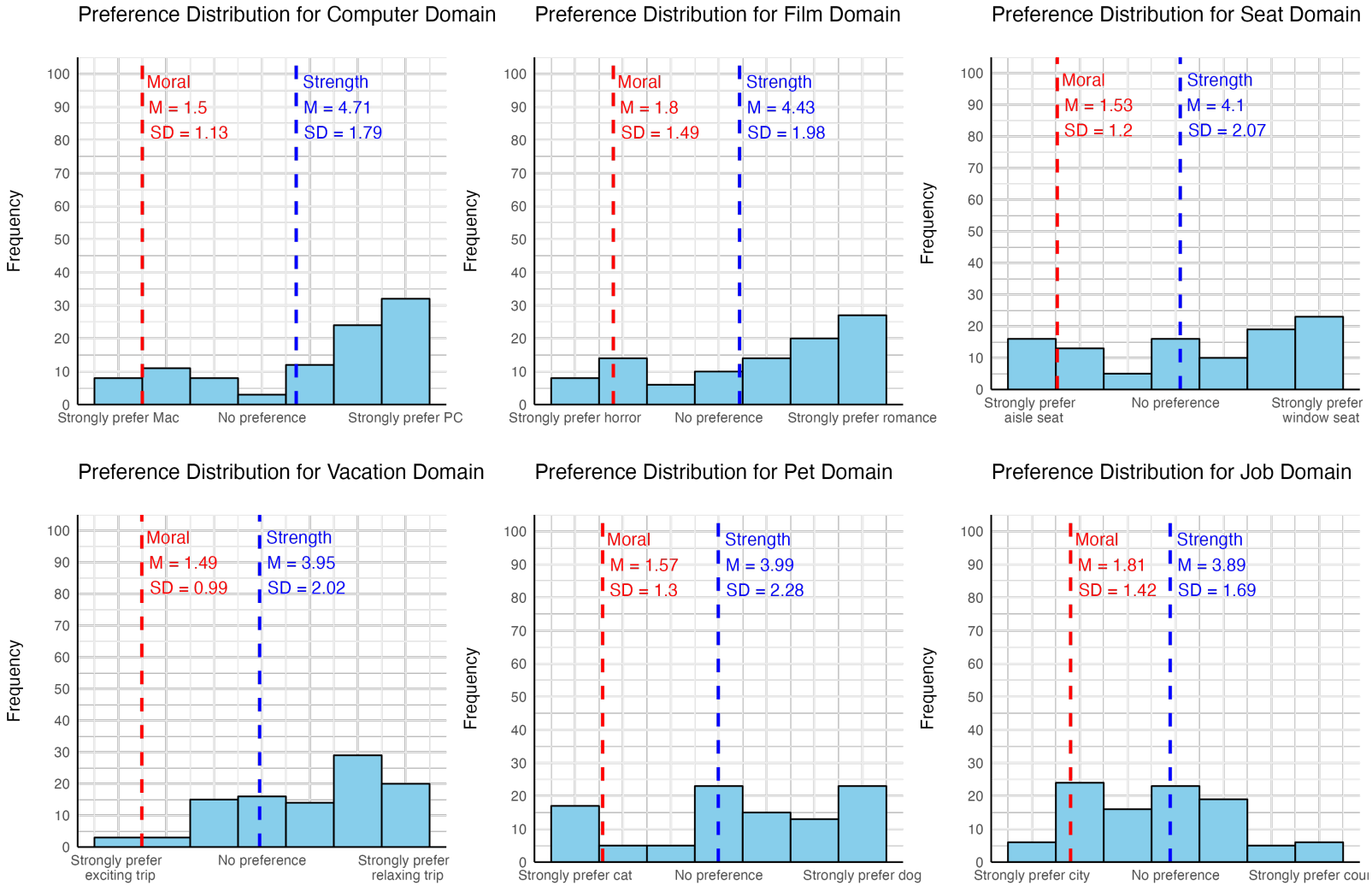
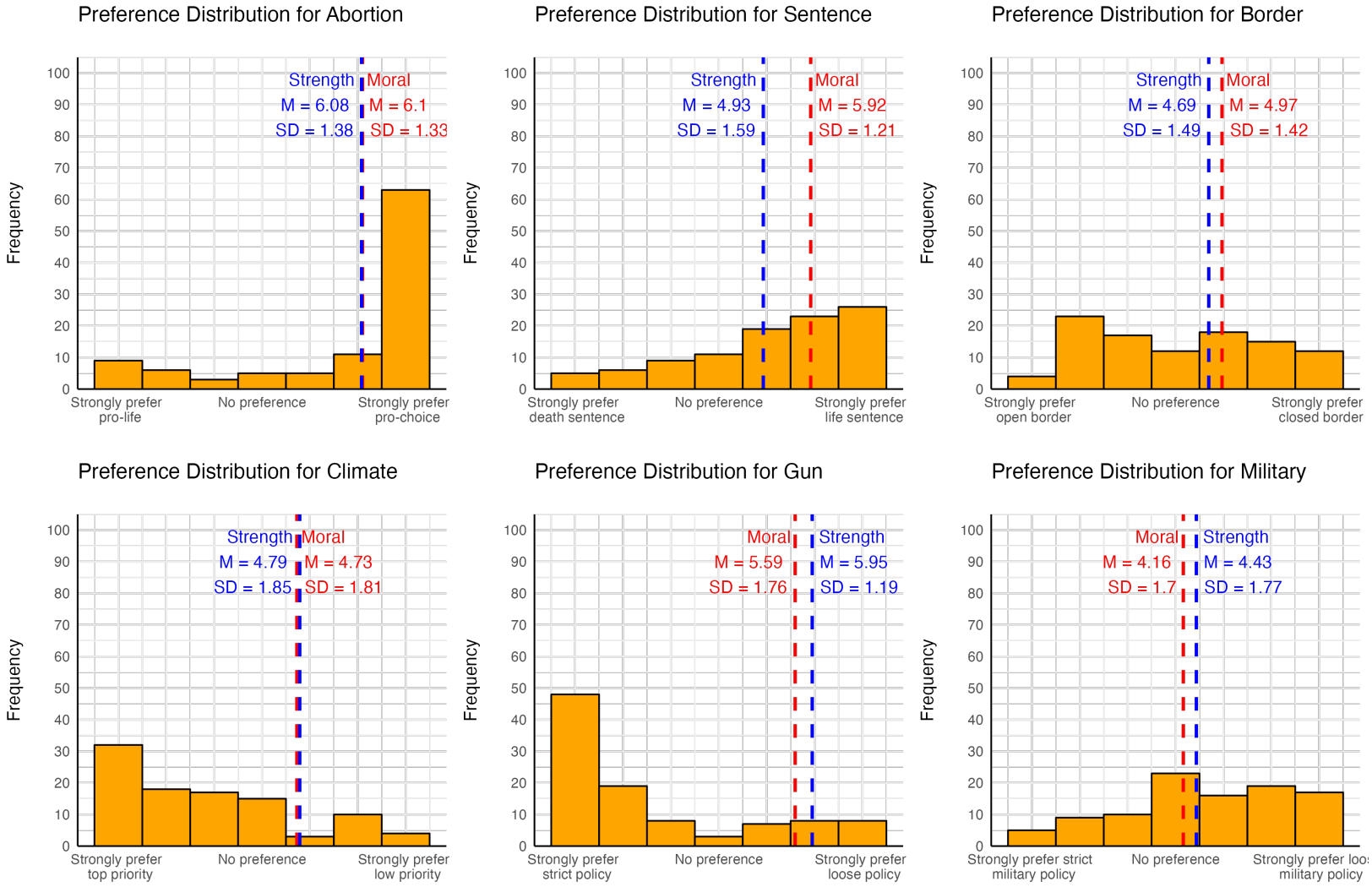


Figure S2B

Preference Distributions and Descriptive Statistics for Normative Scenarios (Study 2)



Supplementary Analyses (Main Text)

Self-Other Discrepancy. As noted in our preregistration, we predicted that there would be greater alignment between the target's preferences and participants' own preferences in normative versus neutral scenarios. First, we calculated the absolute difference between the strength of the target's preference and the strength of the participant's preference for each scenario. Then, we tested our prediction in two ways: first, we averaged these difference scores across all four scenarios and then performed a 2 (Domain: Neutral vs. Normative) x 2 (Assignment: Intuition Favors Option A. vs. Intuition Favors Option B) x 2 (Order: Intuition First vs. Deliberation First) ANOVA with the average difference scores as the dependent variable. Secondly, to generalize our findings to non-sampled stimuli (Yarkoni, 2022), we also replicated these results with a 2 (Decision Type: Normative vs. Neutral) X 2 (Assignment: Intuition Favors Option A vs. Intuition Favors Option B) X 2 (Option Counterbalancing: Deliberation Displayed First vs. Intuition Displayed First) mixed OLS regression model, with participants (N = 800) and scenarios (N = 8) treated as random effects. Replicating our results in the main text, and identically across both models, we observed a significant interaction between domain and assignment (see Table S2C). Results indicated that in the normative scenarios, it did not matter if intuition favored option A (M = 2.05, SE = 0.11) or if intuition favored option B (M = 1.99, SE = 0.11), $t(792) = 0.53$, $p = .594$. However, in neutral scenarios, the assignment of outcomes did matter, such that self-other discrepancy was greater when intuition favored option A (M = 2.49, SE = 0.11) versus option B (M = 2.03, SE = 0.11), $t(793) = 4.59$, $p < .001$. In other words, in normative scenarios, people are more likely to align their preferences with the target's preferences, regardless of whether or not intuition or deliberation favors a particular outcome.

However, in neutral scenarios, the assignment of outcomes to intuition and deliberation adjusts observers further away from targets (i.e., the 'pull' of intuition matters more).

Target Preferences. We also replicated our results in the main text by including the effect of order in the model (see Table S2D) and with participants (N = 800) and scenarios (N = 8) treated as random effects (see Table S2E). Importantly, across all four models, we observed the same significant domain X assignment interaction as noted in the main text, as well no significant interaction between domain, assignment, and participants' own strength of preferences. In general, the pattern of results suggests that the pull of intuition was stronger than the pull of deliberation in the neutral (vs. normative) scenarios, and this pattern of results did not vary significantly depending on participants' own strength of preferences.

Alignment. We also tested Hypothesis 2 (i.e., that observers will think that a person's authentic preference is whatever the observer believes is morally good), in a multilevel model. First, we recoded the dependent variable such that 1 = the preference that is good for republicans (e.g., prolife) and 7 = the preference that is good for democrats (e.g., prochoice) across the normative scenarios. Then, we conducted a 2 (Participant Politics: Republicans vs. Democrats) X 2 (Assignment: Deliberation Favors Option B vs. Intuition Favors Option B) ANOVA, with participants (N = 398) and scenarios (N = 4) treated as random effects. Results were nearly identical to those observed in the main text, in which we observed a significant main effect of assignment, $\chi 2(1) = 12.24$, $p < .001$, and a significant main effect of participant politics, $\chi 2(1) = 16.38$, $p < .001$, but no significant interaction between assignment and politics, $\chi 2(1) = 2.58$, $p = .108$. Supporting Hypothesis 2, independent of whether System 1 or System 2 favored the behavior, Democrats ($M = 4.22$, $SE = 0.11$) were significantly more likely than Republicans ($M = 3.86$, $SE = 0.11$), to believe that the target authentically preferred the option that was 'good' for

democrats, $d = 0.20$, 95% CI [0.06, 0.34]. That is, participants were more likely to believe that the target authentically preferred the option that aligned with whatever they believed was morally good.

Table S2C

ANOVA and Multilevel Regression Results Examining Self-Other Discrepancy

	Model 1 (ANOVA)		Model 2 (Multilevel)	
	$F(1, 792)$	p	χ^2	p
Domain	11.87	< .001	3.35	.067
Assignment	13.12	< .001	13.12	< .001
Domain * Assignment	**8.17**	**.004**	**8.17**	**.004**
Order	3.78	.052	3.78	.052
Order * Domain	3.48	.062	3.48	.062
Order * Assignment	2.79	.095	2.79	.095
Order * Domain * Assignment	0.03	.867	0.03	.867

Table S2D

ANOVA and OLS Regression Results Examining Target Preference

	Model 1		Model 2		
	F(1, 792)	*p*	*b*	*SE*	*p*
Domain	4.71	.030	-0.38	0.21	.078
Assignment	101.41	< .001	0.90	0.21	< .001
Domain * Assignment	**11.97**	**< .001**	**0.90**	**0.30**	**.003**
Order	5.58	.018	0.32	0.21	.133
Order * Domain	2.24	.135	0.45	0.31	.135
Order * Assignment	6.97	.008	-0.44	0.31	.146
Order * Domain * Assignment	0.62	.430	-0.29	0.43	.496
Strength of Preference			-0.24	0.21	.256
Strength * Domain			-0.12	0.28	.664
Strength * Assignment			0.26	0.28	.357
Strength * Order			-0.15	0.19	.423
Strength * Domain * Assignment			0.12	0.39	.750

Table S2E

ANOVA and OLS Regression Results Examining Target Preference (Multilevel)

	Model 1		Model 2		
	χ2	*p*	*b*	*SE*	*p*
Domain	2.55	.110	-0.49	0.23	.033
Assignment	101.41	< .001	0.89	0.21	< .001
Domain * Assignment	**11.97**	**< .001**	**0.94**	**0.30**	**.002**
Order	5.58	.018	0.29	0.21	.167
Order * Domain	2.24	.134	0.49	0.30	.103
Order * Assignment	6.97	.008	-0.38	0.30	.208
Order * Domain * Assignment	0.62	.430	-0.39	0.42	.361
Strength of Preference			0.12	0.02	< .001
Strength * Domain			-0.03	0.03	.509
Strength * Assignment			-0.02	0.03	.337
Strength * Order			0.01	0.02	.590
Strength * Domain * Assignment			0.04	0.04	.314

Study 3

Self-Other Discrepancy. For this dependent measure, we calculated the absolute difference between one's own candidate preference and their predictions of who the target would vote for (range = 0-6). Therefore, greater scores on this dependent measure reflects a greater self-other discrepancy (i.e., lower alignment) between one's own preferences and their predictions of the target's behavior. We then ran an OLS regression model predicting self-other discrepancy scores by assignment (intuition favors Harris vs. intuition favors Trump), belief in the voter's good true self, perceived similarity with the voter, as well as the interactions between assignment X true self beliefs, and assignment X perceived similarity.

Firstly, a greater belief in the voter's good true self predicted a lower discrepancy (i.e., greater alignment) between one's own preferences and their predictions of the voter's behavior, b = -0.32, SE = 0.07, $p < .001$, $\eta^2 = .05$. This means that the more that participants believed the voter had a good true self, the more likely they were to believe that the voter would ultimately vote for their own preferred candidate. Importantly, this effect held when controlling for how similar the voter was perceived to be. We also found support that greater perceived target similarity predicted a lower discrepancy between one's own preferences and their predictions of the voter's behavior , b = -0.27, SE = 0.06, $p < .001$, $\eta^2 = .06$. There was no significant main effect of assignment, b = -0.50, SE = 0.43, p = .245, as well as no significant interactions between assignment and true self beliefs, b = 0.06, SE = 0.10, p = .529, or between assignment and perceived similarity, b = 0.004, SE = 0.08, p = .957.

