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

Abstract

Which mental process reveals one's authentic preference—deliberative reasoning or one's gut impulses? The existing literature offers conflicting answers to this question: Some research suggests that people generally see deliberation as more fundamental, while other work suggests that people see intuition as more fundamental. This paper argues that belief in a true self provides a unifying framework to explain when participants will attribute one's authentic preference to either System 1 or System 2. In line with predictions made by our theory, the results from four experiments (N=3,399 American adults) show that attributions about others' authentic preferences vary predictably across normative and non-normative contexts. Further, we show that the more participants report believing in a good true self, the more their judgments about others adhere to a predictable pattern; and, directly manipulating information about a target's true self changes people's judgments about a target's authentic preferences. By integrating theories of decision conflict and existing research on the true self, this work advances our understanding of how people reason about others' minds, revealing how lay theories about identity can systematically shape social prediction and judgment.

Statement of Limitations

The primary limitation of the current research is generalizability. Although efforts were made to ensure an equal sampling of political liberals and conservatives, the recruitment of North American adults through online participant pools limits the generalizability to other populations. We recommend future research replicate these results in more culturally diverse samples. Additionally, in the present studies, participants were exposed to hypothetical scenarios to maximize internal validity. This, however, limits our understanding of how these effects may

unfold during real, in-person conversations or among people in existing relationships. We recommend future research replicate the results in contexts involving real-world decision conflicts, especially as they occur naturally in daily life.

People frequently experience decision conflict in daily life (Schwartz, 2004). Although decision conflict can be understood in a variety of ways, one common distinction is between the "heart" and "head"—a person's intuition (System 1) pulls in one direction, while deliberation (System 2) pulls in another (e.g., Stanovich & West, 2000; Kahneman 2011; Kahneman & Frederick, 2005). In this paper, we seek to address a simple, yet fundamental question: when people reason about such conflicts between the heart and head, what lay theories (if any) guide their beliefs about what the person authentically prefers; do observers think that deliberation is more reflective of one's authentic preference, or do they instead believe that a target's true desires lie with their automatic, gut impulses?

To illustrate this question, consider a thought experiment proposed by the philosopher, Harry Frankfurt. Frankfurt (1971) asks readers to consider the plight of someone struggling with addiction. The person wants to quit using drugs, trying everything they can to avoid drugs and the urge to use. And yet, the intuitive pull to keep using drugs remains quite powerful. Which desire, Frankfurt asks, should we consider to be the person's authentic preference: Is it their deliberate desire to stop using drugs or their intuitive desire to continue?

Frankfurt (1971) frames this dilemma in terms of a hierarchy of preferences—the notion that some preferences are more central to a person's identity than others. In the example of addiction, many people might think about this situation as the person having a first- (or lower-) order preference to use drugs, but a second- (or higher-) order preference not to. As a result, it is intuitive to say something like, "the desire to use drugs overcame *him*", indicating that the speaker believes that the person's identity is associated with the deliberate decision to not use drugs.

But is this always the case? Do people generally believe that System 2 is most reflective of what the person authentically wants? While research on this exact question is scant, there are many findings which suggest that people are motivated to see themselves as competent, good decision-makers (Steele, 1988), and consequently, may believe one's authentic preference is best revealed through reason and deliberation. For example, studies find that individuals frequently endorse rational processing (e.g., "use a pros and cons list") over intuition (e.g., "follow their gut") as an effective strategy for making satisfying decisions (Schlegel et al., 2013). This is especially true for decisions that are objectively evaluable or complex (Inbar et al., 2010). Further, people rely less on intuition when they are motivated to be accurate (Payne et al., 1988); and when choices feel easier than anticipated, people may even complicate their decision-making to feel they have adequately thought about it (Schrift et al., 2011).

Other research, however, finds that people appear to more strongly identify with their intuition. For example, participants evaluating consumer goods report greater certainty in their decisions when using their "gut feeling" versus "rational analysis" (Maglio & Reich, 2018; 2020). Individuals perceive their intuitive thoughts as providing a great deal of self-insight (Morewedge et al., 2014). Most people believe they should avoid contradicting their initial instincts (Dane et al., 2012; Kruger et al., 2005; Miller & Taylor, 2002) and will avoid deliberative thinking when doing so protects their intuitive preferences (Woolley & Risen, 2018). Finally, some studies show that people are more likely to rely on intuition when they are concerned about appearing authentic (Oktar & Lombrozo, 2022).

Thus, the existing literature points to potentially conflicting views on which decision process—intuition versus deliberation—is perceived to be more reflective of one's authentic

preference. Some studies suggest people identify most with reason and deliberation, while others indicate that people identify most with intuition.

This paper aims to present a more integrative theory of how people reason about conflicts between System 1 and System 2. This question is important to our understanding of social perception because beliefs in authenticity may influence a variety of subsequent beliefs about the person (Rossignac-Milon et al., 2024; Sedikides & Schlegel, 2024). For example, beliefs about whether a person authentically wants to quit using drugs (or not) may inform predictions about how likely the person is to keep using drugs, how they will respond when tempted to use, or how receptive they might be to intervention (Christy & Schlegel, 2024; Collins et al., 2012; Furnham & Henley, 1988).

Here we propose that when people reflect on their own preferences and those of others, they appeal to the concept of a "true self," which in turn, affects which decision process—System 1 vs. System 2—is perceived to be more authentic. In the following sections, we first introduce the concept of the true self and its key attributes. We then develop our integrative theory explaining how true self beliefs inform people's reasoning about decision conflicts. Finally, we present the results of four studies testing our predictions across a variety of decision contexts.

The True Self and Decision-Making

Interestingly, Frankfurt (1971) proposed another version of his thought experiment: Frankfurt now asks readers to consider a doctor who treats patients struggling with addiction. After careful deliberation, the doctor decides that the only way she can truly understand her patients is if she tries drugs herself. Yet, despite this deliberate desire, the doctor has an

immediate, gut aversion to using drugs. Now, which desire, Frankfurt asks, should we consider to be the target's authentic preference?

In this case, many readers might be tempted to say that the doctor's authentic preference is to not take drugs. For example, we might hear someone say, "her instincts saved her", suggesting that her true identity aligns with her intuitive response rather than her deliberate reasoning. This creates an interesting contrast: in the first case (the person struggling with addiction), many view the target's authentic preference as reflected by deliberation over impulse, while in the second case (the concerned doctor), the target's authentic preference seems better reflected by intuition over deliberation.

This comparison reveals something important: our judgments about which mental process reflects someone's authentic preference may depend less on the process itself and more on which outcome we consider normatively preferable. But then, this raises a question: Why are beliefs about another person's preferences based on what is perceived to be normatively good?

We argue that this question can be answered by integrating insights from research on the "true self". Existing research suggests that people often appeal to a deeper, more essential notion of self (Johnson & Boyd, 1995; Koole & Kuhl, 2003; Newman et al., 2014). For example, in *A Christmas Carol*, Ebenezer Scrooge transforms from crotchety miser to joyful humanitarian. Although both instantiations of Scrooge are Scrooge, observers are drawn to the conclusion that the latter Scrooge has discovered his true nature. In other words, not all parts of the self are equally 'self-like'—some aspects of the self are perceived to be more authentic to a person's identity than others.

Indeed, existing research finds that the "true self" differs from ordinary conceptions of the self in several key respects (see Strohminger et al., 2017). And there are two key features of

people's conception of the true self, in particular, which we think help to resolve the apparent contradictions in how people reason about decision conflicts:

The True Self is Unconscious. First, the true self is believed to be largely unconscious and automatic. Existing research finds that traits are thought to be more authentic when rooted in emotions rather than cognitions (Haslam et al., 2004) or behavior (Andersen & Ross, 1984). Newman et al. (2014) demonstrated that people often posit the existence of a true self that differs from explicit thoughts. And several studies have found that more insight into the true self is gained via unconscious thoughts than deliberate reflection (Morewedge et al., 2014).

The automatic nature of the true self may explain why many studies–which typically examine neutral decisions such as selecting consumer products (cf. Maglio & Reich, 2018)–find that intuitive preferences are generally seen as more authentic. Thus, a true self account predicts that, in general, observers will see intuition as more authentic.

H1: Observers will generally see intuition as more authentic.

The True Self is Good. Second, and critically for understanding normatively-valenced decisions, the true self is perceived to be "good" (Bench et al., 2015; Newman et al., 2014; Newman et al., 2015; Strohminger et al., 2017). For example, people perceive a deadbeat dad who starts showing affection for his children as revealing his true self, while viewing a previously involved father who begins neglecting his kids as departing from his true self (Newman et al., 2014). Also, relatives of patients with dementia report that immoral changes yield the largest shifts in the person's essential nature (Strohminger & Nichols, 2014; also see Chen et al., 2016; Molouki & Bartels, 2017).

This belief that the true self is fundamentally good applies when reasoning about both oneself (e.g., Bench et al., 2015; Molouki & Bartels, 2017) and others (e.g., Bench et al., 2015; Newman et al., 2014). And the belief that the true self is fundamentally good is observed cross-culturally, even among misanthropic individuals with pessimistic beliefs about others (De Freitas et al., 2018). These normatively-valenced patterns now have been replicated by multiple labs (Finaly & Starmans, 2022; Lee & Feldman, 2022).