Option counterbalancing. We also conducted our analysis predicting self-other discrepancy scores with option counterbalancing (i.e., whether the head or gut was displayed first) included in the model. Results revealed no significant main effect of order, b = -0.77, SE =

0.45, p = .084, as well as no significant interactions between order and assignment, b = -0.03, SE = 0.24, p = .892, order and true self beliefs, b = 0.11, SE = 0.10, p = .279, and order and similarity, b = 0.05, SE = 0.08, p = .513. Importantly, our main results remain unchanged when controlling for option counterbalancing. First, there was no significant main effect of assignment, b = -0.43, SE = 0.45, p = .341. Instead, greater true self beliefs predicted lower self-other discrepancy scores, b = -0.37, SE = 0.09, p < .001, even when controlling for perceived similarity, which also predicted lower self-other discrepancy scores, b = -0.29, SE = 0.07, p < .001. There were no significant interactions between assignment and true self beliefs, b = 0.05, SE = 0.10, p = .609, or between assignment and perceived similarity, b = 0.003, SE = 0.08, p = .966.

Predicting target behavior. We also conducted our main analyses with predictions about who the target would vote for as the dependent variable, and participants' own voting preferences as a predictor. As a robustness check, we examined three different measures of participants' own voting preferences: (1) the categorical measure of who participants intend to vote for (Harris vs. Trump), (2) participants' party identification as indicated by their Prolific pre-screener (Democrat vs. Republican), and (3) participants' self-identified political orientation (1 = *Extremely liberal*, 7 = *Extremely conservative*). We therefore ran three separate regression models predicting who the target would vote for by the main effect of participants' own voting preferences, assignment, true self beliefs, perceived similarity, as well as the interactions between participant preferences and assignment, participant preferences and true self beliefs, and participant preferences and target similarity (see Table S3).

Results were consistent across all three models. That is, we observed a main effect of participants' own voting preference, such that left-leaning participants (i.e., people who prefer

Harris, pre-screened Democrats, and self-reported liberals) were more likely to believe that the target would vote for Harris; a main effect of true self beliefs, such that the greater the belief in the voter's true self, the more participants thought that the voter would ultimately vote for Harris; no main effect of assignment (whether the target's intuition favored Trump or Harris); and no main effect of perceived similarity (see Table S3). Importantly, and related to our central question (and H4), we observed a significant interaction between participants' own preferences and their belief in the voter's true self, which held after controlling for the interaction between participants' own preferences and their perceived similarity with the target, which was also statistically significant (except for model 3).

Table S3

Supplemental Regression Models Predicting Who the Target Will Vote For (Study 3)

	Model 1			Model 2			Model 3		
	b	*SE*	*p*	*b*	*SE*	*p*	*b*	*SE*	*p*
Participant Preference	1.48	0.51	.004	1.07	0.53	.042	0.24	0.12	.043
Assignment	0.20	0.44	.648	0.30	0.46	.522	0.28	0.47	0.56
True Self	0.33	0.09	< .001	0.34	0.09	< .001	0.52	0.12	< .001
Similarity	0.09	0.08	.262	0.08	0.09	.331	0.04	0.12	.704
Participant Preference X Assignment	-0.42	0.26	.099	-0.21	0.27	.441	-0.07	0.06	.252
Participant Preference X True Self	-0.44	0.11	< .001	-0.39	0.11	< .001	-0.10	0.03	< .001
Participant Preference X Similarity	-0.34	0.08	< .001	-0.27	0.09	.003	-0.04	0.02	.069
Assignment X True Self	-0.09	0.11	.410	-0.12	0.11	.271	-0.07	0.11	.525
Assignment X Similarity	0.12	0.08	.133	0.09	0.09	.291	0.09	0.09	.289

Note. Participant preference for model 1 represents the categorical measure of who participants intend to vote for (Harris vs. Trump). Participant preference for model 2 represents participants' categorical party affiliation as identified by their Prolific pre-screener (Democrat vs. Republican). Lastly, participant preference for model 3 represents participants' continuous political orientation (1 = *Extremely liberal*, 7 = *Extremely conservative*).

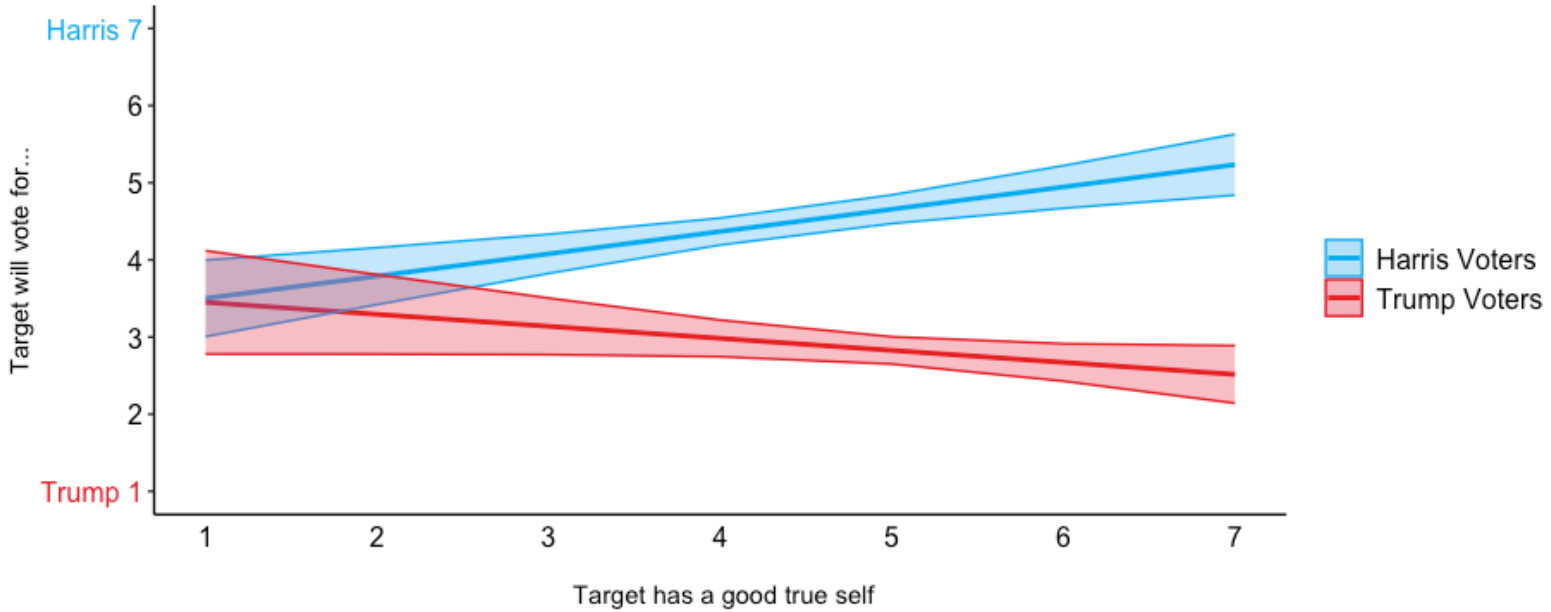
For model 1 (categorical measure) and as shown in Figure S3A, the more that Harris voters believed the voter had a good true self, the more that they thought the target would vote for Harris, $b = 0.29$, $SE = 0.07$, 95% CI [0.15, 0.43], $p < .001$. Conversely, for voters of Donald Trump, the more that participants believed the voter had a good true self, the more that they thought the target would vote for Donald Trump, although the effect was not statistically significant, $b = -0.16$, $SE = 0.08$, 95% CI [-0.32, 0.01], $p = .060$.

Similarly for model 2 (Prolific pre-screener; see Figure S3B), the more that Democrats believed that the voter had a good true self, the more they thought the target would vote for Harris, $b = 0.28$, $SE = 0.07$, 95% CI [0.13, 0.43], $p < .001$. Conversely, the more that Republicans believed that the voter had a good true self, the greater their belief that the target would vote for Donald Trump, although the effect was not statistically significant, $b = -0.11$, $SE = 0.08$, 95% CI [-0.28, 0.05], $p = .181$.

Lastly, for self-reported political orientation (see Figure S3C), we decomposed the simple slopes by looking at the effect of good true self beliefs predicting who the target would vote for at one standard deviation below ($M = 1.77$) and one standard deviation above ($M = 5.98$) the mean of self-reported political orientation. Results revealed that the more that liberal participants believed the voter had a good true self, the more likely they were to believe that the voter would vote for Harris, $b = 0.31$, $SE = 0.07$, 95% CI [0.16, 0.45], $p < .001$. On the other hand, the more that conservatives believed that the voter had a good true self, the more likely they were to believe that the voter would vote for Trump, although the effect was not statistically significant, $b = -0.12$, $SE = 0.08$, 95% CI [-0.28, 0.05], $p = .169$.

Figure S3A

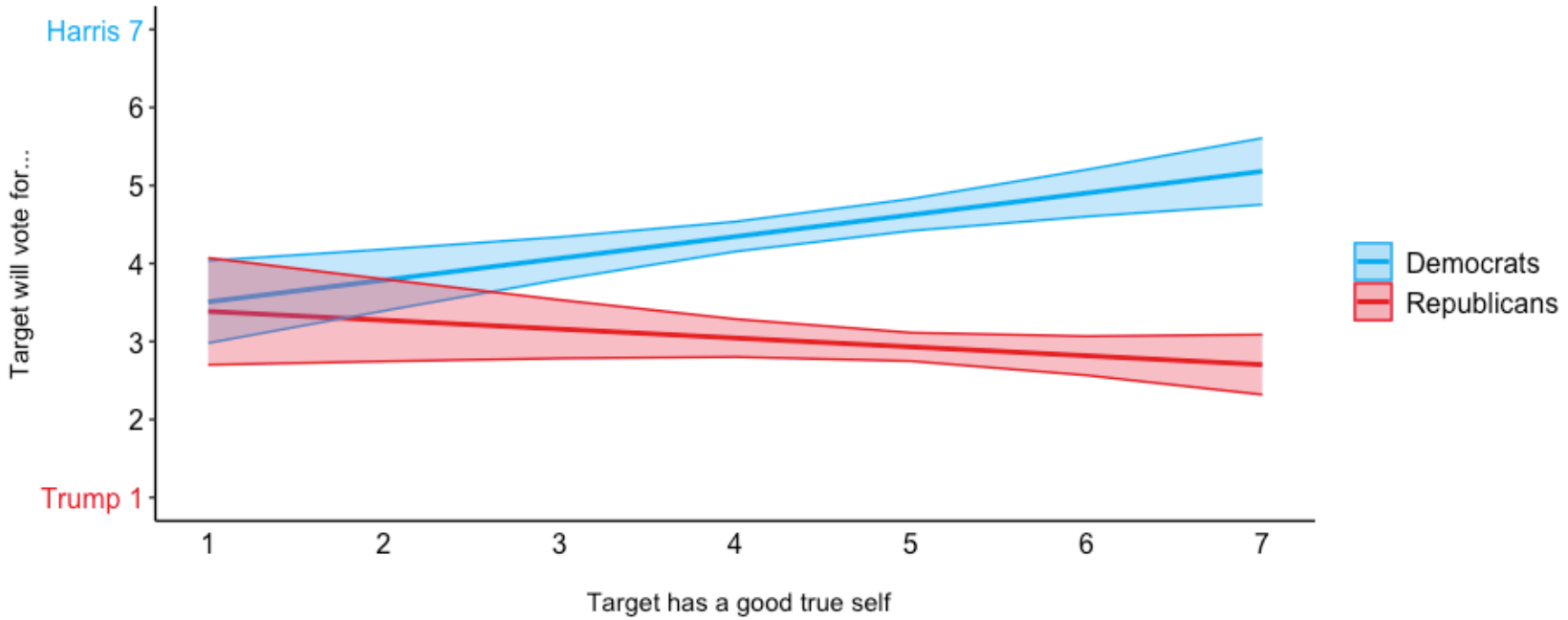
Belief in the Target's Good True Self Facilitates Alignment When Controlling for Perceived Similarity (Model 1)



Note. Error bands represent 95% confidence intervals.

Figure S3B

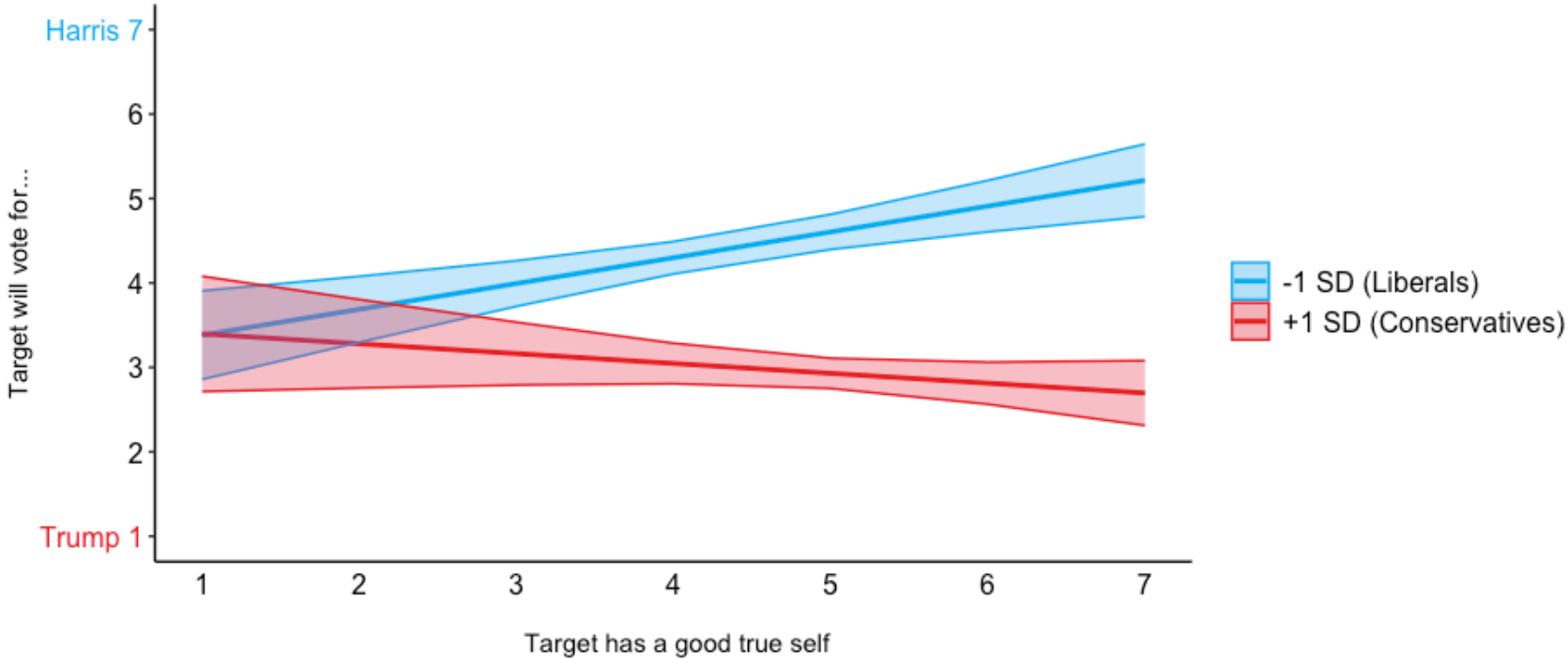
Belief in the Target's Good True Self Facilitates Alignment When Controlling for Perceived Similarity (Model 2)



Note. Error bands represent 95% confidence intervals.

Figure S3C

Belief in the Target's Good True Self Facilitates Alignment When Controlling for Perceived Similarity (Model 3)



Note. Error bands represent 95% confidence intervals.

Study 4

Self-Other Discrepancy. For this analysis, we calculated the absolute difference between the target's abortion preference and the observer's own abortion preference (range = 0-6) and then conducted a 2 (Target True Self: Good vs. Bad) X 2 (Assignment: Intuition Favors Pro-Choice Option vs. Intuition Favors Pro-Life Option) between-subjects ANOVA. Our dependent variable therefore reflects a self-other discrepancy in abortion preferences (i.e., higher scores = lower alignment). Results revealed a significant main effect of the target's true self, $F(1, 795) = 20.81$, $p < .001$, $\eta^2 = .025$, no significant main effect of assignment, $F(1, 795) = 0.40$, $p = .526$, $\eta^2 < .001$, and no significant interaction between the target's true self and assignment, $F(1, 795) = 0.02$, $p = .895$, $\eta^2 < .001$. Supporting H5, participants were more likely to align their beliefs with the target's preferences when the target was described as having a good ($M = 2.13$, $SE = 0.10$) as opposed to bad ($M = 2.75$, $SE = 0.10$) true self, $d = -0.32$, 95% CI [-0.46, -0.18]. That is, independent of whether one's intuition or deliberation favored a particular outcome, beliefs in a good true self explained the alignment between one's own beliefs and another's preferences.

Option counterbalancing. We also conducted our analyses predicting whether the target had a true preference (yes vs. no), as well as self-other discrepancy scores, with option counterbalancing (i.e., whether the pro-life or pro-choice option was displayed first) included in the model. A logistic regression model predicting whether or not the target had an authentic preference revealed only a significant main effect of true self beliefs, $b = 0.64$, $SE = 0.29$, $p = .030$. Participants were 1.89 times more likely, 95% CI [1.07, 3.37], to believe that the target had an authentic preference when the target was described as having a true self that was fundamentally good ($\hat{p} = 66.0\%$, $SE = 2.4\%$) as opposed to a true self that was fundamentally bad

($\hat{p}$ = 51.3%, *SE* = 2.5%). No other main effects or interactions were statistically significant (all *p*s > .352).

We then conducted a 2 (Target True Self: Good vs. Bad) X 2 (Assignment: Intuition Favors Pro-Choice Option vs. Intuition Favors Pro-Life Option) X 2 (Order: Pro-Life Option Displayed First vs. Pro-Choice Option Displayed First) between-subjects ANOVA examining mean differences of self-other discrepancy scores in abortion beliefs (range = 0-6). Similar to our results in the main text, we found a significant main effect of true self beliefs, $F(1, 791) = 20.07$, $p < .001$, $\eta^2 = .024$, such that participants were more likely to align their beliefs with the target's preferences when the target was described as having a good (*M* = 2.14, *SE* = 0.10) as opposed to a bad (*M* = 2.75, *SE* = 010) true self, *d* = -0.32, 95% CI [-0.46, -0.18]. Although we did not find a significant main effect of assignment, $F(1, 791) = 0.42$, $p = .518$, $\eta^2 < .001$, order, $F(1, 791) = 0.05$, $p = .822$, $\eta^2 < .001$, as well as no significant interaction between assignment and true self, $F(1, 791) = 0.01$, $p = .907$, $\eta^2 < .001$, or order and true self, $F(1, 791) = 0.19$, $p = .664$, $\eta^2 < .001$, we did observe a significant interaction between assignment and order, $F(1, 791) = 6.99$, $p = .008$, $\eta^2 = .010$. This interaction effect could be interpreted as a 'matching' effect between the target's intuition and whether the intuitive option was presented first. Specifically, when the target's intuition favored the pro-choice option, participants were more likely to align their beliefs with the target's preferences when the pro-choice option was presented first (*M* = 2.20, *SE* = 0.14) compared to when the pro-life option was presented first (*M* = 2.59, *SE* = 0.14), $t(791) = -2.01$, $p = .044$, *d* = -0.20, 95% CI [-0.40, -0.004]. Conversely, when the target's intuition favored the pro-life option, participants were more likely to engage in alignment when the pro-life option was presented first (*M* = 2.32, *SE* = 0.14), compared to when the pro-choice

option was presented first, (M = 2.65, SE = 0.14), although this effect was not statistically significant, $t(791) = -1.73$, $p = .084$, $d = -0.17$, 95% CI [-0.37, 0.02].

Subset Analysis. As indicated in our preregistration, we also conducted our main analysis after removing those who indicated that the target did not have an authentic preference (N after removal = 468). A 2 (Target True Self: Good vs. Bad) X 2 (Assignment: Intuition Favors Pro-Choice Option vs. Intuition Favors Pro-Life Option) between-subjects ANOVA revealed only a significant main effect of the target's true self, $F(1, 464) = 16.19$, $p < .001$, $\eta^2 = .033$. Simple effects revealed that participants were more likely to align their abortion beliefs with the target's abortion preferences when the target had a good (M = 2.14, SE = 0.13) as opposed to a bad (M = 2.94, SE = 0.15) true self, $t(464) = -4.02$, $p < .001$, $d = -0.37$, 95% CI [-0.56, -0.19].

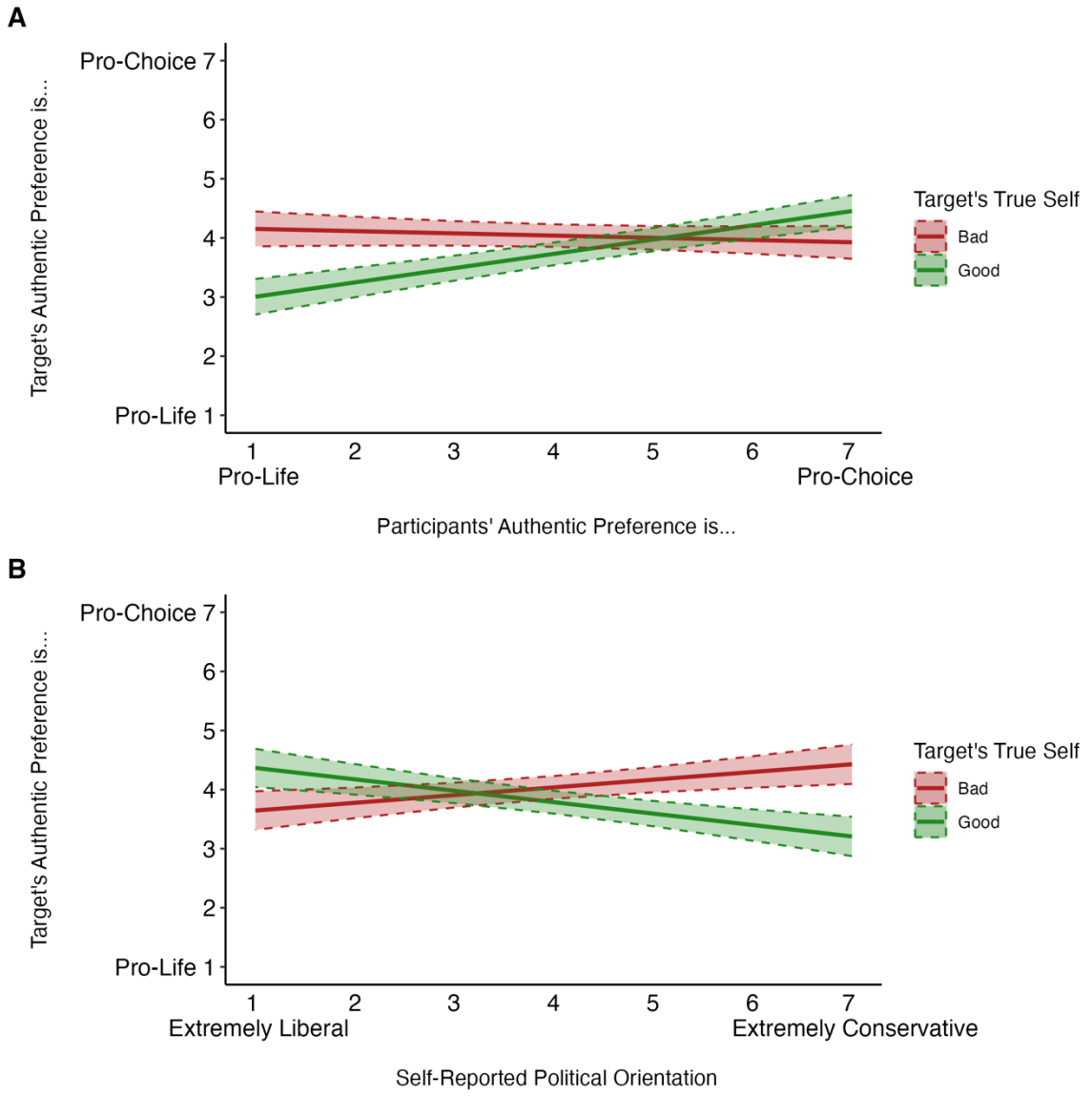
Predicting target's authentic preference. We also conducted our main analysis with the target's authentic abortion preferences (1=*Pro-Life*, 7 = *Pro-Choice*) as the dependent variable, and participants' own beliefs as a predictor. As a robustness check, we examined two different measures of participants' own beliefs: (1) the continuous measure of participants' abortion beliefs (1 = *Pro-life*, 7 = *Pro-Choice*) and (2) participants' self-identified political orientation (1 = *Extremely liberal*, 7 = *Extremely conservative*). We therefore ran two separate regression models predicting the target's authentic abortion preferences by the main effect of participants' own beliefs, the target's true self (good vs. bad), assignment (intuition favors pro-choice vs. intuition favors pro-life), and all two-way interactions (see Table S4 and Figure S4).

Results were consistent across both models. In model 1, we observed a significant interaction between participants' abortion beliefs (continuous) and the target's true self, b = 0.28, SE = 0.05, $p < .001$, such that participants were more likely to align their beliefs with the target's

authentic preferences when the target had a good, $b = 0.24$, $SE = 0.04$, $p < .001$, as opposed to a bad, $b = -0.04$, $SE = 0.04$, $p = .300$, true self. In model 2, we observed a significant interaction between participants' political orientation (continuous) and the target's true self, $b = -0.32$, $SE = 0.06$, $p < .001$, such that participants were more likely to align their beliefs with the target's authentic preferences when the target had a good, $b = -0.19$, $SE = 0.04$, $p < .001$, as opposed to a bad, $b = 0.13$, $SE = 0.04$, $p = .004$, true self.

Figure S4

People Align Their Beliefs With Another's Preferences Only When The Target Has A Good (Vs. Bad) True Self (Study 4)



Note. Error bands represent 95% confidence intervals. All figures display the interaction between participants' abortion beliefs and the target's true self. Figure S4A displays results for Model 1 (IV = continuous measure of participant abortion beliefs), and S4B displays Model 2 (continuous measure of participants' political orientation).

Table S4

Supplemental Regression Models Predicting Who the Target Will Vote For (Study 4)

	Model 1			Model 2		
	b	*SE*	*p*	*b*	*SE*	*p*
Intercept	4.19	0.22	< .001	3.48	0.27	< .001
Participant Beliefs	-0.03	0.04	.527	0.15	0.06	.008
True Self	-1.37	0.29	< .001	1.11	0.32	< .001
Assignment	-0.01	0.29	.983	0.07	0.32	.828
Participant Beliefs X True Self	0.28	0.05	< .001	-0.32	0.06	< .001
Participant Beliefs X Assignment	-0.02	0.05	.711	-0.04	0.06	.560
True Self X Assignment	-0.10	0.27	.701	-0.13	0.28	.631

Note. Participant beliefs for model 1 represents participants' self-reported abortion beliefs (1 = *pro-life*, 7 = *pro-choice*). Participant beliefs for model 2 represents participants' self-reported continuous political orientation (1 = *Extremely liberal*, 7 = *Extremely conservative*).

Supplemental Study 5: Individual Differences in Normative Beliefs

The goal of Supplemental Study 5 was to test the prediction that, if people are predisposed to believe that others authentically prefer whatever is normatively good, then what constitutes "good" behavior should depend on participants' own view of right and wrong (H2). In other words, individual differences in normative beliefs should predict differences in people's beliefs about others' authentic preferences.

Method

Participants

We aimed to recruit 400 adults on Amazon's Mechanical Turk. A total of 409 adults (42.5% female; M_{age} = 37.3, SD = 12.2) took the survey were paid a nominal fee for participating. This gave us 95% power (with α = .05) to detect a moderate sized effect (η^2 = .051).

Procedure

In Supplemental Study 5, we moved to a design where, instead of manipulating the normative valence of the behavior experimentally, we designed each decision scenario such that conservatives would be more likely to regard one decision option as "good" and liberals would be more likely to regard the other option as "good" (e.g., Graham et al., 2009; Newman et al., 2014). We then asked which type of processing (intuition vs. deliberation) best reflected the target's authentic preference.

Specifically, participants evaluated four vignettes featuring a target who was having conflicting feelings about a decision between two options: their intuition was pulling them towards one option, while deliberation was pulling them towards the other option (see Figure S5A). For example, in one scenario participants read about an individual who had conflicted thoughts about global warming. In one version, her "gut" feeling was to not believe in global

warming, but when she deliberated, she felt that she should view global warming as a serious problem. We hypothesized that in this scenario, liberals should say that deliberation reflects the target's authentic preference, while conservatives should say that intuition reflects the target's authentic preference. Conversely, in a matched scenario in which the target's "gut" feeling was to believe in global warming, but when she deliberated, she felt that she should not view global warming as a serious problem, we expected that liberals should say that intuition reflects the target's authentic preference, while conservatives should say that deliberation reflects the target's authentic preference.