Why is the true self good? This is likely due to several factors (see the General Discussion for further elaboration). Consistent with principles of naïve realism (Ross & Ward, 1996), belief in a "good" true self may reflect a general tendency to see one's own moral values as inherent and fundamental (Baron & Spranca, 1997). Additionally, a good true self may reflect the tendency to mentally represent entities in terms of an ideal or essence (see Christy et al., 2019; Newman et al., 2015). Indeed, research has found people believe that the essence of many entities is drawn toward normatively good outcomes (De Freitas et al. 2017). Finally, belief in a good true self could reflect widespread religious beliefs. The notion of a deeper, unconscious part of identity that is morally good is consistent with many conceptions of soul across a variety of religious traditions (Böttigheimer & Widenka, 2023; Richert & Smith, 2012).

Regardless of the ultimate reason *why* the true self is perceived to be good, a key prediction of our theory is that normative valence—whether observers themselves think of the behavior as good or bad—will also have an effect on people's beliefs about authentic preferences. Specifically, for normatively-valenced decisions (like trying an illegal drug), people will think a person's authentic preference is whatever the observer believes is normatively good.

H2: For normative decisions, observers will generally think that a person's authentic preference is whatever the observer believes is normatively good.

In sum, these two attributes of the true self—its unconscious nature and perceived goodness—provide a framework for understanding when people will see intuition versus deliberation as more revealing of one's authentic preferences.

Normative Valence and Egocentric Projection

In fact, recent research has found some support for H1 and H2. Garrison et al. (2023) provided participants with hypothetical scenarios about self-control in which a target (self or other) was conflicted: their impulse favored one behavior (e.g., watching Netflix), but self-control favored another (e.g., studying for an exam). The results indicated that, overall, people think impulsive actions are more authentic than self-controlled actions, particularly when judging others. Importantly though, Garrison et al. (2023) also find that this effect is moderated by domain "positivity." When impulsive behaviors have a positive valence, being impulsive is seen as more authentic. However, when self-control is positive (e.g., not losing one's temper, or not flirting with a stranger when one is in a committed relationship), deliberative self-control is seen as more authentic. Although the authors operationalized domain valence via positivity and negativity, positivity also tended to correspond with normative valence in these studies (see p. 1655). Thus, these patterns are consistent with a true self account and hypotheses H1 and H2.[1]

[1] Garrison et al. (2023) exclusively examine cases of self-control (by design). Self-control and morality do overlap in cases of "moral self-control," when people override a selfish impulse or desire (e.g., cheating on a spouse) to enact a less selfish moral value or goal (e.g., remaining faithful; Hofmann et al, 2018). These tend to capture clear "right" versus "wrong" decisions. However, expanding to moral decision conflicts more broadly would include many "right" versus "right" dilemmas (Kidder, 2009), such as resolving the tension between loyalty and fairness in a whistleblowing dilemma (Waytz et al., 2013), or decisions surrounding controversial moral values (Navarick, 2013; Turiel et al., 1991), such as abortion, gun control, vaccination, and voting. Further, their manipulation of 'positivity'

However, despite consistent patterns in Garrison et al. (2023) (and Newman et al., 2014), existing work has not directly linked beliefs in the true self with how people reason about decision conflicts. Thus, alternative explanations remain. For example, perhaps people think good behaviors are more authentic simply because of base rates—most people generally want to engage in good behaviors (Sun et al., 2023). Or perhaps the effect stems from egocentric projection (Todd & Tamir, 2024). For example, because most participants see adultery as wrong (Pew Research Center, 2006), they say a target authentically prefers to avoid adultery. In other words, there may be an effect of positivity for reasons unrelated to the true self.

Therefore, the primary goal of our work is to "rule in" the true self as an important organizing framework for reasoning about decision conflict. In this respect, the present studies go beyond existing research in several important ways:

Direct Evidence for the True Self. Even though previous results are consistent with a true self account, there is no direct support for it as the mechanism by which people determine which preference is more authentic. Why is this important? There is substantial evidence supporting egocentric mentalizing—the tendency to use our own beliefs as a starting place for understanding others' beliefs (see Todd & Tamir, 2024 for review). Therefore, it could be that the "positivity" effect observed in previous research is simply another case of egocentric projection.

One contrasting prediction that the true self account (versus egocentric projection) makes has to do with the strength of observers' preferences. One could argue, for example, that the reason why observers think that a person authentically prefers the 'good' option is simply due to the strength of one's preference—observers have stronger preferences regarding normative

generally does seem to overlap with being a normatively correct option, but some instances were morally ambiguous (e.g., crying at work).

issues, and thus are more likely to project their preferences onto targets in these cases. In contrast, if the proposed effects are indeed driven by beliefs about the true self, then observers will generally see intuition, as well as the 'good' or 'right' option, as the more authentic preference regardless of the strength of one's own preferences.

H3: Observers will generally see intuition, as well as the normatively 'good' option, as the target's authentic preference regardless of the strength of observers' own preferences.

A second unique prediction of the true self account has to do with explicit beliefs in a true self. The more that participants believe in the notion of a good true self, the more likely they should be to say that deep down, others prefer normatively "good" options. In other words, explicit beliefs in a true self should moderate judgments about others' authentic preferences.

H4: Explicit beliefs in a true self should moderate judgments about others' authentic preferences.

A third unique prediction made by a true self account is that directly manipulating information about a target's true self should affect observers' beliefs about someone's authentic preferences. Specifically, providing information that a target's true self is inherently (good) bad should (increase) decrease the extent to which people believe that deep, down the target prefers the normatively-good option.

H5: Providing information that a target's true self is inherently (good) bad should (increase) decrease the extent to which people believe that that person authentically prefers the normatively-good option.

In sum, while existing research has found patterns that are consistent with a true self account, further work is needed to establish that indeed, beliefs in the true self provide an overarching framework for how people reason about the preferences of others. Doing so is important because it advances our understanding of how people reason about others' minds by showing how abstract beliefs about identity systematically influence social prediction and judgment.

Why do we need the true self as an explanatory framework? After all, there is a wealth of evidence supporting egocentric mentalizing, or the tendency to use our own beliefs and preferences as a starting place for understanding others' beliefs and preferences (see Todd & Tamir, 2024 for review). While egocentric projection undeniably plays a role in how people reason about others' preferences, we argue that the true self account offers an important (and complementary) perspective that helps to explain patterns of data that are difficult to account for through projection alone. For example, although projection is consistent with aligning one's own normative beliefs with a target, it would not explain why intuition should have primacy over deliberation, nor would it explain why people's beliefs in a good true self would be related to their social judgments.

Throughout this paper we examine predictions from both accounts—true self and egocentric projection—and find evidence that the true self explains predictions above-and-beyond egocentric projection, while finding evidence for both. We suggest that rather than

competing explanations, egocentric projection and true self beliefs represent distinct, but interacting mechanisms that people flexibly deploy to understand others' preferences. This theoretical integration not only helps explain existing findings, but generates novel predictions about when and how people will align their own preferences with predictions about others.

Beyond this theoretical contribution, this research has practical applications related to person perception and decision-making. People often make character judgments based on others' decision processes (e.g., Barasch et al., 2014; Critcher et al., 2013; Tetlock et al., 2000), which they then use to predict behavior: How will undecided voters ultimately vote? What does someone who lashes out in anger authentically desire? Will seemingly conflicted Americans ultimately support or prevent immigrant deportation? These findings may therefore also be linked to systematic mispredictions about consequential behavior like voting and determining stances on contentious social issues.

The Current Research

In Study 1, we build off Garrison et al. (2023) and show that overall, people think that authentic preferences are best reflected in intuition (H1). However, when the decision is normative, people think that a target's authentic preference is revealed by whichever mental process is pulling towards the normatively good option (H2). In Study 2, we examined how these effects interact with observers' strength of preference, showing that the strength of observers' preferences about various issues does not affect judgments of the target's authentic preferences (H3).

In Study 3, we examine how self-reported beliefs in a "good true self" moderate beliefs about others' authentic preferences, even when controlling for perceived target similarity (H4). And in Study 4, we directly manipulate information about a target's true self, showing it has a

causal effect on people's beliefs about the target's authentic preferences (H5). Table 1 outlines our hypotheses and results from Studies 1-4.

Transparency and Openness

In all studies, the target sample size was determined prior to data collection based on a sensitivity power analysis. We report all data exclusions (if any), all manipulations, and all measures in every study (Simmons et al., 2011). The data, code, and materials for all studies, as well as the preregistrations[2] and Supplemental Online Materials (SOM) are accessible on our Open Science Framework (OSF) repository at https://osf.io/aw439/?view_only=6e56a66f3d6a464898f5a14b6431bb52. Data were analyzed using RStudio, version 4.2.1 (Posit Team, 2024).

[2] Note that we deviate from our preregistered analyses in Studies 1-4 in the following ways: In Study 1, for our test of "Participants' Justifications", we average across scenarios as opposed to running eight individual chi-square tests of independence. In Studies 2-4, our dependent variable is the target's authentic preference as opposed to self-other discrepancy scores. We deviated from our preregistered analyses because, in hindsight, the analyses we report in the main text are a more accurate and clearer test of our hypotheses. However, our preregistered analyses were consistent with our predictions as well (see SOM; Study 1, pp. 3-5; Study 2, pp. 19-20; Study 3, p. 24; Study 4, p. 32).