Between-subjects, we then counterbalanced which decision outcome (i.e., the option that was good for liberals versus the option that was good for conservatives) was favored by the target's intuition versus the target's deliberation. Participants were asked to indicate the person's authentic preference by choosing between Option A, Option B, or "neither." Similar to Study 1 in the main text, after indicating the person's authentic preferences, participants were asked, "When trying to determine [his/her] true preference in this decision, [he/she] should … (1) Listen to their "gut" feeling, (2) Think for a long time, (3) Choose based on what other people would choose, (4) Choose based on what other people say they should choose, or (5) Other (please specify)."

After evaluating the four vignettes, participants indicated their own stance on each vignette's focal issue. Participants were asked to select the option from each vignette that was closest to their own views or select "indifferent/no opinion." For example, "When it comes to abortion…I am closer to pro-choice; I am closer to pro-life; Indifferent/no opinion." Finally, participants completed the following binary forced-choice item: "Generally speaking, which is closer to your social worldview?" by selecting "liberal" or "conservative."

Figure S5A

Vignettes used in Supplemental Study 5

Option that is "good" for conservatives listed first

Suppose that Matt is having conflicting feelings about supporting abortion.
- His "gut" feeling is to be "pro-life" and support a fetus' right to be born.
- When he thinks for a long time, he feels he should be "pro-choice" and support a woman's right to choose.

Suppose that Sarah is having conflicting feelings about global warming.
- Her "gut" feeling is to ***not*** believe in global warming.
- When she thinks for a long time, she feels she should view global warming as a serious problem.

Suppose that Alex is having conflicting feelings about his sexual life.
- His "gut" feeling is to be monogamous.
- When he thinks for a long time, he feels he should be sexually promiscuous and have sex with multiple partners.

Suppose that Emily is having conflicting feelings about her country.
- Her "gut" feeling is to be patriotic and openly express love for her country.
- When she thinks for a long time, she feels she should be deeply critical of her country.

Results

Manipulation check. As a check of our manipulation, we first compared the number of times that self-identified liberals and conservatives endorsed the "good for liberals" versus "good for conservatives" options from each vignette as being closer to their own stance on the corresponding issue. We coded each endorsement of the "good-for liberals" option as -1, each endorsement of the "good-for conservatives" option as +1, and "indifferent/no opinion" as 0, and summed across the four vignettes (range = -4 to +4). Confirming the validity of our manipulation, self-identified conservatives (n = 140) tended to select the "good for conservatives" option as closer to their worldview (M = 0.92, SD = 2.03) while self-identified liberals (n = 269) tended to select the "good for liberals" option as closer to their worldview (M = -2.00, SD = 1.64), $t(235) = 14.75$, $p < .001$, $d = 1.65$.

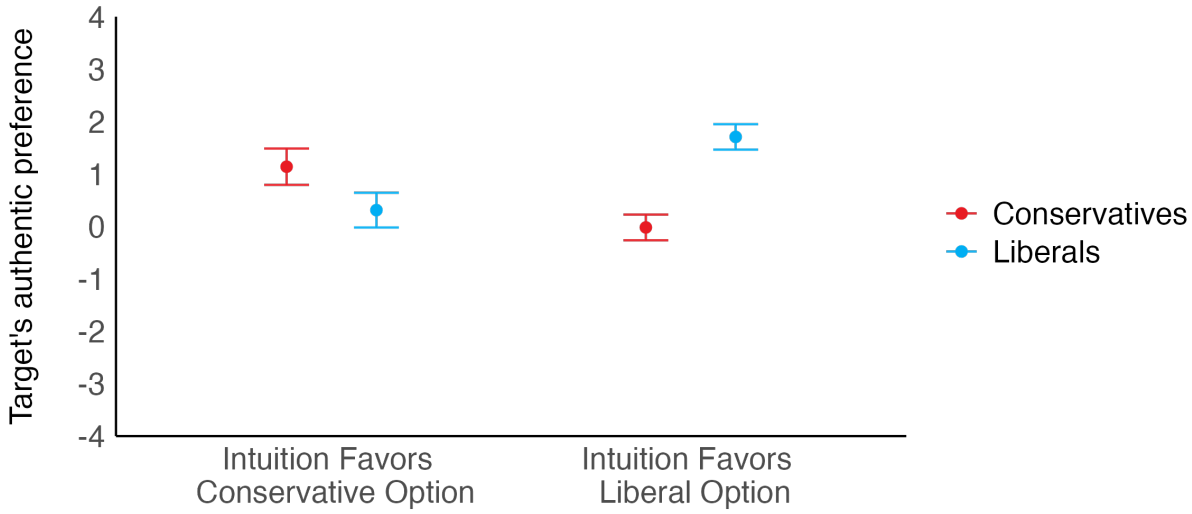
Endorsing authentic preferences. We recoded responses to each vignette as 1 (*Intuitive*), -1 (*Deliberate*), and 0 (*Neither*). We then summed the scores across vignettes to create a statistic that varied from "-4" (selected the deliberate option for all vignettes) to "4" (selected the intuitive option for all vignettes). We then performed a 2 (Participant Politics: Liberal vs. Conservative) X 2 (Assignment of Outcome) X 2 (Order of Presentation) ANOVA. Supporting H2, this analysis revealed a significant interaction between participant politics and the assignment of the outcomes to intuition versus deliberation, $F(1, 401) = 18.52$, $p < .001$, $\eta^2 = .042$.

As shown in Figure S5B, when the "good for liberal" options were presented as favored by intuition, liberals ($M = 1.71$, $SE = 0.24$) were significantly more likely than conservatives ($M = 0.31$, $SE = 0.33$) to say that the target truly preferred the intuitive option, $t(401) = 3.40$, $p = .001$, $d = 0.49$, 95% CI [0.21, 0.78]. Conversely, when the "good for conservative" options were presented as favored by intuition, conservatives ($M = 1.14$, $SE = 0.35$) were significantly more likely than liberals ($M = -0.02$, $SE = 0.25$) to say that the target truly preferred the intuitive option, $t(401) = 2.73$, $p = .010$, $d = 0.41$, 95% CI [0.11, 0.71]. In other words, supporting H2, participants said that the target truly preferred the option that aligned with their normative beliefs, regardless of whether it was favored by intuition or deliberation.

In line with H1, we also observed a significant, but small main effect of assignment, such that participants were in general more likely to say that intuition constituted the targets' authentic preferences ($M = 0.79$, $SD = 2.92$, one-sample t-test comparison to "0", $t(408) = 5.50$, $p < .001$).

Figure S5B

Individual Differences in Normative Beliefs Predict Beliefs about Authentic Preferences (Supplemental Study 5)



Note. Error bars represent the standard error of the mean. The y-axis ranges from -4 (*selected deliberation for all four vignettes*) to +4 (*selected intuition for all four vignettes*).

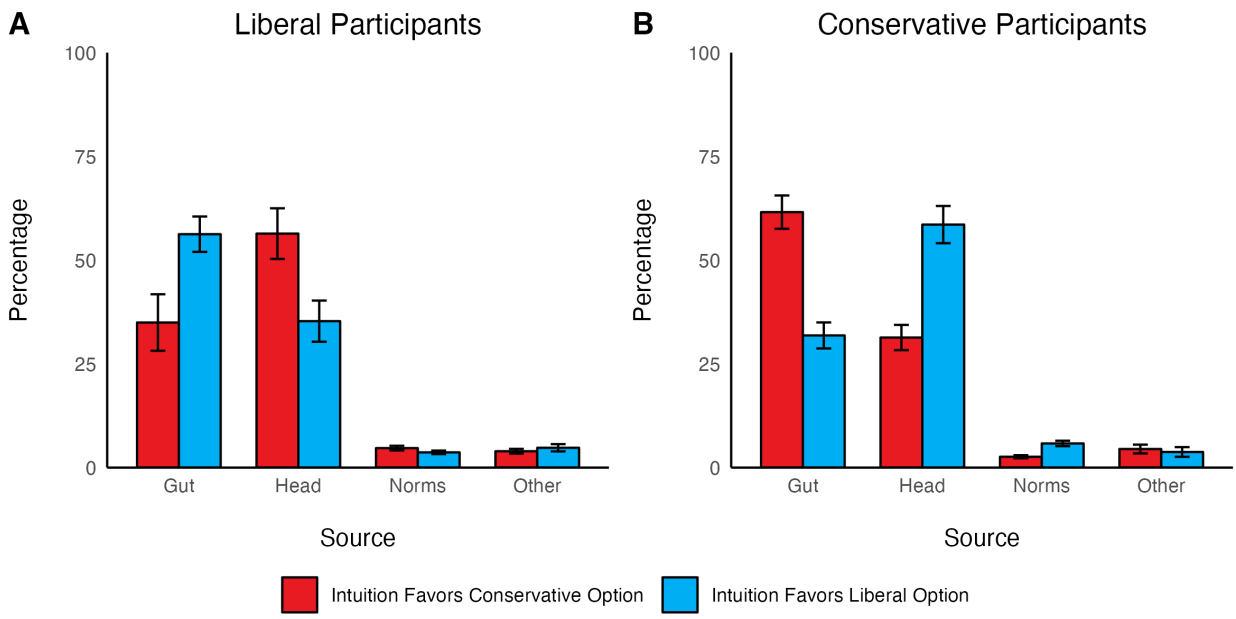
Source of authentic preferences. For each of the four scenarios, we calculated the percentage of participants who indicated that the target should listen to their gut, head, or norms (i.e., what other people would choose or say they should choose) to determine their authentic preference, and then averaged these responses across all four scenarios. We then statistically compared the average proportion selecting each justification across scenarios by Assignment (Intuition Favors Conservative Option vs. Intuition Favors Liberal Option), separately for self-identified liberals and conservatives. Replicating the effects of Study 1 and as shown in Figure

S5C, we found that for the exact same scenarios, liberal and conservative participants came to diverging views on the source of the target's authentic preferences.

When the target's intuition favored the liberal option (e.g., pro-choice), liberal participants were significantly more likely to say that, in trying to determine their authentic preference, the target should listen to their gut (M = 56.2%, SD = 8.5%) as opposed to their head (M = 35.3%, SD = 9.9%), $t(88.74) = 12.10$, $p < .001$. However, for the exact same scenarios (e.g., intuition favors pro-choice), conservative participants were significantly more likely to say that the target should instead listen to their head (M = 58.6%, SD = 9.0%) instead of their gut (M = 31.8%, SD = 6.3%), $t(59.89) = 14.14$, $p < .001$. Conversely, when the target's intuition favored the conservative option (e.g., pro-life), liberal participants were significantly more likely to say that the target should listen to their head (M = 56.4%, SD = 12.2%) instead of their gut (M = 35.0%, SD = 13.6%), $t(88.55) = 8.77$, $p < .001$. But conservative participants held the opposite view: for these scenarios (e.g., intuition favors pro-life), conservative participants were more likely to say that the target should listen to their gut (M = 61.6%, SD = 8.0%) instead of their head (M = 31.3%, SD = 6.1%), $t(51.16) = 16.57$, $p < .001$. Importantly, both liberal and conservative participants were unlikely (Ms < 6.0%, SDs < 1.50%) to indicate that descriptive or injunctive norms were the source of one's authentic preferences. Thus, the results are consistent with the idea that people hold a cognitive theory of authentic preferences, believing that one's authentic preferences are best reflected in a particular decision process (intuition vs. deliberation), while being unaware that their judgements shift dramatically based on whether an outcome is believed to be good or bad.

Figure S5C

Individual Differences in Normative Beliefs Predict Beliefs about the Source of One's Authentic Preferences (Supplemental Study 5)



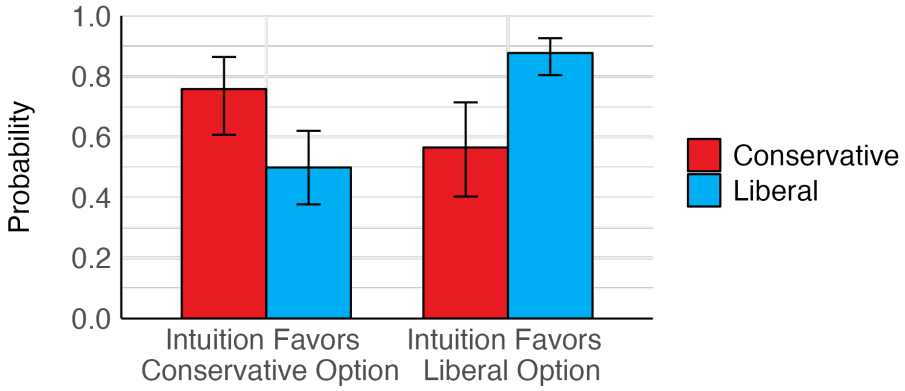
Note. Error bars represent the standard error of the mean across scenarios.

Endorsing authentic preferences (multilevel model). As in Study 1 in our main text, to generalize our findings to non-sampled stimuli (Yarkoni, 2022), we also replicated these results with a 2 (Participant Politics: Liberal vs. Conservative) X 2 (Assignment of Outcome) generalized mixed logistic regression model predicting whether the target truly preferred the intuitive or deliberative option, with participants (N = 402) and scenarios (N = 4) treated as random effects. After removing the seven cases where participants indicated that the target did not have an authentic preference (N = 402), the results of the model revealed a non-significant main effect of participant politics, $\chi^2(1) = 0.92$, $p = .338$, and a significant main effect of

assignment, $\chi^2(1) = 9.04$, $p = .003$, both of which were qualified by a significant Participant Politics X Assignment interaction, $\chi^2(1) = 20.63$, $p < .001$ (see Figure S5D). When one's intuition favored the option that was "good for liberals", liberals were significantly more likely ($P_{\text{authentic preference = intuition}} = 87.7\%$, $SE = 0.03$) than conservatives ($P_{\text{authentic preference = intuition}} = 56.5\%$, $SE = .08$) to believe that the target truly preferred the intuitive option, $OR = 5.52$, 95% CI [2.35, 12.23], $p < .001$. However, when the "good for conservative" options were presented as favored by intuition, conservatives were significantly more likely ($P_{\text{authentic preference = intuition}} = 75.8\%$, $SE = .07$) than liberals ($P_{\text{authentic preference = intuition}} = 50.0\%$, $SE = .06$) to believe that the target truly preferred the intuitive option, $OR = 3.14$, 95% CI [1.33, 7.44], $p = .009$. In other words, participants said that the target truly preferred the option that aligned with their normative beliefs, regardless of whether or not it was favored by intuition or deliberation (see "truepreferences_supplemental_study5.R").

Figure S5D

Individual Differences in Normative Beliefs Predict Beliefs about Authentic Preferences (Supplemental Study 5)



Note. Error bars represent the standard error of the predicted probabilities.

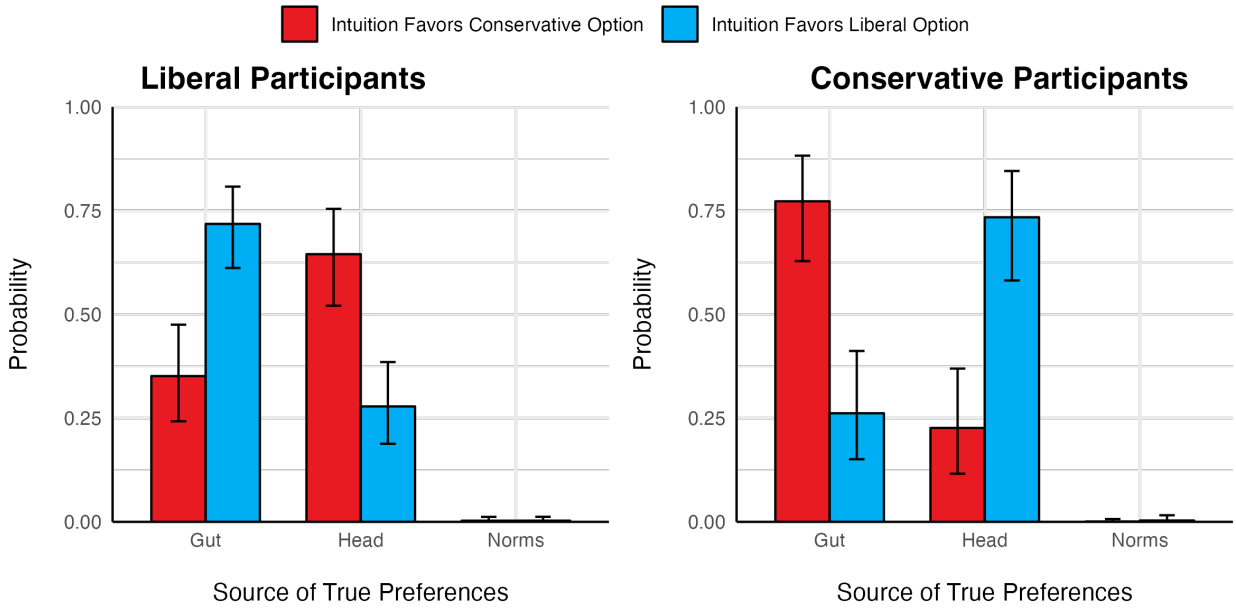
Source of authentic preferences (multilevel model). We also ran a 2 (Participant Politics) X 2 (Assignment of Outcome) mixed multinomial regression model predicting whether or not participants indicated the gut, head, or norms as the source of one's authentic preference, with participants (N = 399) and Scenarios (N = 4) treated as random effects (see Figure S5E). Replicating our "Source" analysis, when one's intuition favored the option that was "good for liberals", liberals were significantly more likely to indicate the gut (P = 71.8%, SE = 4.9%) as opposed to the head (P = 27.8%, SE = 4.9%) as the source of one's authentic preference, Z = 6.32, OR = 1.55, 95% CI [1.35, 1.78], $p < .001$. However, conservatives held the opposite view: when one's intuition favored the option that was "good for liberals", conservatives were significantly more likely to indicate the head (P = 73.4%, SE = 6.4%) as opposed to the gut (P = 26.2%, SE = 6.3%) as the source of one's authentic preference, Z = 5.28, OR = 1.60, 95% CI [1.35, 1.91], $p < .001$.

On the other hand, when one's intuition favored the option that was "good for conservatives," liberals were now significantly more likely to indicate the head (P = 64.5%, SE = 6.0%) as opposed to the gut (P = 35.1%, SE = 6.0%) as the source of one's authentic preference, Z = 3.43, OR = 1.34, 95% CI [1.14, 1.58], $p < .001$. The same was true for conservatives: when one's intuition favored the option that was "good for conservatives", conservatives were now significantly more likely to indicate the gut (P = 77.3%, SE = 6.2%) as opposed to the head (P = 22.6%, SE = 6.2%) as the source of one's authentic preference, Z = 6.26, OR = 1.73, 95% CI [1.45, 2.05], $p < .001$. Importantly, both liberal (P = 0.29%, SE = 0.22%) and conservative participants (P = 0.21%, SE = 0.21%) were unlikely to indicate that either descriptive or injunctive norms were the source of one's authentic preferences. Thus, the results are consistent with the idea that people hold a cognitive theory of authentic preferences, believing that one's

true preferences are best reflected in a particular decision process (intuition vs. deliberation), while being unaware that their judgements shift dramatically based on whether an outcome is believed to be good or bad.

Figure S5E

Individual Differences in Normative Beliefs Predict Beliefs about the Source of One's Authentic Preferences (Supplemental Study 5)



Note. Error bars represent the standard error of the predicted probabilities.

Supplemental Study 5 Discussion

In Supplemental Study 5 and supporting H2, we found that, for normative scenarios, participants are more likely to believe that a target truly prefers the good option, independent of whether that option was preferred by the target's intuition versus deliberation. Furthermore, this effect depends on participants' personal definitions of right and wrong, and thus systematically varies based on individual differences in normative beliefs.

Supplemental Study 6: Moderation by Belief in a True Self (2020 Election)

Supplemental Study 6 had three main goals. First, we sought to examine a downstream consequence of this phenomenon by examining behavioral predictions. Based on our theorizing, we predicted that when evaluating a prospective voter who is conflicted between two candidates (e.g., intuition prefers one candidate, while deliberation prefers the other), observers would think that deep-down, the conflicted voter would vote for (and authentically preferred) their own preferred candidate, independent of the mental processes (intuition versus deliberation) underlying the voter's decision.

Second, we offer a more direct test of the hypothesized role of beliefs in the true self. Specifically, we leverage a moderation-of-process approach (Pirlott & MacKinnon, 2016; Spencer et al., 2005) testing whether individual differences in belief in the true self moderate the current effect. That is, if true self beliefs are driving the tendency to believe that targets' authentic preferences reside with the normatively good option, then these results should be stronger (weaker) for participants higher (lower) in belief in the true self.

Finally, we included a new control condition in which participants were told that "part of him" prefers one candidate, whereas "another part" prefers another candidate without invoking intuition versus deliberation. We predicted that this would result in lower agreement between participants' own preferences and their beliefs about the target because there is no potential hierarchical arrangement to which participants might appeal to justify their preferred response.

Method

Participants

We recruited 600 MTurk participants with a goal of 100 participants per cell (N = 608 MTurk workers; 36.4% female; M_{age} = 37.05, SD = 10.99). This gave us 95% power (with α =

.05) to detect a small effect (η^2 = .021). We aimed to recruit roughly an equal number of Biden and Trump supporters using pre-existing Democrat and Republican panels on MTurk. This study was conducted in August 2020, prior to the 2020 Presidential election.

Procedure

We randomly assigned participants to one of three conditions. In each condition, participants read a vignette featuring a voter who was having conflicting feelings about the 2020 presidential election: part of the voter favored Joe Biden, while another part favored Donald Trump. The first two conditions framed the voter's preferences in terms of dual-processes: (1) the voter's intuition favored Biden, while their deliberation favored Trump, or (2) the voter's intuition favored Trump, while their deliberation favored Biden. In an Unspecified condition, we indicated that the target was conflicted, but did not specify which candidate was preferred by intuition (vs. deliberation) (i.e., "Part of him favors Donald Trump. Part of him favors Joe Biden"). The order of these statements was counterbalanced within condition.

Immediately following the vignette, participants made predictions about who the voter would ultimately vote for in the 2020 election on a 7-point scale ranging from 1 (*Definitely Biden*) to 7 (*Definitely Trump*) with a midpoint of 4 (*I don't know*). Participants also indicated which candidate they believed the voter truly preferred using a separate 7-point scale with the same labels. The order of the scale points was counterbalanced across participants, such that participants were randomly assigned to see both scales with Trump on the left or to see both scales with Biden on the left. Participants indicated who they planned to vote for in the 2020 presidential election: Joe Biden, Donald Trump, or "Other / Not sure." Finally, participants completed a three-item scale measuring the extent to which they endorsed the notion of a true self (i.e., "People have a true inner-nature that sometimes contrasts with their behavior"; "People

have a core personality that can be very different from what they show to others or even themselves"; "People have a "true" self that captures who they really are") using a 7-point scale ranging from 1 (*Strongly disagree*) to 7 (*Strongly agree*) (adapted from Newman et al., 2010; M = 5.44, SD = 0.99; 90.3% > the midpoint of the scale; ω = .79).[1]

Results

Alignment. In our primary analyses, we excluded 74 participants who indicated "Other / Not sure" regarding their own voting preference. This left 534 participants, who were evenly split by their intention to vote for Joe Biden (N = 264) versus Donald Trump (N = 270). For both dependent measures (i.e., prediction of the target's voting behavior and prediction of the target's true candidate preference), we first ran a 3 (Assignment: Gut = Trump, Gut = Biden, and Unspecified) X 2 (Candidate Preference: Biden vs. Trump) between-subjects ANOVA. For the measure of the target's voting behavior, we observed a large and significant main effect of participants' candidate preferences, $F(1, 528) = 122.91$, $p < .001$, $\eta^2 = .19$, such that Trump voters were significantly more likely to report that the target would vote for Trump (M = 5.10, SE = 0.10) compared to Biden voters (M = 3.46, SE = 0.11), d = 0.96, 95% CI [0.78, 1.14]. There was no main effect of assignment, $F(2, 528) = 1.15$, $p = .319$, $\eta^2 = .003$, and no significant interaction between one's own candidate preferences and assignment, $F(2, 528) = 0.81$, $p = .446$, $\eta^2 = .002$, That is, participants aligned their beliefs with the target's behavior, independent of the decision conflict of the voter (see Figure S6A).

Results were similar for the measure of the target's authentic preference. We observed a large and significant main effect of participants' candidate preferences, $F(1, 528) = 89.65$, $p <$

[1] We note that we did not specify "morally good" in these items to avoid a potential demand effect. However, existing research has documented that people see the true self as morally good (see Strohminger et al., 2017 for a review).

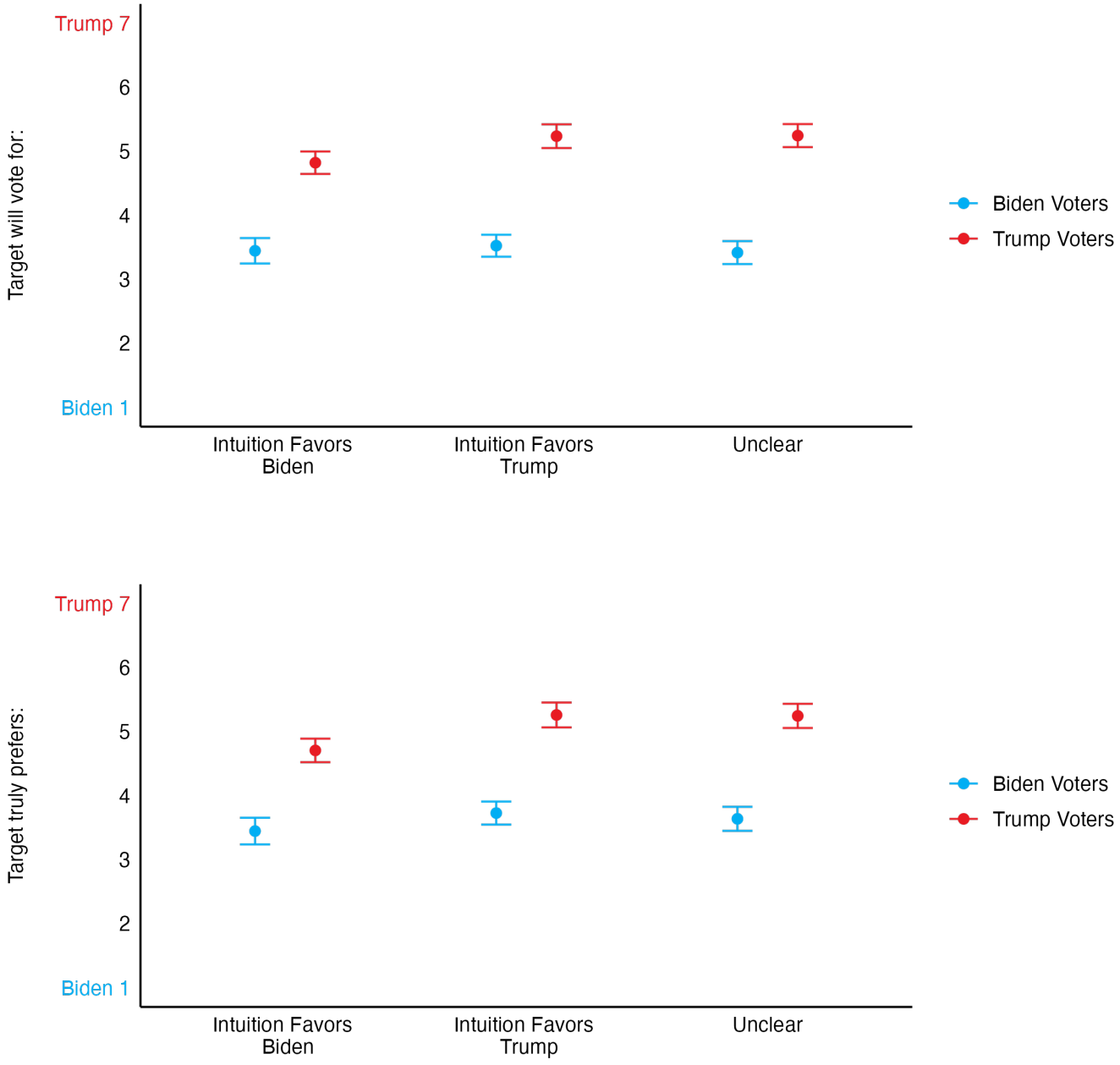
.001, η^2 = .14, such that Trump voters were significantly more likely to report that the target would vote for Trump (M = 5.07, SE = 0.11) compared to Biden voters (M = 3.60, SE = 0.11), d = 0.82, 95% CI [0.64, 0.91]. There was a marginally significant main effect of assignment, F(2, 528) = 2.95, p = .053, η^2 = .01, and no significant interaction between one's own candidate preferences and assignment, F(2, 528) = 0.45, p = .639, η^2 = .001 (see Figure S6A).

Moderation by belief in the true self. We then ran a linear regression model with assignment (i.e., Gut = Trump; Gut = Biden; Unspecified), participants' own candidate preferences (coded as a binary variable; 1 = Trump, 0 = Biden), true self beliefs, and all of their two-way interactions as predictors.

For the measure of the target's voting behavior and supporting H4, we observed the predicted interaction between participants' candidate preferences and their belief in the true self, b = 0.56, t(524) = 3.56, p < .001. For supporters of Donald Trump, the more that participants endorsed a belief in the true self, the more that they thought the target would vote for Donald Trump, b = 0.33, SE = 0.11, 95% CI [0.10, 0.56]. Conversely, for supporters of Joe Biden, the more that participants endorsed a belief in the true self, the more that they thought the target would vote for Joe Biden, b = -0.23, SE = 0.10, 95% CI [-0.43, -0.02]. To further explore the significant interaction, we used the Johnson-Neyman technique (Johnson & Fay, 1950; Spiller et al., 2013) to map the zones of significance (see Figure S6B). This analysis revealed that participants projected their own candidate preferences onto the target voter when their belief in the true self score exceeded 3.57 out of 7.