Table 1

Overview of Experiments 1 – 4.

Study	Key Hypotheses	Key Results
1	H1: Observers will generally see intuition as more authentic.	In line with the true self being automatic and unconscious, intuition was perceived to be more fundamental in revealing a target's authentic preferences.
	H2: For normative decisions, observers will think that a person's authentic preference is whatever the observer believes is normatively good.	In line with the true self being fundamentally good, when intuition favored the 'bad' behavior, participants reported that deliberation was more authentic. However, when deliberation favored the 'bad' behavior, participants reported that intuition was more authentic.
2	H3: Observers will generally see intuition, as well as the normatively 'good' option, as the target's authentic preference regardless of the strength of observers' own preferences.	The strength of participants' own preferences did not change reasoning about authentic preferences.
3	H4: Explicit beliefs in a true self should moderate judgments about others' authentic preferences.	The more that participants believed that a conflicted voter had a "good true self," the more likely they were to believe that the target would vote for their preferred candidate, above and beyond perceptions of target similarity.
4	H5: Providing information that a target's true self is inherently bad should reduce the extent to which people believe that that person authentically prefers the normatively-good option.	When the target's true self is described as fundamentally bad, participants are less likely to say that target authentically prefers the normatively-good option.

Study 1: Normative vs. Neutral Scenarios

In Study 1, participants read about a target who was conflicted—their intuition favored one option, while deliberation favored another. Participants were then asked to indicate which option reflected the target's authentic preference.

In this study, following Garrison et al. (2023), we varied both the type of decision (normative vs. neutral) as well as the type of target (oneself vs. another person). By doing so, we were able to simultaneously test hypotheses H1 and H2. According to a true self account, we should observe the following: First, intuition should generally be seen as more reflective of one's authentic preference (H1); Second, in normative domains, whichever mental process is associated with the 'good' option should be seen as reflective of one's authentic preference (H2). Consistent with a true self account, we expected these patterns to hold for both self and others. Indeed, existing research shows similar patterns of beliefs for the self and others regarding a true self (see Strohminger et al., 2017).

Our studies also differed from previous research in other important ways: First, we manipulated the decision process (intuition vs. deliberation) using the exact same scenarios (e.g., deciding whether or not to steal something). This is an important departure from Garrison et al. (2023) who used different scenarios/contexts for good vs. bad outcomes—in other words, the effect of "positivity" in prior work may have resulted from different decision contexts rather than positivity per se. Second, Garrison et al. (2023) only examine cases of self-control, which can be normative in nature, but not always (Hofmann et al., 2018). Therefore, in Study 1 (as well as the other studies) we look at normative issues that extend beyond self-control.

Finally, we examined people's justifications. It may be that people appeal to normative values in guiding their judgments about others (e.g., "the target prefers X because it is the right

thing to do"). Or, it may be that they appeal to the primacy of a particular decision process ("the target prefers X because intuition is more revealing of one's authentic preference"). The latter would suggest a more elaborate process whereby observers' feel the need to rationalize their judgments by appealing to the primacy of intuition versus deliberation (see Kunda, 1990). In other words, even though people may be tempted to say that others share their preferences–especially when it comes to normative or moral issues (Cusimano & Lombrozo, 2021)–it is not clear what types of evidential requirements are necessary to justify or rationalize that position.

Method

Participants

We aimed to recruit 1200 adults on Prolific Academic in a pre-registered design (https://aspredicted.org/3RQ_FB6). A total of 1199 adults (44.7% female, 1.7% other, M_{age} = 38.8, *SD* = 12.8) took the survey. This gave us 95% power (with α = .05) to detect a small effect (*d* = 0.21).

Procedure

Participants were randomly assigned to one of sixteen between-subjects conditions in a 2 (Decision Type: Normative vs. Neutral) x 2 (Target: Self vs. Other) x 2 (Assignment: Deliberation Favors Action vs. Intuition Favors Action) x 2 (Option Counterbalancing: Deliberation Displayed First vs. Intuition Displayed First) design.

Each participant read four vignettes (order randomized) featuring a target who was deciding between two options (see Figure 1 for a sample vignette). Half of the participants evaluated four normative scenarios while the other half evaluated four neutral scenarios (see Appendix A for all vignettes). And half of participants evaluated scenarios where *another person* was the target, while the other half evaluated scenarios where *they* were the target.

Figure 1

Sample Vignette from Study 1

Suppose that someone is feeling fatigued at work. They are having conflicting feelings about taking an illegal drug that increases alertness.

Their "gut" feeling is to **not** take the drug.

When they think for a long time, their feeling is **to take** the drug.

What is the person's true preference in the decision? That is, what option do they truly prefer deep down?

The four normative scenarios each featured a decision between a normatively good option and a normatively bad option: taking an illegal a drug vs. not; stealing a sandwich vs. not; switching to a new supplier that uses child labor vs. not; and recommending an environmentally harmful product vs. not. The four neutral scenarios were designed to semantically and structurally mirror the normative scenarios, but lacked a normative component: having an additional cup of coffee vs. not; buying a sandwich vs. not; switching to a new supplier vs. not; and recommending a product vs. not.

We pretested the scenarios with a separate sample of Prolific participants (N = 100) and asked them to what extent the decision involved a conflict between right vs. wrong, on a scale from 1 = *Not at all* to 7 = *Very much so.* The results confirmed that participants were more likely to view the normative scenarios as involving a conflict between something that was right vs. wrong (M = 5.29, SD = 1.74) more so than the neutral scenarios (M = 3.66, SD = 1.17), $t(99) = 6.78$, $p < .001$, d = 0.68, 95% CI [0.46, 0.90] (see "truepreferences_study1_domainpretest" on our OSF page).

For each vignette, participants were asked, "What is [*the person's/your*] true preference in the decision? That is, what option do [*they/you*] truly prefer deep down?" to which they responded on a scale from 1 (*Definitely not [take the drug]*) to 7 (*Definitely [take the drug]*), anchored at 4 (*Neither/no true preference*).

Further, we manipulated which option was favored by the target's intuition and which option was favored by deliberation. For example, in the taking the drug vignette, for half of the participants, taking the drug was the intuitively-favored option (while not taking the drug was the deliberatively-favored option), while for the other half of participants, taking the drug was the deliberatively-favored option (while not taking the drug was the intuitively-favored option). Finally, we counterbalanced whether the intuitively-favored option or the deliberatively-favored option was presented first. The hypothesized results for H1 and H2 were unaffected by this factor and thus we collapsed across the option counterbalancing conditions.[3]

After indicating the target's authentic preference, participants were then asked, "Why did you provide that response?" and were given four responses to choose from: 1) Because in general, a person's true preference is best revealed by their automatic, gut instinct; 2) Because in general, a person's true preference is best revealed by their careful, deliberative thoughts; 3) Because it is

[3] We did observe a small but significant interaction between assignment and option counterbalancing, $F(1, 1183) = 4.07$, $p = .044$, $\eta_p^2 = .003$, 90% CI [.000, .011], 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, $t(1138) = 3.59$, $p < .001$, $d = 0.29$, 95% CI [0.13, 0.45]. On the other hand, intuition ($M = 3.31$, $SE = 0.09$) was not more indicative of authentic preferences than deliberation ($M = 3.22$, $SE = 0.09$) when intuition was ordered first, $t(1187) = 0.74$, $p = .457$, $d = 0.06$, 95% CI [-0.10, 0.22]. However, considering that the main results of the model remain unchanged when the option counterbalancing term is included in the model, the results reported here collapse across the option counterbalancing conditions (but see the Supplemental Online Materials (pp. 2-3) and supplementary file "truepreferences_study1_supplementary_additional.R" on our OSF page to examine the results with the option counterbalancing main effect and interaction terms included).

the right thing to do; and 4) Other (please specify).[4] Finally, in this and all subsequent studies, participants provided basic demographic information.

Results

We first averaged the responses from the four vignettes to create a single measure of authentic preferences where 1 = the preference to not engage in the behavior and 7 = the preference to engage in the behavior. We then conducted a 2 (Decision Type: Normative vs. Neutral) x 2 (Target: Self vs. Other) x 2 (Assignment: Deliberation Favors Action vs. Intuition Favors Action) ANOVA.[5]

We observed a main effect of assignment (Deliberation Favors Action vs. Intuition Favors Action), $F(1, 1191) = 9.34$, $p = .002$, with a small effect size, $\eta_p^2 = .008$, 90% CI [.002, .019]. 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" ($M = 3.44$, $SE = 0.61$) versus one's "head" ($M = 3.17$, $SE = 0.61$), $t(1191) = 3.06$, $p = .002$, $d = 0.18$, 95% CI [0.06, 0.29].

We also observed a significant main effect of decision type (Normative vs. Neutral), $F(1, 1191) = 249.09$, $p < .001$, here with a large effect size, $\eta_p^2 = .173$, 90% CI [.142, .205]. Participants reported that the target authentically preferred to engage in the behavior significantly

[4] In two additional studies (see Supplemental Studies S1A and S1B in the Supplemental Online Materials; pp. 8-10) we further investigated the role of people's normative beliefs in their lay theories about others' preferences without first providing participants with scenarios. Our results replicate what we find in the main text, which is that people do not readily endorse norms as the basis for determining authentic preferences.