Figure S6A

Participants Align Their Own Candidate Preferences with Others' Authentic Preferences and Subsequent Behavior (Supplemental Study 6)



Note. Error bars represent the standard error of the mean

Neither of the main effects for candidate preference (b = -1.38, t(524) = -1.53, p = .126), belief in the true self (b = -0.14, t(524) = -1.09, p = .276), nor assignment ($b_{\text{Gut=Trump vs. Gut = Biden}}$ = 2.04, t(524) = 1.94, p = .053) were statistically significant. There was a statistically significant interaction between assignment and true self beliefs, which can be interpreted as an effect of intuition: b = -0.40, t(524) = 2.06, p = .040. Supporting H1, when the target's gut favored Biden, both Republicans and Democrats were more likely to say that the target would vote for Biden, and when the target's gut favored Trump, both Republicans and Democrats were more likely to say that the target would vote for Trump.

The same analysis on beliefs about the target's authentic preference yielded a small main effect of participants' own candidate preferences, b = -1.87, t(524) = 1.98, p = .048, which was qualified by the predicted interaction between participants' candidate preference and their belief in the true self (see Figure S6B), b = 0.62, t(524) = 3.76, p < .001. Supporting H4, simple slopes indicated that, for supporters of Donald Trump, the more that participants endorsed a belief in the true self, the more that they thought the target, deep down, preferred Donald Trump, b = 0.36, SE = 0.12, 95% CI [0.11, 0.60]. Conversely, for supporters of Joe Biden, the more that participants endorsed a belief in the true self, the more that they thought the target, deep down, preferred Joe Biden, b = -0.26, SE = 0.11, 95% CI [-0.48, -0.05]. For this interaction, participants projected their own candidate preferences onto the target voter when their belief in the true self score exceeded 4.01 out of 7. We did not observe any other significant main effects or interactions.

Dual-process framing facilitates more confident predictions. We predicted that when the voter's preferences are framed in the context of dual-process, observers would be more likely to project their own normative beliefs onto the voter compared to when the underlying mental

processes are left ambiguous. We designed the two dependent measures (i.e., predictions of who the voter would vote for and who the target truly preferred) such that values closer to the scale's endpoints indicated greater response certainty (i.e., 1 = *Biden*, 7 = *Trump*). Therefore, we would expect more extreme predictions of candidate preferences and voting behavior (i.e., greater absolute distance from the scale's midpoint; range = 0-3) when mental processes are explicitly assigned, since, under this dual-process framing, participants may be tempted to conclude that one decision process is more central than the other. Note that these two dependent variables are the same from the "Alignment" and "Moderation by belief in the true self" sections, but are recoded to reflect response certainty (i.e., greater scores = response is closer to either Trump or Biden).

The data supported this prediction. A planned contrast revealed that compared to when the decision conflict was ambiguous (M = 1.43, SD = 1.08), participants made more confident predictions of voting behavior when the decision conflict was framed in dual-process (i.e., either the voter's intuition favored Biden and their deliberation favored Trump, or the voter's deliberation favored Biden and their intuition favored Trump; (M = 1.70, SD = 0.99), t(336.60) = 2.83, p = .005, d = 0.27, 95 % CI [0.09, 0.45]. Likewise, compared to when the decision conflict was ambiguous (M = 1.36, SD = 1.16), participants made more confident predictions of the voter's authentic candidate preference when the voter's preference was framed in terms of dual processes (M = 1.79, SD = 0.98), t(315.09) = 4.21, p < .001, d = 0.41, 95% CI [0.23, 0.59].

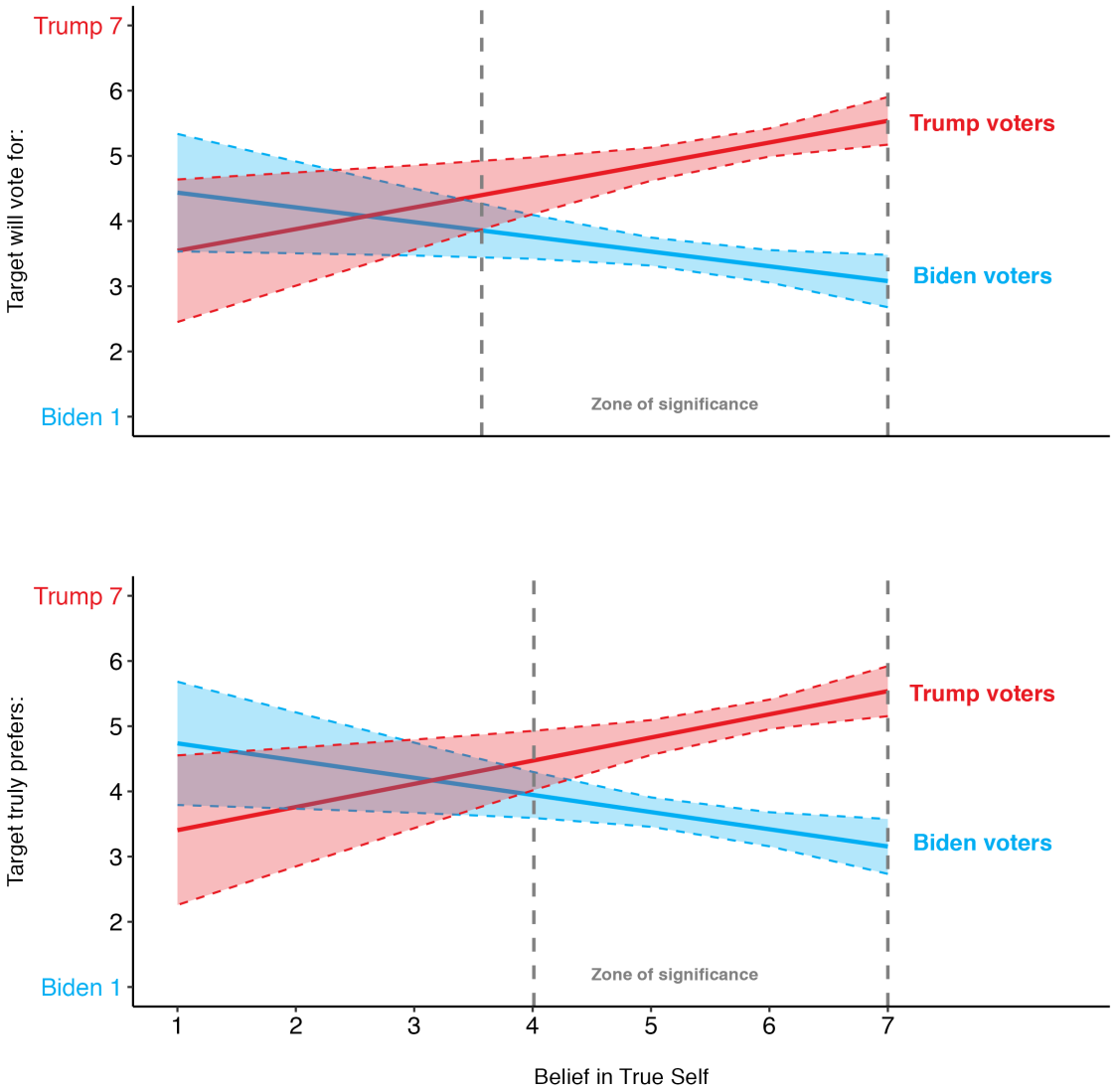
Supplemental Study 6 Discussion

Using data collected prior to the 2020 Presidential election, we directly tested the role of belief in the true self as facilitating the alignment between one's candidate preference (i.e., whether a participant intended to vote for Biden or Trump) and another's political preferences

(i.e., whether or not a conflicted target would vote for and authentically preferred Biden or Trump). Firstly, although participants aligned their beliefs with the preferences of the conflicted voter in all conditions, participants made more confident or extreme predictions when the conflict was framed in a dual-process vs. unspecified manner. That is, we observed more extreme predictions of candidate preferences and voting behavior when mental processes were explicitly assigned, since, under this dual-process framing, participants may be tempted to conclude that one decision process is more central than the other. Importantly, the alignment between predictions of the voter and participants' own candidate preferences was moderated by belief in the true self. The more that participants believed in the existence of a true self, the more they predicted that the target's authentic preference was aligned with their own preferences, independent of intuition or deliberation.

Figure S6B

Belief in the True Self Facilitates Egocentric Alignment Between One's Own Candidate Preferences and Others' Authentic Preferences and Subsequent Behavior (Supplemental Study 6)



Note. Gray dotted lines represent levels of belief in a true self for which participants significantly projected their own preferences onto their predictions of the target's true preference and voting behavior (Johnson-Neyman technique). Colored error bands represent 95% confidence intervals.

Supplemental Study 7: Moderation by Dual Processes

We reasoned that when a decision conflict is framed in terms of a dual-process, observers can appeal to the notion of a true self—i.e., that one type of mental process is more central than the other. This is consistent with Kunda's (1990) suggestion that an individual's tendency to engage in motivated reasoning is constrained by the extent to which the individual is able to marshal evidence in support of their position. When there's a broader, hierarchical arrangement of intuition vs. deliberation that one can appeal to, individuals are able to justify and rationalize the belief that, because the target has a "good true self", they truly prefer to listen to the decision process that is pulling towards an outcome an observer believes is normatively "good" or "right." In this way, individuals are able to justify and rationalize the belief that "deep down, others share my view."

However, when the underlying mental processes are not defined, or there is no stated conflict, it is more difficult to justify that one option is more reflective of the person's true self than the other option. That is, in these cases, observers may not be able to justify their motivated reasoning, and thus, may be less likely to conclude that a target ultimately shares their preferences, deep down. Therefore, we expected the effect of observers' own normative beliefs on their predictions about the target would occur in the dual process cases more so than in the other conditions.

Moreover, we wanted to examine a downstream consequence of this phenomenon using a behavioral measure. Specifically, we explored whether believing that the target shared their beliefs deep-down would encourage participants to try to persuade the target to adopt their own position on the issue. Empirical research shows that the closer one is to achieving a goal, the more motivation and resources are mobilized towards achieving that goal (Kivetz et al., 2006;

see also Hull, 1932). Applied to the current context, when participants believe that a target shares their beliefs "deep down," they may perceive the target as being closer to accepting their beliefs or viewpoints.

This perceived proximity to the goal may enhance their motivation to persuade the conflicted target, because the target will be viewed as already leaning towards, and truly preferring, the "right" or "good" option. On the other hand, when a target unambiguously holds a position contrary to one's own, or is conflicted but it is not clear which option is associated with which decision process, observers may perceive these targets as less persuadable because the gap between their current stance and what is deemed "good" or "right" is greater in the former or made uncertain in the latter. We therefore explored whether participants in the dual-process conditions – where intuition explicitly favors one option and deliberation another – were more likely to see the target as persuadable and to try to persuade them compared to cases where the target was either not conflicted, or was conflicted but the option that was associated with intuition or deliberation was not specified.

Method

Participants

We recruited a sample of 1,004 adults from Prolific Academic (46.4% female, M_{age} = 39.58, SD = 13.81) in a preregistered design (https://aspredicted.org/T3S_RSP). This gave us 95% power (with α = .05) to detect a small effect (η^2 = .020). We aimed to recruit roughly an equal number of pro-choice and pro-life participants. Because in the U.S., abortion attitudes correlate with political ideology (e.g., Osborne et al., 2022), we recruited from pre-existing Democrat and Republican panels on Prolific Academic.

Procedure

Participants were first told "In this study, we would like your help coding some in-person interviews that were conducted earlier this year. You will be shown a snippet of an interview where we asked someone what they think about abortion. Please read the interview and respond to the questions." Then, participants were randomly assigned to read one of five snippets of text. In two of the scenarios, the target was conflicted such that deliberation favored one option (e.g., pro-life) but intuition favored the other (e.g., pro-choice). In two additional scenarios, the target was not conflicted and only favored one option. In a final scenario, the target was portrayed as conflicted, but we did not specify which option was associated with which decision process. The texts were as follows:

Interviewer: "So, would you consider yourself pro-choice or pro-life?"

Participant:

[Dual-Process Condition where Intuition = Pro-life]: "Hmm, well…My head is telling me one thing, while my gut tells me another. My gut reaction is that abortion just feels wrong. It's important to protect the unborn baby. But when I think about it for a while, I also think it's important that women should decide for themselves and abortion bans deny that. So, I guess you could say that I'm pretty conflicted about the whole issue. I'm willing to listen to arguments on both sides."

[Dual-Process Condition where Intuition = Pro-choice]: "Hmm, well…My head is telling me one thing, while my gut tells me another. My gut reaction is that abortion bans just feel wrong. It's important that women should decide for themselves. But when I think about it for a while, I also think it is important to protect the life of an unborn baby. So, I guess you could say that I'm pretty conflicted about the whole issue. I'm willing to listen to arguments on both sides."

[Unambiguous Condition where Preference = Pro-life]: "Hmm, well… Abortion is just wrong. It's important to protect the unborn baby. That said, I'm willing to listen to arguments on both sides."

[Unambiguous Condition where Preference= Pro-choice]: "Hmm, well… Abortion bans are just wrong. It's important that women should decide for themselves and abortion bans deny that. That said, I'm willing to listen to arguments on both sides."

> [Ambiguous Condition]: "Hmm, well…My head is telling me one thing, while my gut tells me another. So, I guess you could say that I'm pretty conflicted about the whole issue. That said, I'm willing to listen to arguments on both sides."

All participants were then asked to respond to the following: "I think this person's true preference is... 1 (*Pro-choice*), 4 (*Indifferent*), 7 (*Pro-life*). Participants were also asked, "How likely is it that this person could be persuaded to see abortion in the same way that you do?" on a 7-point scale from 1 (*Extremely Unlikely*) to 7 (*Extremely Likely*), and "Would you like to send this participant a note to try to persuade them of your view on abortion?" coded as 1 (*Yes*) and 0 (*No*). If participants responded "Yes", they were presented with a text box and the prompt, "Please type your note in the space provided below."

Finally, participants indicated their own beliefs about abortion, "Now, please tell us about your own views. When it comes to abortion... (I am closer to pro-choice, I am closer to pro-life, I am indifferent/have no opinion)." In this sample, 61.3% of participants indicated that they were "closer to pro-choice," 31.2% indicated they were "closer to pro-life", and 7.5% indicated "I am indifferent/have no opinion." For our primary analysis, and as indicated in our preregistration, we exclude participants who indicated that they were indifferent on the issue of abortion (n = 75). Participants also provided basic demographic information including age and gender as well as their political orientation, ranging from 1 (*Extremely liberal*) to 7 (*Extremely conservative*).

Results

True preferences. We first conducted a 2 (Participant Beliefs: Pro-Choice vs. Pro-Life) X 5 (Assignment: Intuition Favors Pro-Life, Intuition Favors Pro-Choice, Unambiguously Pro-Life, Unambiguously Pro-Choice, Ambiguous) between-subjects ANOVA on Likert ratings of the target's beliefs. In line with our predictions, we observed a significant Participant Beliefs X Assignment interaction, $F(4, 915) = 3.20$, $p = .013$, $\eta^2 = .013$. In the Intuition Favors Pro-Life

condition, pro-choice participants (M = 3.74, SE = 0.11) were significantly more likely than pro-life participants (M = 3.24, SE = 0.16) to say that the target was truly pro-choice, $t(915) = 2.62$, $p = .008$, $d = 0.42$, 95% CI [0.10, 0.73]. Similarly, in the Intuition Favors Pro-Choice condition, pro-choice participants (M = 3.96, SE = 0.11) were again significantly more likely than pro-life participants (M = 3.56, SE = 0.15) to say that the target was pro-choice, $t(915) = 2.16$, $p = .031$, $d = 0.33$, 95% CI [0.03, 0.63]. Thus, in the two Dual-Process conditions, participants said that the target authentically preferred the option that aligned with their normative beliefs, independent of whether it was favored by intuition or deliberation.

In contrast, there were no significant differences in ratings of the target's authentic preference when the target's stated position was unambiguous (only pro-life or pro-choice) or completely ambiguous (a conflict was stated, but the underlying mental processes was not defined), all ps > .28, see Table S7A. There was also a significant main effect of assignment, $F(4, 915) = 163.81$, $p < .001$, $\eta^2 = .52$, but no significant main effect of the observer's own preference, $F(1, 915) = 1.06$, $p = .304$, $\eta^2 < .001$.

Table S7A

Marginal Means of Condition and Participant Beliefs in Predicting Ratings of the Target's Beliefs and Likelihood of Being Persuaded (Supplemental Study S7A)

	Dual-Process		**Unambiguous**		**Ambiguous**
Target Beliefs: (higher numbers = pro-choice)	Intuition Favors Pro-Life (*n* = 185)	Intuition Favors Pro-Choice (*n* = 186)	Pro-life (*n* = 189)	Pro-choice (*n* = 179)	Ambiguous (*n* = 186)
Pro-choice	3.74 (0.11)[a]	3.96 (0.11)[b]	1.83 (0.11)	5.76 (0.11)	3.55 (0.11)
Pro-life	3.24 (0.16)[a]	3.56 (0.15)[b]	2.02 (0.16)	5.95 (0.15)	3.63 (0.15)
Likelihood of being persuaded? (higher numbers = greater likelihood)					
Pro-choice	5.06 (0.13)	5.02 (0.13)	2.79 (0.13)[c]	5.68 (0.14)[d]	4.81 (0.13)
Pro-life	5.19 (0.19)	5.44 (0.18)	4.48 (0.19)[c]	3.53 (0.18)[d]	4.95 (0.19)
% sending note	18.92%	19.89%	12.17%	10.61%	13.98%

Note. Standard errors reported in parentheses. Superscripts with matching letters indicate that means significantly differ at $p < .05$.

Persuasion. We then conducted a 2 (Participant Beliefs: Pro-Choice vs. Pro-Life) X 5 (Assignment: Intuition Favors Pro-Life, Intuition Favors Pro-Choice, Unambiguously Pro-Life, Unambiguously Pro-Choice, Ambiguous) between-subjects ANOVA on the extent to which participants believed the target could be persuaded. We observed a significant main effect of Assignment, $F(4, 913) = 51.90$, $p < .001$, $\eta^2 = .163$, and no significant main effect of Participant Beliefs, $F(1, 913) = 0.16$, $p = .692$, $\eta^2 < .001$, which was qualified by a significant Participant

Beliefs X Assignment interaction, $F(4, 913) = 38.22$, $p < .001$, $\eta^2 = .120$. As shown in Table S7A, in the Intuition Favors Pro-Life condition, pro-choice ($M = 5.06$, $SE = 0.13$) and pro-life participants ($M = 5.19$, $SE = 0.19$) did not significantly differ in the extent to which they believed the target could be persuaded, $t(913) = -0.55$, $p = .580$, $d = -0.09$, 95% CI [-0.40, 0.22]. Similarly, in the Intuition Favors Pro-Choice condition, pro-choice ($M = 5.02$, $SE = 0.13$) and pro-life participants ($M = 5.44$, $SE = 0.18$) did not significantly differ in the extent to which they believed the target could be persuaded, $t(913) = -1.91$, $p = .056$, $d = -0.29$, 95% CI [-0.59, 0.01]. Instead, participants were significantly less likely to believe the target could be persuaded when the target unambiguously endorsed a viewpoint that diverged from their own. That is, when the target was unambiguously pro-life, participants who were themselves pro-choice ($M = 2.79$, $SE = 0.13$) were significantly less likely than participants who were pro-life ($M = 4.48$, $SE = 0.19$) to believe the target could be persuaded to adopt their view point, $t(913) = -7.50$, $p < .001$, $d = -1.17$, 95% CI [-1.49, -0.86]. And, when the target was unambiguously pro-choice, participants who were themselves pro-life ($M = 3.53$, $SE = 0.18$) were significantly less likely than participants who were pro-choice ($M = 5.68$, $SE = 0.14$) to believe the target could be persuaded to adopt their view point, $t(913) = -9.62$, $p < .001$, $d = -1.49$, 95% CI [-1.80, -1.18]. Pro-choice ($M = 4.81$, $SE = 0.13$) and pro-life participants ($M = 4.95$, $SE = 0.15$) did not significantly differ in the extent to which they believed the target could be persuaded when the target was ambiguously conflicted, $t(913) = -0.60$, $p = .546$, $d = -0.09$, 95% CI [-0.40, 0.21].

A planned contrast indicated that participants in the two combined Dual-Process conditions ($M = 5.13$, $SD = 1.35$) believed that the target could be persuaded significantly more so than those in the two combined unambiguous conditions ($M = 4.09$, $SD = 2.01$), $t(639.93) = 8.27$, $p < .001$, $d = 0.64$, 95% CI [0.46, 0.76], as well as those in the ambiguous condition ($M =$

4.86, *SD* = 1.25), *t*(395.48) = 2.38, *p* = .018, *d* = 0.21, 95% CI [0.03, 0.39]. And participants were in fact more likely to send the participant a note trying to persuade them in the combined two Dual-Process conditions (19.41%, *n* = 72) than those in the two combined unambiguous conditions (11.41%, *n* = 42), $\chi^2(1) = 9.05$, *N* = 739, *p* =.003, but not significantly more than those in the ambiguous condition (13.98%, *n* = 26), $\chi^2(1) = 2.52$, *N* = 557, *p* =.113 (see the supplementary file "notes to participant_abortion.doc" on our OSF page for further details regarding the actual notes that participants sent).

Robustness Checks

True preference subset analysis. We also asked participants a binary measure of whether they believed the target had a true preference, coded as 0 (*No*) or 1 (*Yes*). Importantly, the results were nearly identical when including only those who believed the target had a true preference (*n* = 541). That is, a 2 (Participant Beliefs: Pro-choice vs. Pro-life) X 5 (Assignment: Intuition Favors Pro-Life, Intuition Favors Pro-Choice, Unambiguously Pro-Life, Unambiguously Pro-Choice, Ambiguous) between-subjects ANOVA on Likert ratings of the target's beliefs revealed a significant Participant Beliefs X Assignment interaction, *F*(4, 531) = 3.31, *p* = .011, $\eta^2 = .001$. In the Intuition Favors Pro-Life condition, pro-choice participants (*M* = 3.48, *SE* = 0.18) were significantly more likely than pro-life participants (*M* = 2.66, *SE* = 0.24) to say that the target was authentically pro-choice, *t*(531) = 2.80, *p* = .005, *d* = 0.64, 95% CI [0.19, 1.10]. Similarly, in the Intuition Favors Pro-Choice condition, pro-choice participants (*M* = 3.87, *SE* = 0.19) were again significantly more likely than pro-life participants (*M* = 3.24, *SE* = 0.26) to say that the target was authentically pro-choice, *t*(531) = 1.99, *p* = .047, *d* = 0.49, 95% CI [0.005, 0.98]. Thus, in the two Dual-Process conditions, participants said that target authentically preferred the option that aligned with their normative beliefs, independent of whether it was

favored by intuition or deliberation. In contrast, there were no significant differences in ratings of the target's authentic preference when the target's stated position was unambiguous (only pro-life or pro-choice) or completely ambiguous (a conflict was stated, but the underlying mental processes was not defined), all *p*s > .28. There was also a significant main effect of assignment, $F(4, 531) = 240.38$, $p < .001$, $\eta^2 = .64$, but no significant main effect of the observer's own preference, $F(1, 531) = 1.63$, $p = .203$, $\eta^2 = .001$.

Continuous political moderator predicting true preference. We also conducted a linear OLS regression model predicting the target's true preference by participants' self-reported political orientation (1 = *extremely liberal* to 7 = *extremely conservative*), assignment (Intuition Favors Pro-Life, Intuition Favors Pro-Choice, Unambiguously Pro-Life, Unambiguously Pro-Choice, Ambiguous), as well as their interaction (see Table S7B), which revealed a significant main effect of assignment, $F(4, 915) = 249.88$ $p < .001$, $\eta^2 = .52$, and a non-significant main effect of political orientation, $F(1, 915) = 1.57$, $p = .210$, $\eta^2 = .001$, which were qualified by a significant political orientation X assignment interaction, $F(4, 915) = 4.43$, $p = .001$, $\eta^2 = .01$. To decompose the interaction, we contrasted the marginal means of the target's authentic preference by those high (+1 SD above the mean) vs. those low (-1 SD below the mean) of participants' political orientation within each assignment condition. In the Intuition Favors Pro-Life condition, although more liberal participants ($M = 3.68$, $SE = 0.12$) rated the target's authentic preference as more pro-choice compared to more conservative participants ($M = 3.48$, $SE = 0.13$), this difference was not statistically significant, $t(915) = 1.18$, $p = .238$, $d = 0.17$, 95% CI [-0.11, 0.45]. Similarly, in the Intuition Favors Pro-Choice condition, although more liberal participants ($M = 3.89$, $SE = 0.13$) rated the target's authentic preference as more pro-choice compared to more conservative participants ($M = 3.75$, $SE = 0.12$), this difference was not statistically

significant, $t(915) = 0.79$, $p = .431$, $d = 0.12$, 95% CI [-0.18, 0.41]. Instead, the effect of the interaction was mainly driven by those in the Unambiguously Pro-Life condition. That is, when the target was unambiguously pro-life, more liberal participants were significantly less likely to believe that the target's authentic preference was pro-choice ($M = 1.60$, $SE = 0.12$) compared to more conservative participants ($M = 2.23$, $SE = 0.13$), $t(915) = -3.53$, $p < 001$, $d = -0.53$, 95% CI [-0.82, -0.23]. There were no significant differences in the target's authentic preference by participants' political orientation for those in the unambiguously pro-choice condition, $t(915) = 0.03$, $p = .857$, or those in the ambiguous condition, $t(915) = 1.14$, $p = .255$ (see "truepreferences_supplemental_study7.R" on our OSF page).

Continuous political moderator predicting persuasion. We also conducted a linear OLS regression model predicting the extent to which participants believed the target could be persuaded by participants' self-reported political orientation (1 = *extremely liberal* to 7 = *extremely conservative*), assignment (Intuition Favors Pro-Life, Intuition Favors Pro-Choice, Unambiguously Pro-Life, Unambiguously Pro-Choice, Ambiguous), as well as their interaction (see Table S7B). This analysis revealed a significant main effect of assignment, $F(4, 913) = 50.03$, $p < .001$, $\eta^2 = .16$, and a non-significant main effect of political orientation, $F(1, 913) = 0.96$, $p = .328$, $\eta^2 = .001$, which were qualified by a significant political orientation X assignment interaction, $F(4, 913) = 30.14$, $p < .001$, $\eta^2 = .10$. To decompose the interaction, we contrasted the marginal means of the target's authentic preference by those high (+1 SD above the mean) vs. those low (-1 SD below the mean) of participants' political orientation within each assignment condition. Similar to the analysis with participant beliefs (pro-choice vs. pro-life) as a predictor, more liberal and more conservative participants did not differ in the extent to which they believed the target could be persuaded in the Intuition Favors Pro-Life condition, $t(913) =$

1.66, $p = .098$, in the Intuition Favors Pro-Choice condition, $t(913) = 1.148$, $p = .140$, or in the ambiguous condition, $t(913) = 1.05$, $p = .296$. Instead, when the target was unambiguously pro-life, more liberal participants were significantly less likely ($M = 2.67$, $SE = 0.15$) to believe that the target could be persuaded compared to more conservative participants ($M = 4.11$, $SE = 0.16$), $t(913) = -6.55$, $p < .001$, $d = -0.98$, 95% CI [-1.28, -0.66]. Conversely, when the target was unambiguously pro-choice, more liberal participants were significantly more likely ($M = 5.85$, $SE = 0.16$) to believe that the target could be persuaded compared to more conservative participants ($M = 4.00$, $SE = 0.14$), $t(913) = 8.52$, $p < .001$, $d = 1.26$, 95% CI [0.96, 1.55].

Although pro-life participants were significantly more likely to be conservative ($M = 5.70$, $SD = 1.63$) compared to those who were pro-choice ($M = 2.74$, $SD = 1.20$), $t(807.3) = 31.23$, $p < .001$, it is important to note that political orientation is an imperfect measure of abortion beliefs ($r = .68$), and thus, results here should be interpreted with caution as measurement noise may reduce the size of point estimates (Schmidt & Hunter, 1999).

Table S7B

Marginal Means of Condition and Participant Politics (Continuous) in Predicting Ratings of the Target's Authentic Beliefs (Supplemental Study 7)

	Dual Process		Unambiguous		Ambiguous
	Intuition Favors Pro-Life (n = 185)	Intuition Favors Pro-Choice (n = 186)	Pro-life (n = 189)	Pro-choice (n = 179)	Ambiguous (n = 186)
Target Beliefs: (higher numbers = pro-choice)					
More Conservative	3.48 (0.13)	3.75 (0.12)	2.23 (0.13) [a]	5.85 (0.12)	3.67 (0.12)
More Liberal	3.68 (0.12)	3.89 (0.13)	1.60 (0.12) [a]	5.82 (0.13)	3.47 (0.12)
Likelihood of being persuaded? (higher numbers = greater likelihood)					
More Conservative	5.28 (0.15)	5.32 (0.15)	4.11 (0.16) [b]	4.00 (0.15) [c]	4.97 (0.15)
More Liberal	4.93 (0.15)	4.99 (0.16)	2.67 (0.15) [b]	5.85 (0.16) [c]	4.75 (0.15)

Note. Standard errors reported in parentheses. Superscripts with matching letters indicate that means significantly differ at $p < .05$. "More Conservative" represents the predicted effect for participants who scored one standard deviation above the mean of political orientation (M = 5.79) and "More Liberal" represents the predicted effect for participants one standard deviation below the mean of political orientation (M = 1.69).