[5] 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 (Target: Self vs. Other) x 2 (Assignment: Deliberation Favors Action vs. Intuition Favors Action) mixed OLS regression model predicting the target's authentic preference, with participants (N = 1182) and scenarios (N = 8) treated as random effects. Results were nearly identical to our main results. The write-up of these results is available in the Supplemental Online Materials for this (p. 6) and all other studies where this supplemental analysis was applicable (pp. 7-8, p. 14, pp. 19-23, and pp. 44-47).

more when the behavior was neutral (M = 3.98, SE = 0.61), than when the behavior was normatively "bad" (M = 2.63, SE = 0.61), $t(1191) = 15.78$, $p < .001$, $d = 0.91$, 95% CI [0.79, 1.03], regardless of whether the outcome was favored by intuition or deliberation.

Importantly, however, we did not observe a statistically significant interaction between target and assignment, $F(1, 1191) = 0.43$, $p = .511$, $\eta_p^2 < .001$, 90% CI [.000, .004], or between target, assignment, and decision type, $F(1, 1191) = 2.40$, $p = .121$, $\eta_p^2 = .002$, 90% CI [.000, .008][6]. That is, we did not find evidence that intuition is more reflective of one's authentic preference for others, but deliberation is more reflective of one's authentic preference for the self.

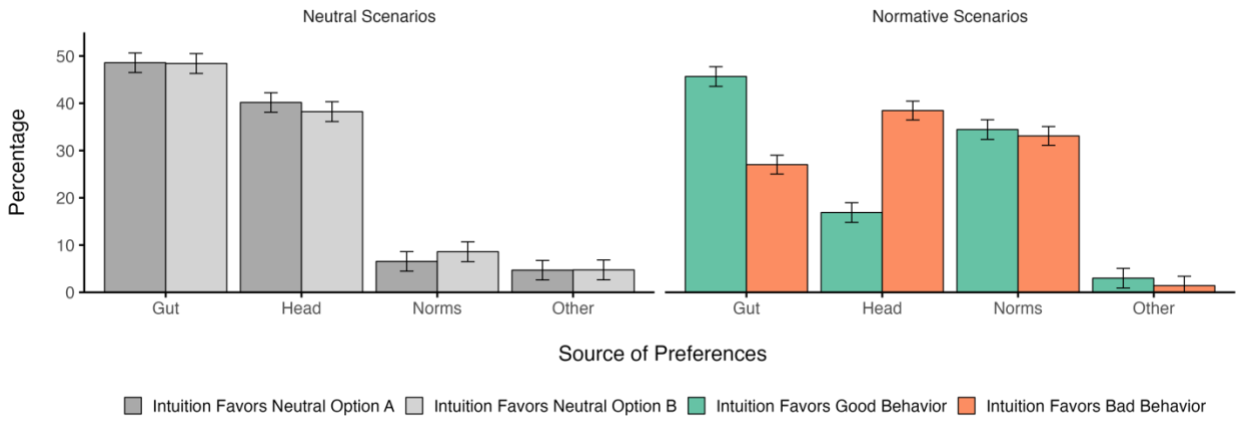
Participants' Justifications. For each participant, we calculated the percentage of scenarios for which they attributed the target's authentic preference to each of the four possible sources – intuition ("gut"), deliberation ("head"), the right thing to do ("norms"), or alternative reasons ("other"). For this analysis (and consistent with our pre-registration), we omitted n = 277 participants who indicated that the target did not have an authentic preference in each of the scenarios (i.e., by selecting "4 - Neither/no true preference"), leaving N = 922. We then ran a 4 (Source: Gut, Head, Norms, and Other) X 2 (Domain: Neutral vs. Normative) X 2 (Assignment: Deliberation Favors Action vs. Intuition Favors Action) mixed ANOVA, with Source as the within-subjects factor, and Domain and Assignment as between-subjects factors. Results revealed a significant three-way interaction between Source, Domain, and Assignment, F(3,

[6] Although not hypothesized and not central to our argument, we also observed a significant main effect of target (Self vs. Other), $F(1, 1191) = 10.33$, $p = .001$, $\eta_p^2 = .009$, 90% CI [.002, .019], which was qualified by a significant decision type X target interaction, $F(1, 1191) = 17.48$, $p < .001$, $\eta_p^2 = .014$, 90% CI [.005, .028]. 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. $M_{neutral} = 3.94$, $SE_{neutral} = 0.09$), $t(1191) = -8.21$, $p < .001$, $d = -0.67$, 95 % CI [-0.83, -0.51], this effect was stronger when the target was oneself ($M_{normative} = 2.31$, $SE_{normative} = 0.09$ vs. $M_{neutral} = 4.02$, $SE_{neutral} = 0.09$), $t(1191) = -14.10$, $p < .001$, $d = -1.15$, 95% CI [-1.32, -0.99]. No other main effects or interactions were observed.

2754) = 13.39, $p < .001$, $\eta_G^2 = .014$, 90% CI [.007, .022]. As shown in Figure 2, for the neutral scenarios, participants were significantly more likely to indicate the gut ($M = 48.50$, $SE = 1.47$) as opposed to the head ($M = 39.20$, $SE = 1.47$) as the source of one's authentic preferences, $t(3672) = 4.46$, $p < .001$, $d = 0.30$, 95% CI [0.17, 0.43]. Similarly, for the normative scenarios, participants were again significantly more likely to indicate the gut ($M = 36.35$, $SE = 1.44$) as opposed to the head ($M = 27.68$, $SE = 1.44$) as the source of one's authentic preferences, $t(3672) = 4.26$, $p < .001$, $d = 0.28$, 95% CI [0.15, 0.40].

Figure 2

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



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

Importantly, 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 ($M =$

38.46, *SE* = 1.99) as opposed to the gut (*M* = 27.02, *SE* = 1.99), *t*(3672) = 4.07, *p* < .001, *d* = 0.37, 95% CI [0.19, 0.54]. 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 (*M* = 45.67, *SE* = 2.08) as opposed to the head (*M* = 16.89, *SE* = 2.08), *t*(3672) = 9.77, *p* < .001, *d* = 0.92, 95% CI [0.73, 1.11][7].

Discussion

Supporting H1, participants were more likely to report that authentic preferences were best revealed by intuition. Supporting H2, when the decision conflict was normative in nature, participants were more likely to indicate that authentic preferences were best revealed by whatever mental process was pulling towards the 'good' outcome. Importantly, a similar pattern of results was observed when the target was oneself and when the target was another person. Examining participants' justifications also revealed that, in the normative scenarios, the majority appealed to the primacy of intuition or deliberation, rather than recognizing the role of normative valence.

Study 2: Addressing the Strength of Preference Alternative Explanation

In Study 1, and in line with the true self being automatic and unconscious, we found that intuition was perceived to be more fundamental in revealing a target's authentic preferences. However, in line with the true self being fundamentally good, we found that when intuition favored a 'bad' behavior, deliberation was perceived to be the source of one's authentic

[7] We also replicated this finding with individual differences in normative beliefs, but due to space limitations, we present the study in the SOM (pp. 38-47).

preferences. In other words, observers thought that a person's authentic preference was whatever they believed to be normatively good. Perhaps, however, this difference arises because observers simply have stronger preferences in normative domains. Indeed, egocentric projection is more likely to occur for stronger beliefs (Brenner & Bilgin, 2011; Marks & Miller, 1985) and whether people adjust away from their own perspective is dependent on the strength with which self-referential information is activated (Todd & Tamir, 2024). Therefore, it could be that the patterns in Study 1 are explained by well-established mechanisms of egocentric projection.

To test this, Study 2 used a similar design as Study 1, but we instead designed each normative 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). In addition, we measured the strength of observers own preferences. Consistent with a true self account (and H3), we predicted that observers will generally see intuition, as well as the normatively 'good' option, as the target's authentic preference regardless of the strength of observers' own preferences. This experiment was preregistered (https://aspredicted.org/mgj6-tpkk.pdf)[8].

Method

Participants

We recruited 800 Prolific participants (63.1% female, 0.7% other/non-binary; M_{age} = 41.29, SD = 13.48), with the goal of having 100 participants per cell. This gave us 95% power (with α = .05) to detect a small effect (d = 0.26). We aimed to recruit an approximately equal

[8] We acknowledge that there are two errors in our preregistration. Although we state that the purpose of Study 2 is to test if alignment is greater in normative vs. neutral scenarios, our main purpose is to test H3: That observers will generally see intuition, as well as the normatively 'good' option, as the target's authentic preference regardless of the strength of observers' own preferences. Second, we note that our labelling of assignment from "intuition favors option A vs. intuition favors option B" should be corrected to "deliberation favors option B vs. intuition favors option B".

number of liberal and conservative participants using pre-existing Democrat (n = 400) and Republican (n = 400) pre-screeners on Prolific.

Procedure

Participants were randomly assigned to one of eight between-subjects conditions in a 2 (Decision Type: Normative vs. Neutral) X 2 (Assignment: Deliberation Favors Option B vs. Intuition Favors Option B) X 2 (Option Counterbalancing: Deliberation Displayed First vs. Intuition Displayed First) between-subjects design. Similar to Study 1, each participant read four vignettes (order randomized) featuring a target who was deciding between two options (see Appendix B for all vignettes). For example, one vignette in the neutral condition read:

> Suppose that someone is deciding to adopt their first pet. They are deciding between a cat or a dog. Their "gut" feeling is to get a cat. When they think for a long time, their feeling is to get a dog.

The four normative scenarios each featured a decision between an option that liberals tend to view as "good" and an option that conservatives tend to view as "good" (Graham et al., 2009; Newman et al., 2014; see also Pew Research Center, 2022, 2023) (see SOM, pp. 11-18 for more information on how scenarios were pre-tested and selected). The four normative scenarios were gun control (supporting strict or lenient gun regulations), immigration (supporting open or closed borders), climate change (viewing it as a top or low priority), and abortion (being pro-choice vs. pro-life). The four neutral scenarios featured a decision between two options that were bipartisan but tended to have bimodal preferences: adopting a pet (cat vs. dog), buying a computer (Mac vs. PC), choosing an airplane seat (window vs. aisle), and watching a film (horror vs. romance).

For each vignette, participants were asked, "What does the person truly prefer, deep down?" to which they responded on a scale from 1 (*Strongly prefers [the cat]*) to 7 (*Strongly prefers [the dog]*), anchored at 4 (*Indifferent/No preference*). At the end of the experiment, participants were asked to "please tell us about your preferences" for each scenario, to which they responded on a scale from 1 (*Strongly prefer [cats]*) to 7 (*Strongly prefer [dogs]*), anchored at 4 (*Indifferent/No preference*). Finally, participants completed an open-ended question to assess "what shaped your decisions about the people you read about in the study", a few demographic items (i.e., age, gender), and their political orientation (1 = *Extremely liberal*, 4 = *Moderate: Middle of the Road*, 7 = *Extremely conservative*).

Results

First, we averaged participants' responses across the four vignettes they read to create a single measure of authentic preferences where 1 = the preference to engage in behavior A and 7 = the preference to engage in behavior B. Then, we conducted a 2 (Domain: Neutral vs. Normative) X 2 (Assignment: Deliberation Favors Option B vs. Intuition Favors Option B) ANOVA. Results revealed a small significant main effect of domain, $F(1, 796) = 4.67$, $p = .031$, $\eta_p^2 = .006$, 90% CI [.000, .018], and a large significant main effect of assignment, $F(1, 796) = 99.77$, $p < .001$, $\eta_p^2 = .111$, 90% CI [.079, .146], both of which were qualified by a significant domain X assignment interaction, $F(1, 796) = 11.74$, $p < .001$, $\eta_p^2 = .015$, 90% CI [.004, .031]. Supporting Hypothesis 1, in the neutral scenarios (e.g., choosing a cat or dog, Mac or a PC), participants were much more likely to say that the target's authentic preference was to engage in the behavior favored by intuition ($M = 4.85$, $SE = 0.11$) versus deliberation ($M = 3.40$, $SE = 0.11$), $t(796) = 9.50$, $p < .001$, $d = 0.95$, 95% CI [0.75, 1.15]. Similarly, in the normative scenarios (e.g., supporting open or closed borders, taking a pro-choice or pro-life stance),

participants were significantly more likely to say that the target's authentic preference was to engage in the behavior favored by intuition (M = 4.24, SE = 0.11) versus deliberation (M = 3.54, SE = 0.11), $t(796) = 4.62$, $p < .001$, but the effect was much smaller, d = 0.46, 95% CI [0.26, 0.66].

To test Hypothesis 2 (i.e., that observers will think that a person's authentic preference is whatever the observer believes is normatively good), 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. Results revealed a significant main effect of assignment, $F(1, 394) = 12.24$, $p < .001$, $\eta_p^2 = .033$, 90% CI [.010, .067], and a significant main effect of participant politics, $F(1, 394) = 16.37$, $p < .001$, $\eta_p^2 = .040$, 90% CI [.014, .076], but no significant interaction between assignment and participant politics, $F(1, 394) = 2.58$, $p = .109$, $\eta_p^2 = .007$, 90% CI [.000, .026]. Supporting Hypothesis 2, independent of whether System 1 or System 2 favored the behavior, Democrats (M = 4.21, SE = 0.06) were significantly more likely than Republicans (M = 3.85, SE = 0.06), to believe that the target authentically preferred the option that was 'good' for democrats, $t(394) = 4.04$, $p < .001$, d = 0.41, 95% CI [0.21, 0.60]. That is, participants were more likely to believe that the target authentically preferred the option that aligned with whatever they believed was normatively good or right.

We next examined whether this difference arises because observers simply have stronger preferences in normative (compared to neutral) domains (H3). To examine if participants were more likely to have stronger preferences in normative compared to neutral scenarios, we calculated the absolute differences between one's own preferences and the mid-point of the

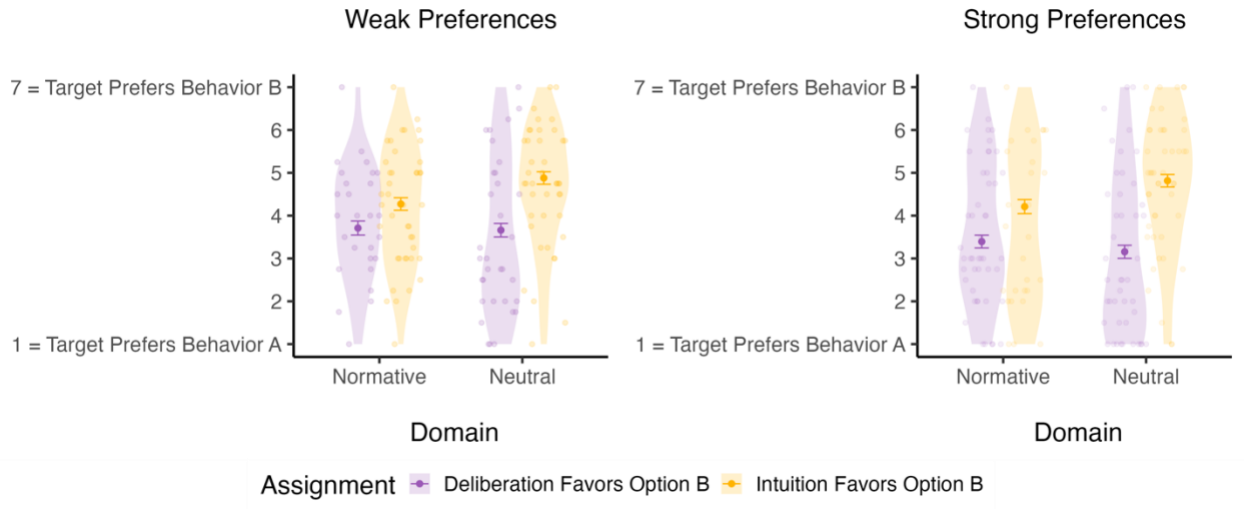
scale, averaged across the four scenarios. Importantly, participants did not significantly differ in their absolute differences from the midpoint of the scale (i.e., 4) in normative (M = 2.12, SD = 0.55) versus neutral (M = 2.13, SD = 0.58) scenarios, t(796.18) = -0.36, p = .719, d = -0.03, 95% CI [-0.16, 0.11].

Next, we included participants' own strength of preferences in the model. We conducted an OLS regression model predicting the target's authentic preferences by Domain (Neutral vs. Normative), Assignment (Deliberation Favors Option B vs. Intuition Favors Option B), the strength of participants' own preference, as well as all two- and three-way interactions.

Results (see Figure 3) revealed no significant main effect of participants' own strength of preference, b = -0.28, SE = 0.20, 95% CI [-0.67, 0.11], t(792) = -1.39, p = .164, as well as no significant interactions between strength of preference and domain, b = -0.17, SE = 0.28, 95% CI [-0.72, 0.39], t(792) = -0.59, p = .554, assignment, b = 0.22, SE = 0.28, 95% CI [-0.33, 0.78], t(792) = 0.80, p = .424, or between strength of preference, domain, and assignment, b = 0.17, SE = 0.39, 95% CI [-0.59, 0.93], t(792) = 0.43, p = .667. Similar to the results testing Hypothesis 1, we observed only a significant main effect of assignment, b = 0.69, SE = 0.15, 95% CI [0.39, 0.99], t(792) = 4.48, p < .001, and a significant interaction between domain and assignment, b = 0.75, SE = 0.22, 95% CI [0.32, 1.17], t(792) = 3.47, p < .001. In other words, participants believed that intuition, as well as the normatively 'good' option, was the target's authentic preference, regardless of the strength of participants' own preferences.

Figure 3

Authenticity Judgments Between Normative and Neutral Scenarios Persists Regardless of the Strength of Observers' Own Preferences (Study 2)



Note. Error bars represent standard error of the mean.

Discussion

According to a true self account, we should observe that people will generally see intuition, as well as the normatively 'good' option, as the target's authentic preference. Study 2 replicated these effects using a different paradigm (than Study 1) in which the normative scenarios differentially appealed to Republicans vs. Democrats. Further, supporting H3, this study showed that these effects are not accounted for by differences in the strength of observers' preferences across normative vs. neutral domains. In other words, it does not appear to be the case that normative domains evoke stronger preferences, which, in turn, drives stronger egocentric projection.