Supplemental Study 7 Discussion

When the target was conflicted, pro-life and pro-choice individuals were more likely to believe that a target authentically preferred the option that aligned with their own normative beliefs. In other words, when there's a broader, hierarchical arrangement of intuition versus deliberation that one can appeal to, individuals are more likely to believe that the target authentically prefers the option that they believe is "right". However, when the target was not conflicted, or the assignment of intuition versus deliberation to a specific outcome was not clearly specified, participants were less likely to say that a target authentically preferred the option that aligned with their own normative beliefs. This pattern of results is consistent with the notion that participants use the language of dual-process to justify their beliefs about others' preferences.

We also found a downstream consequence of this effect: When the target's intuition pulled in one direction and their deliberation in another, participants were more likely to believe that the target could be persuaded to adopt their viewpoint compared to when the target was not conflicted or when the target was conflicted, but dual processes were not specified.

Supplemental Study 8: Predicting Health Behavior

In a final study, we examined how the present effect may impact people's predictions about others' health behaviors. Participants were told about an individual who was conflicted about whether to get a COVID-19 vaccine: their deliberative (intuitive) thoughts favored vaccination, while their intuitive (deliberative) thoughts favored not getting the vaccine. Subsequently, participants were asked to predict whether the target would get a COVID-19

vaccine and to explain the reason for their prediction. The design and analysis plan were pre-registered (https://aspredicted.org/TYB_BJ6).

Method

Participants

We recruited 400 adults from Prolific Academic (N = 406, 68.9% female; M_{age} = 35.88, SD = 13.43). This gave us 95% power (with α = .05) to detect a small effect (η^2 = .031). We aimed to recruit roughly an equal number of pro-vaccine and anti-vaccine participants. Because in the U.S., vaccination attitudes correlate with political ideology (Pew Research Center, 2021), we recruited from pre-existing Democrat and Republican panels on Prolific Academic.

Procedure

Participants read a vignette featuring an individual who was having conflicting feelings about getting a COVID-19 vaccine. Participants were presented with one of two scenarios. In Scenario A, participants read the following: Chris is having conflicting feelings about getting a COVID-19 vaccine. When he listens to his gut intuition, he doesn't want to get the vaccine. When he thinks about it for a long time, he wants to get the vaccine.

In Scenario B, the assignment of the options to intuition versus deliberation was reversed: Chris is having conflicting feelings about getting a COVID-19 vaccine. When he listens to his gut intuition, he wants to get the vaccine. When he thinks about it for a long time, he doesn't want to get the vaccine.

All participants were then asked, "What do you think Chris will do?" and were presented with a seven-point Likert scale, ranging from 1 (*Definitely not get vaccine*) to 7 (*Definitely get vaccine*). Between-subjects, the order of the statements for each of the scenarios as well as the side of the Likert scale anchors were counterbalanced. We hypothesized that participants' own

vaccination preferences would guide their beliefs about what the target individual would do, regardless of whether their preference was informed by deliberation or intuition.

On the next page, participants were provided with the following prompt: "Please explain your response in the space provided below:" and were given a free-response text box in which they could write whatever they wished. We investigated whether, despite participants' own preferences informing their predictions about the targets, only a few participants would explain their reasoning as such.

At the end of the study, participants were asked to indicate whether they, personally, had or planned to get a vaccine (1=Yes, 2=No, 3=Undecided), who they voted for in the 2020 election (1=Joe Biden, 2=Donald Trump, 3=Other/Prefer not to say), as well as basic demographic information (i.e., age and gender).

Results

We first examined the distribution of participants' own vaccination status: 76.6% (n = 307) said they had gotten/planned to get the vaccine, 21.9% (n = 88) they would not get the vaccine, 1.5% (n = 6) said they were undecided, and 1.2% (n = 5) did not respond to the question. For our primary analysis we excluded the "undecided" and "unknown" participants (n = 11).

We then performed a 2 (Scenario: A vs. B) X 2 (Vaccination Status: Pro vs. Anti) X 2 (Counterbalanced Order) ANOVA (see Figure S8A), which indicated a significant main effect of Vaccination Status, $F(1, 387) = 39.59$, $p < .001$, $\eta^2 = .089$, and no main effect or interaction with Scenario type (both Fs < 1 and ps > .40). In other words, regardless of whether getting the vaccine was favored by intuition or by deliberation, participants who were themselves pro-vaccine reported that the target would get the vaccine (M = 4.51, SE = 0.08) significantly more

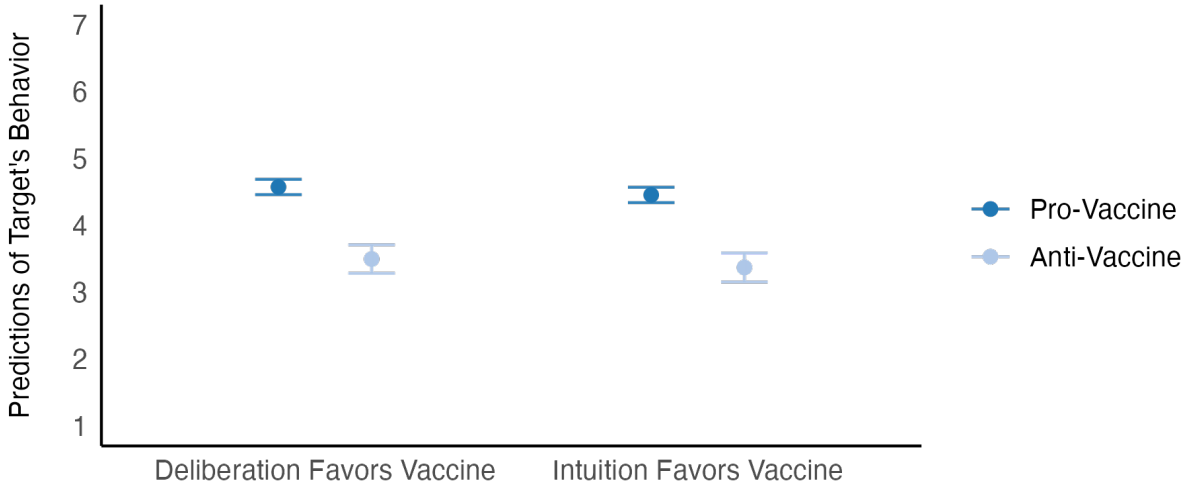
so than participants who were anti-vaccine (M = 3.43, SE = 0.15), d = 0.76, 95% CI [0.52, 1.01]. There was also a significant Scenario X Order interaction, $F(1, 387) = 13.01$, $p < .001$, $\eta^2 = .029$. Because order covaried with the placement of the Likert anchors, this result is most readily interpreted as a bias toward the right side of the scale. Importantly, this "side bias" did not significantly interact with the effect of participants' own vaccination status, $F(1, 387) = 0.07$, $p = .784$.

The results indicated a strong effect of participants' own vaccination preferences on their prediction about the target. Interestingly, when we examined the free-response data, we observed that participants rarely appealed to their own normative beliefs, and instead said things like "In the long run, we listen to our heads instead of our gut feeling because we think it's our best choice," or "Gut feelings usually win when deciding something."

Participants' explanations generally fell into one of 6 categories: 32% of responses captured that it is best to go with "gut", 23% of responses captured that it is best to go with deliberation; 9% of responses claimed that indecision results in getting the vaccine; 4% of responses claimed that indecision results in not getting the vaccine; 12% appealed to norms (e.g., "In the end he will get the vaccine because most of the people are vaccinated"), and 20% were classified as "other" (e.g., redescription of scenario, circular logic, nonsensical, blank (see Supplement file "coding of vaccine responses.xlsx"). Therefore, the majority of participants (55%) provided post-hoc explanations that explicitly appealed to the type of decision process (intuition versus deliberation) rather than to norms or their own normative beliefs.

Figure S8A

Participants Project Their Own Vaccine Preferences When Predicting Others' Behavior (Supplemental Study 8)



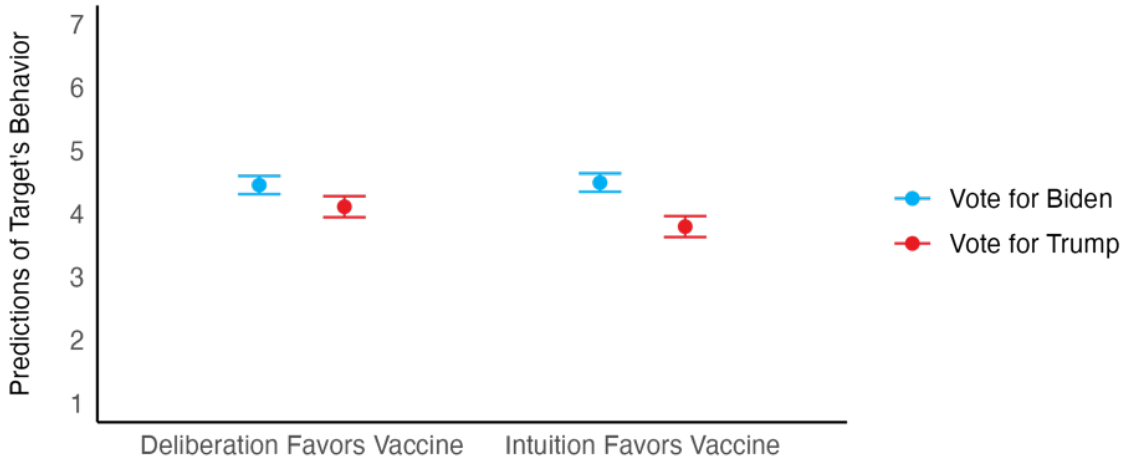
Note. Error bars represent the standard error of the mean. The y-axis ranges from 1 (*Definitely not get vaccine*) to 7 (*Definitely get vaccine*).

Voting preference as predictor. We also replicated the results with Voting Preference (Biden vs. Trump) as opposed to Vaccination Status (Pro vs. Anti) as a predictor of the target's behavior (see Figure S8B and "truepreferences_supplemental_study8.R"). That is, after removing participants who did not prefer to vote for either Biden or Trump (or preferred not to say, $n = 45$), a 2 (Scenario: A vs. B) X 2 (Voting Preference: Biden vs. Trump) X 2 (Counterbalanced Order) ANOVA revealed a main effect of Voting Preference, $F(1, 353) = 10.85$, $p = .001$, $\eta^2 = .028$, and no main effect or interaction with Scenario type (Fs < 1.30, ps > .06). Regardless of whether getting the vaccine was favored by intuition or by deliberation,

participants who intended to vote for Biden reported that the target would get the vaccine (M = 4.47, SE = 0.10) significantly more so than participants who intended to vote for Trump (M = 3.95, SE = 0.12), $t(353) = 3.33$, $p = .001$, $d = 0.36$, 95% CI [0.14, 0.57].

Figure S8B

Participants Project Their Own Voting Preferences When Predicting Others' Behavior (Supplemental Study 8)



Note. Error bars represent the standard error of the mean. The y-axis ranges from 1 (*Definitely not get vaccine*) to 7 (*Definitely get vaccine*).

Supplemental Study 8 Discussion

The results of Supplemental Study 8 demonstrate a strong effect of participants' own vaccination preferences on their predictions about whether a target would get the vaccine themselves. That is, independent of whether getting the vaccine was favored by intuition or deliberation, participants believed that the target would do what they themselves thought was the

right thing to do. Importantly, however, participants were not aware that their judgements were guided by their normative beliefs; when asked to explain their response, more than half doubled-down on the primacy of intuition or deliberation, and only one in ten appealed to norms.

References

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, *96*(5), 1029-1046. https://doi.org/10.1037/a0015141

Hull, C. L. (1932). The goal-gradient hypothesis and maze learning. *Psychological Review, 39*(1), 25–43. https://doi.org/10.1037/h0072640

Johnson, P. O., & Fay, L. C. (1950). The Johnson-Neyman technique, its theory and application. *Psychometrika*, *15*(4), 349–367. https://doi.org/10.1007/BF02288864

Kivetz, R., Urminsky, O., & Zheng, Y. (2006). The goal-gradient hypothesis resurrected: Purchase acceleration, illusionary goal progress, and customer retention. *Journal of Marketing Research*, *43*(1), 39-58. https://doi.org/10.1509/jmkr.43.1.39

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108*(3), 480–498. https://doi.org/10.1037/0033-2909.108.3.480

Newman, G. E., Bloom, P., & Knobe, J. (2014). Value judgments and the true self. *Personality and Social Psychology Bulletin*, *40*(2), 203–216. https://doi.org/10.1177/0146167213508791

Pew Research Center. (2021, June 2). *Most Americans favor the death penalty despite concerns about its administration*. https://www.pewresearch.org/politics/2021/06/02/most-americans-favor-the-death-penalty-despite-concerns-about-its-administration/

Pew Research Center. (2022, May 6). *America's abortion quandary*. https://www.pewresearch.org/religion/2022/05/06/americas-abortion-quandary/

Pew Research Center. (2024, June 6). *Cultural issues and the 2024 election*. https://www.pewresearch.org/politics/2024/06/06/cultural-issues-and-the-2024-election/

Pew Research Center. (2024, March 7). *How Americans view the coronavirus (COVID-19) vaccines amid declining levels of concern*. https://www.pewresearch.org/science/2024/03/07/how-americans-view-the-coronavirus-covid-19-vaccines-amid-declining-levels-of-concern/

Pirlott, A. G., & MacKinnon, D. P. (2016). Design approaches to experimental mediation. *Journal of experimental social psychology*, *66*, 29-38 https://doi.org/10.1016/j.jesp.2015.09.012

Schmidt, F. L., & Hunter, J. E. (1999). Theory testing and measurement error. *Intelligence*, *27*, 183-198. https://doi.org/10.1016/S0160-2896(99)00024-0

Spencer, S. J., Zanna, M. P., & Fong, G. T. (2005). Establishing a causal chain: why experiments are often more effective than mediational analyses in examining psychological processes. *Journal of Personality and Social Psychology*, *89*(6), 845-851. https://doi.org/10.1037/0022-3514.89.6.845

Spiller, S. A., Fitzsimons, G. J., Lynch Jr, J. G., & McClelland, G. H. (2013). Spotlights, floodlights, and the magic number zero: Simple effects tests in moderated regression. *Journal of Marketing Research*, *50*(2), 277-288.

https://doi.org/10.1509/jmr.12.0420

Strohminger, N., Knobe, J., & Newman, G. (2017). The true self: A psychological concept distinct from the self. *Perspectives on Psychological Science*, *12*(4), 551–560. https://doi.org/10.1177/1745691616689495

Yarkoni, T. (2022). The generalizability crisis. *Behavioral and Brain Sciences*, *45*, e1. https://doi.org/10.1017/S0140525X20001685