Study 3: Moderation by Belief in a Good True Self

In Study 3, we examined how people's explicit beliefs in a good true self affect their predictions about others' authentic preferences. The decision context was voting. Just prior to the 2024 Presidential Election (Harris vs. Trump), American participants (Democrats and Republicans) were told about a voter who was conflicted—their "head" favored one candidate, while their "gut" favored another. Given that the 2024 election was highly polarized and many prospective voters regarded it as a moral issue (Pew Research Center, 2024), we predicted that beliefs in a true self would moderate the extent to which participants' own political preferences influenced their beliefs about the target: Consistent with Hypothesis 4, the more that participants believe in a good true self, the more they should say that they target authentically prefers the "good" candidate deep down. Additionally, we asked about how similar observers felt to the target. Prior research suggests that perceived similarity is a strong predictor of whether or not an individual engages in egocentric projection (Ames, 2004; Todd & Tamir, 2024; Wang et al., 2023). The more similar a target is to oneself, the more people project their own values and preferences on to that person. However, if the true self mechanism is distinct from previously documented forms of projection, then we should observe separate effects of perceived similarity and beliefs in a good true self on predictions about the voter.

We suggest that this is a context in which beliefs about authentic preferences are ecologically-valid and important. Indeed, in 2024 there were many "undecided" voters (Pew Research Center, 2024), and large discrepancies between Democrat-leaning (e.g., Gooding, 2024) vs. Republican-leaning (e.g., Mallon, 2024) news outlets in their predictions about how those undecided voters were "truly leaning." Understanding the source of such discrepancies may improve accuracy in forecasting as well as potentially improve dialogue across politically divided

groups. This experiment was pre-registered and conducted in the context of the 2024 Presidential Election (https://aspredicted.org/t4gd-fn62.pdf)[9].

Method

Participants

We recruited 600 Prolific participants (59.1% female, 0.7% other/non-binary; M_{age} = 40.05, SD = 12.77), which gave us 95% power (with α = .05) to detect a small effect (η_p^2 = .034). We aimed to recruit an approximately equal number of Harris and Trump supporters using pre-existing Democrat (n = 301) and Republican (n = 299) pre-screeners on Prolific. This study was conducted on November 4, 2024, one day prior to the 2024 Presidential election.

Procedure

We randomly assigned participants to one of two conditions. In one condition, the voter's intuition favored Harris, while their deliberation favored Trump. In the other condition, the voter's intuition favored Trump, while their deliberation favored Harris. The order of these statements was counterbalanced within condition[10]. For example, in one condition, participants read:

> As the 2024 election approaches, Jesse is having conflicting feelings about the Presidential candidates.
>
> - When they listen to their gut intuition, they favor Donald Trump.
> - When they think about it for a long time, they favor Kamala Harris.

[9] Note that in our preregistration, we state that the purpose of Study 3 is to examine if true self beliefs are a stronger predictor of authentic preferences compared to perceived target similarity. However, our argument has since changed to acknowledge that true self beliefs are a complementary, and not competing, mechanism of projection.

[10] As indicated in our preregistration, we collapsed across the counterbalancing conditions (i.e., whether the head or gut was displayed first) in our primary analysis. Our results remain unchanged when including option counterbalancing in the model (see SOM, p. 24).

Immediately following the vignette, participants made predictions about who the target would ultimately vote for in the 2024 election on a 7-point scale ranging from 1 (*Definitely Trump*) to 7 (*Definitely Harris*) with a midpoint of 4 (*I don't know*). The order of the scale points was counterbalanced across participants, such that participants were randomly assigned to see both scales with *Definitely Trump* on the left or to see both scales with *Definitely Harris* on the left.

To assess perceptions of a good true self and perceived similarity, we used measures from prior research. Participants completed an item measuring the extent to which the voter had a good true self ("To what extent do you believe Jesse is a good person, deep down?"; Newman et al., 2015). They also completed an item assessing how similar they were to the voter ("To what extent do you believe Jesse is similar to you?"; Ames, 2004). Both items were counterbalanced across participants and measured on a scale from 1 (*Not at all*) to 7 (*Very*) anchored at 4 (*Somewhat*).

Finally, participants completed basic demographic information (i.e., age, gender) as well as three measures of their political orientation and voting preferences. Our main measure of voting preferences was a continuous scale asking participants who they plan to vote for in the upcoming election on a scale from 1 (*Definitely Trump*) to 7 (*Definitely Harris*) anchored at 4 (*I don't know*). A second forced-choice measure also asked about voting preferences and was used to exclude participants who were undecided or unwilling to report their preferred candidate. We also measured participants' political orientation on a scale from 1 (*Extremely liberal*) to 7 (*Extremely conservative*) anchored at 4 (*Moderate: Middle of the Road).* This measure was used as a robustness check in Supplementary Analyses (see pp. 25-31).

Results

As indicated in our preregistration, we excluded 36 participants who indicated "Prefer Not to Say / Undecided" regarding their own voting preference. This left 564 participants, who

were split by their intention to vote for Kamala Harris (N = 300) versus Donald Trump (N = 264).

We then conducted an OLS regression model predicting who the target would vote for by the main effect of participants' own voting preferences, condition, true self beliefs, perceived similarity, as well as the interactions between participant preferences and condition, participant preferences and true self beliefs, and participant preferences and target similarity.

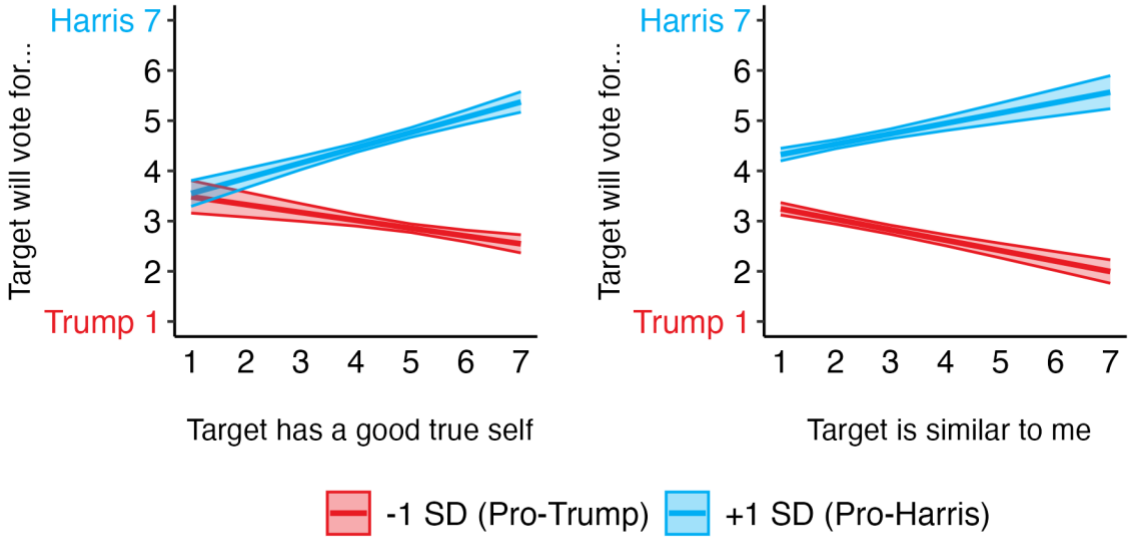
Controlling for perceived similarity to the target, we observed a main effect of participants' voting preferences, b = -0.29, SE = 0.09, 95% CI [-0.46, -0.12], $t(554) = -3.30$, p = .001, η_p^2 = .257, 90% CI [.208, .306] and the predicted interaction between participants' voting preferences and belief in the target's good true self, b = 0.08, SE = 0.02, 95% CI [0.04, 0.12], $t(554) = 4.36$, $p < .001$, η_p^2 = .082, 90% CI [.049, .121]. As shown in Figure 4, the more that Harris supporters believed the target had a good true self, the more that they thought the target would vote for Kamala Harris, b = 0.30, SE = 0.07, 95% CI [0.16, 0.44], $t(554) = 4.29$, $p < .001$. Conversely, the more that Trump supporters believed the voter had a good true self, the more that they thought the target would vote for Donald Trump, b = -0.16, SE = 0.08, 95% CI [-0.31, -0.002], $t(554) = 1.99$, p = .047.

Supporting previous research finding that perceived similarity predicts egocentric projection (Ames, 2004; Todd & Tamir, 2024; Wang et al., 2023), and in line with our theorizing that true self beliefs are a complementary mechanism to previously established predictors of egocentric projection, we also observed a significant interaction between participants' voting preferences and participants' perceived similarity to the target (see Figure 4), b = 0.07, SE = 0.01, 95% CI [0.04, 0.10], $t(554) = 4.95$, $p < .001$, η_p^2 = .043, 90% CI [.020, .074]. The more that Harris supporters perceived the target to be similar to them, the more that they thought the target

would vote for Kamala Harris, $b = 0.21$, $SE = 0.07$, 95% CI [0.07, 0.34], $t(554) = 3.05$, $p < .001$. Conversely, the more that Trump supporters perceived the target to be similar to them, the more that they thought the target would vote for Donald Trump, $b = -0.21$, $SE = 0.05$, 95% CI [-0.31, -0.11], $t(554) = -4.16$, $p = .002$.

Figure 4

Belief in the Target's Good True Self and Perceived Similarity Predict Voting Judgements (Study 3)



Note. Error bands represent standard errors.

Discussion

In Study 3, we examined how people's explicit beliefs in a good true self affected their predictions about who a conflicted target would ultimately vote for in the 2024 Presidential election. In line with our theorizing (H4), we observed that the more participants believed that the

conflicted voter had a good true self, the more likely they were to believe that the voter would vote for the "good" candidate.

Note that we also replicated these results in an earlier study that was conducted just prior to the 2020 U.S. Presidential Election (Biden vs. Trump). In this study, belief in the true self was measured as an individual difference (e.g., "People have a "true" self that captures who they really are") (see SOM pp. 48-56).

Study 4: Manipulating the True Self

The aim of Study 4 was to test H5. Specifically, we examined whether manipulating information about a target's true self would affect observers' judgments about the target's authentic preference. The previous studies built on the logic that people generally believe that deep down, others are "good"—it just happens that what people define as good varies from one person to the next. This predicts, however, that providing information that the target is not good deep down, should in turn reduce the extent to which people think the target authentically prefers normatively good outcomes.

To test this hypothesis, we examined another controversial and consequential domain: abortion. We described an individual who was conflicted in their beliefs—their head favored one stance on abortion, while their gut favored another. However, between-subjects, we also manipulated additional information about the target—i.e., that deep down they were good, or deep down they were bad. We predicted that there should only be an alignment between observers' own stance on abortion and their predictions about the target when they believed that "deep down" the target was good. This experiment was pre-registered (https://aspredicted.org/9n87-7978.pdf)

Method

Participants

We recruited 800 Prolific participants (49.9% female, 0.6% other/non-binary; M_{age} = 36.68, SD = 13.84), which gave us 95% power (with α = .05) to detect a small effect (d = 0.25). A total of 799 participants completed the dependent measures. We pre-screened and recruited an equal number of pro-choice (n = 399) and pro-life (n = 400) participants.

Procedure

Participants were randomly assigned to one of eight between-subjects conditions in a 2 (Assignment: Intuition Favors Pro-Life vs. Intuition Favors Pro-Choice) X 2 (Target True Self: Good vs. Bad) X 2 (Option Counterbalancing: Pro-Choice Option Displayed First vs. Pro-Life Option Displayed First) between-subjects design.

Participants first read about a target, named John, who was having conflicting feelings regarding his stance on abortion. For half of the participants, John's intuition pulled towards being pro-life (i.e., that "it's important to preserve every potential life") but his deliberation pulled towards being pro-choice (i.e., that "personal choice should take precedence, and women should be able to decide for themselves"). For the other half of participants, John's intuition pulled towards being pro-choice, but his deliberation pulled towards being pro-life. The order of these statements was counterbalanced within condition.[11]

To manipulate true self beliefs, participants were then randomly assigned to read one of two snippets about John (adapted from Newman et al., 2015; words in parentheses represent phrasing in the 'bad' true self condition):

> Ever since John was born, it was clear that there was something distinctive about his personality. He sometimes did bad (good) things to other people, but deep down in his very essence, he was a fundamentally good (bad) person. At the very core of his being, he has a profound compassion (no compassion) for other people and a genuine concern (no concern at all) about their well-being. He seems to be fundamentally driven toward a life of good

[11] As indicated in our preregistration, we collapsed across the counterbalancing conditions (i.e., whether the head or gut was displayed first) in our primary analysis. Our results remain unchanged when including option counterbalancing in the model (see SOM, p. 32).

(bad).

Immediately following the vignette, participants were first asked, "Deep down, does John have a true preference regarding abortion" and indicated their response on a binary Yes/No item. Then they indicated what John's true preference was on a scale from 1 (*Pro-Life*) to 7 (*Pro-Choice*) anchored at 4 (*Indifferent/No Preference*). Participants also indicated their own preferences on the same scale. The order of the scale points was counterbalanced across participants, such that participants were randomly assigned to see both scales with *Pro-Life* on the left or to see both scales with *Pro-Choice* on the left.

Finally, participants completed basic demographic information (i.e., age, gender) as well as their political orientation on a scale from 1 (*Extremely liberal*) to 7 (*Extremely conservative*) anchored at 4 (*Moderate: Middle of the Road).* In an open-ended manner, they were also asked what they believed the purpose of the study was and to provide any other additional comments.

Results

Manipulation Check. As a manipulation check, participants indicated whether deep down, John's true self was 1 (*Fundamentally bad*) to 7 (*Fundamentally good*). Indeed, participants perceived the target's true self to be more fundamentally good in the good true self condition ($M = 5.78$, $SD = 1.45$) compared to those in the bad true self condition ($M = 2.43$, $SD = 1.84$), $t(758.42) = 28.56$, $p < .001$, $d = 2.02$, 95% CI [1.85, 2.19].

Main Analysis. For our main analysis, we 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 (Pre-screened Participant Beliefs: Pro-Choice vs. Pro-Life) between-subjects ANOVA on the target's authentic preferences (1 = *Pro-Life* to 7 = *Pro-Choice*)[12]. Results revealed a

[12] As indicated in our preregistration, we also conducted this analysis removing those who stated that the target did not have a true preference. Importantly, the results remain the same when removing these participants (see SOM p.

significant main effect of participants' abortion beliefs, $F(1, 792) = 5.13$, $p = .024$, $\eta_p^2 = .006$, 90% CI [.000, .018], as well as a significant interaction between participants' abortion beliefs and the target's true self, $F(1, 792) = 20.28$, $p < .001$, $\eta_p^2 = .025$, 90% CI [.010, .045].

As shown in Figure 5, simple effects revealed that pro-choice participants were significantly more likely to believe that the target authentically preferred to be pro-choice when the target was described as having a good ($M = 4.26$, $SE = 0.14$) as opposed to a bad ($M = 3.87$, $SE = 0.14$) true self, $t(792) = 1.99$, $p = .047$, $d = 0.20$, 95% CI [0.002, 0.40]. Similarly, pro-life participants were significantly more likely to believe that the target authentically preferred to be pro-life when the target was described as having a good ($M = 3.32$, $SE = 0.14$) as opposed to a bad ($M = 4.18$, $SE = 0.14$) true self, $t(792) = -4.39$, $p < .001$, $d = -0.44$, 95% CI [-0.64, -0.24]. There were no other significant main effects or interactions (Fs < 2.92, ps > .08). 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.

34). We also conducted this analysis with self-other discrepancy (i.e., the absolute difference between the target's abortion preference and the observer's own abortion preference) as the dependent variable (SOM p. 32), as well as participants self-reported political orientation and abortion beliefs (continuously measured) as a predictor. Results were consistent with the notion that observers aligned their preferences with the target's preferences only when the target was described as having a good true self (see SOM pp. 32-37).

Figure 5

People Are More Likely to Align Their Beliefs With Another's Preferences When Others Have a Good (vs. Bad) True Self.

Note. Error bars represent standard errors.

Discussion

Our theory predicts that observers align their beliefs about a target's authentic preferences based on the belief that the target is good, deep down. Therefore, presenting information that the target is not good should disrupt that alignment (H5). Consistent with this prediction, pro-choice participants were more likely to believe that the target was pro-choice, and pro-life participants were more likely to believe that the target was pro-life, only when the target had a good (versus) bad true self. In other words, observers were more likely to align their beliefs with another's preferences only when the target was believed to have a good true self.

General Discussion

The present studies were motivated by a basic question: When people experience decision conflict, which mental process is seen as more authentic: Do people favor the products of reflection and deliberative thought, or do they instead believe that a target's authentic preference lies with their automatic, gut impulses? The answer to this question, we argue, is rooted in people's beliefs in a true self, which we unpacked in five hypotheses, all of which were supported empirically.

Hypothesis 1 was that people will generally tend to believe that a target's authentic preference lies with intuition. We found support for this prediction in Studies 1 and 2. Hypothesis 2 held that for normative domains, people will say that the target authentically prefers the "good" option (according to their own definition of right vs. wrong). We found support for this prediction in Studies 1-4.

In Study 1 we found that these patterns arise both when people are reasoning about themselves and another person, which is consistent with research that the true self is believed to be good both for oneself and for others (Bench et al., 2015; De Freitas & Cikara, 2018; Heiphetz et al., 2017). Our next set of hypotheses tested unique predictions made by a true self account. There is substantial evidence supporting egocentric mentalizing—the tendency to use our own preferences as a starting place for understanding others' beliefs and preferences (see Todd & Tamir, 2024 for review). While egocentric projection undeniably plays a role in how people reason about others' preferences, we argued that the true self account offers an important (and complementary) perspective that helps to explain patterns of data that are difficult to account for through projection alone.

Hypothesis 3 proposed that the strength of participants' own preferences will not change reasoning about authentic preferences–we found support for this in Study 2. Further, supporting

Hypothesis 4, in Study 3, we found that beliefs in a good true self predict beliefs about a target's authentic preferences above-and-beyond perceived similarity (cf. Ames, 2004; Wang et al., 2023). Finally, supporting Hypothesis 5, Study 4 showed that manipulating information about a target's true self affects observers' beliefs about that target's authentic preferences—if the target is believed to be bad deep down, observers no longer expect them to authentically prefer the good outcome.

Together, this empirical package contributes to the existing literature in several ways: First, we make novel connections between two theories to address new empirical questions. Broadly, we show that the notion of a good true self provides a more general framework for making predictions about when a given decision process will be perceived as reflective of one's authentic preference. Direct evidence for the link between theories of decision conflict and beliefs in the true self is (to our knowledge) novel to this paper.

Second, our work reconciles seemingly conflicting findings regarding whether intuition or deliberation is perceived as more authentic. One line of research has tended to look at decision conflict from the perspective of norms—i.e., when someone is conflicted, which outcome or type of reasoning do people think the person *should* go with. In general, this research finds that people see deliberation as better suited to more difficult or consequential decisions (Inbar et al., 2010; Pachur & Spaar, 2015). Interestingly, however, when researchers have simply asked people what is most revealing of what they and others prefer, people tend to overwhelmingly rely on their intuition when making decisions (Maglio & Reich, 2020; Morewedge et al., 2014; Oktar & Lombrozo, 2022).

Our research contributes to this literature by conceptualizing a moderator that explains conflicting predictions in the literature. Specifically, we identify an important and novel source

of heterogeneity in people's beliefs about whether intuition vs. deliberation is more authentic: individual differences in one's normative beliefs. Although work by Garrison et al. (2023) finds that the 'positivity' of the domain moderates whether observers think that one's impulse or self-control is more authentic, we extend this work to show that for the exact same scenarios, two people can come to opposite conclusions regarding what a conflicted target authentically prefers.

Finally, these findings may help to generate novel predictions and explain other forms of alignment effects found in the literature. For example, past research has found that liberal and conservative Christians each think Jesus Christ would hold their (mutually incompatible) political views (Ross et al., 2012). Rogers et al. (2017) similarly demonstrate that individuals believe that others will come to share their own point of view in the future (i.e., the *belief in a favorable-future* effect). It may be that the notion of a "good true self" plays an important role in these phenomena as well. For example, religious people may believe that Jesus would agree with them (Ross et al., 2012) precisely because they believe that Jesus has a "good true self."

We suggest that egocentric projection and beliefs in a good true self are not competing mechanisms, but rather complementary ones. What our data show is that egocentric projection and beliefs in the true self may operate in tandem, such that people project their own beliefs about what is "good" onto other people's true selves. This in turn allows people to assume that others do indeed share their preferences "deep down," even when those targets are conflicted. This theoretical integration not only helps explain existing findings, but generates novel predictions about when and how people will align their own preferences with predictions about others. Additionally, this insight is important because it suggests potential interventions. For example, efforts to debunk the notion of a true self may, to some degree, limit the extent to which people believe that others share the same viewpoint as them on consequential issues.

Is Dual Process Necessary?

In the current studies we examine conflicts between System 1 vs. System 2. However, one might wonder whether the same patterns would arise even if System 1 vs. 2 were never mentioned or tied to specific outcomes. In other words, how important is dual process to the patterns we document here?

In addition to the four experiments we report in the main text, we conducted an additional four supplemental studies (see SOM, pp. 38-75). In two of those supplemental studies (Supplemental Studies 6 and 7, pp. 48-68), we included conditions in which we did not assign the decision conflict to either System 1 vs. System 2 and instead said simply that "part of him prefers option A, while another part prefers option B" or more broadly that their "head was telling them one thing but their gut was telling them another" without assigning the decision outcome to System 1 or System 2.

In short, dual process does seem to matter somewhat to whether or not participants say that the target prefers the "good" outcome, deep down. In Supplemental Study 7 (pp. 57-68), participants are less likely to show reliable differences based on their own normative beliefs when the decision outcomes are not assigned to System 1 versus System 2. However, it remains unclear why this is the case.

One explanation might be related to Kunda's theorizing about boundary conditions and a "reasonableness constraint." Kunda (1990) suggested that an individual's tendency to engage in motivated reasoning is constrained by the extent to which one is able to marshal evidence in support of their position: "People motivated to arrive at a particular conclusion attempt to be rational and to construct a justification of their desired conclusion that would persuade a dispassionate observer. They draw the desired conclusion only if they can muster up the

evidence necessary to support it" (pp. 482-483). In other words, motivated reasoning is subject to "reasonableness" constraints (Boiney et al., 1997; Kunda & Sanitioso, 1989; Snyder et al., 1979; Woolley et al., 2021). Thus, even though people may be tempted to say that others share their preferences–especially when it comes to normative or moral issues—they may be less likely to do so if they cannot appeal to a justification beyond their own preferences.

This suggests that a dual process distinction may be just one instance of a broader phenomenon. Other conceptual dichotomies—such as nature vs. nurture—might produce similar effects by appealing to notions of a good true self. For example, in nature vs. nurture conflicts, people might attribute a normatively-good outcome option to someone's "natural tendencies" rather than "learned behaviors." These alternative frameworks represent promising avenues for future research.

Why is the True Self Good?

Another question is why people perceive the true self to be fundamentally good in the first place. As outlined in the Introduction, previous work has considered several explanations (De Freitas et al., 2017; De Freitas et al., 2018, Newman et al., 2014; Newman et al., 2015). One explanation is psychological essentialism. Under this view, beliefs in the true self arise because of a more general belief that entities have a true, essential nature (Gelman, 2003; Keil, 1989; Newman & Keil, 2008; Newman et al., 2015). A second explanation (which is compatible with essentialism) has to do with the notion that people may have an implicit motivation to see their own values as stable, essential aspects of the world. Indeed, past research has documented several instances in which people are motivated to endorse abstract values that do not directly benefit them personally (e.g., Jost et al., 2004). Therefore, it may be very difficult for people to view behaviors that they regard as immoral as fundamental and unchanging.

While the present studies were not designed to differentiate between these various explanations, the supplemental studies do lend credence to the latter "motivated" explanation of the good-true-self effect. If participants' beliefs were driven only by a cognitive mechanism, such as psychological essentialism, then we should have observed the same pattern of results regardless of whether the various decision outcomes were assigned to different mental processes. However, as we discuss above, when the outcomes were not assigned to intuition vs. deliberation and the person was simply described as "conflicted," participants were less likely to say that deep down, the target preferred the same option as them (see Supplemental Study 7, pp. 57-68). This suggests that the more general bias to see the true self as good might be importantly linked to motivated reasoning, and the ability or not to justify that position to others (cf. Kunda, 1990).

Limitations and Future Directions

Despite the advances made by the current studies, there are several limitations that may shape the conclusions drawn from the current work (see Table 2). We recommend that future research replicate the findings in culturally-diverse samples, in real world contexts, and link predictions of other's behavior to their actual behavior.

Future research may also examine how other individual differences may interact with the current effect. For example, do these effects emerge among individuals who view others negatively? Some research has found that even misanthropic individuals believe that the true self is fundamentally good (De Freitas et al., 2018), and thus, we predict that this effect might occur for misanthropes as well.

Conclusions

Which decision process is seen as more authentic–intuition or deliberation? The present studies demonstrate why the answer to this question is more complicated than it may initially

appear. Observers ground their beliefs about others' authentic preferences in a more elaborate theory of mind which invokes a notion of a good true self. Because of this, determining which mental process is seen as more fundamental changes depending on the values of the observer and the extent to which they believe a target has a good true self.

Table 2

Table of limitations across studies

	Limitation	Future research
External Validity		
Western/convenience samples	All participants were recruited through online samples. Although efforts were made to ensure an equal balance of political liberals and conservatives, inferences can only be made to a North American population.	While evidence has emerged to suggest that the belief in a good true self is consistent across other cultures (De Freitas et al., 2018), future research can replicate the findings with a representative sample as well as with other populations.
Ecological validity	Due to the hypothetical nature of the scenarios, participants were made aware of the internal conflict of the target. However, unless explicitly stated, people may not know about the content of another's internal conflict in real-world scenarios.	Future research can replicate the findings in real-world contexts, such as political polling.
Construct Validity		
Accuracy of others' behavior	Although participants were more likely to align their beliefs with another's behavior, it is unclear how accurate their perceptions are.	Future research will need to link these predictions to actual behavior (e.g., real health decisions, voting behavior, attitudinal change).
Statistical Validity		
Restriction of range	Mean true self scores were above the mid-point of the scale (Study 3: $M = 4.61$, $SD = 1.31$; 48.8% > midpoint of the scale), and thus, we did not have high statistical power to make precise inferences about individuals on the low-end of true self beliefs.	Future research can recruit a large sample to more powerfully capture the full range of the construct.
Internal Validity		
Mechanism of true self beliefs	The mechanisms by which true self beliefs come to inform people's theories of preferences remain unclear.	Future research should examine whether the tendency to see one's own normative beliefs as inherent and fundamental, as well as the tendency to mentally represent various entities in terms of an ideal/essence contribute to this effect (see Chen et al., 2016; Christy et al., 2019; Newman et al., 2015; Newman & Knobe, 2019 for further discussion).

References

Ames, D. R. (2004). Strategies for social inference: A similarity contingency model of projection and stereotyping in attribute prevalence estimates. *Journal of Personality and Social Psychology, 87*(5), 573–585. https://doi.org/10.1037/0022-3514.87.5.573

Andersen, S. M., & Ross, L. (1984). Self-knowledge and social inference: I. The impact of cognitive/affective and behavioral data. *Journal of Personality and Social Psychology*, *46*(2), 280-293. https://doi.org/10.1037/0022-3514.46.2.280

Barasch, A., Levine, E. E., Berman, J. Z., & Small, D. A. (2014). Selfish or selfless? On the signal value of emotion in altruistic behavior. *Journal of Personality and Social Psychology, 107*(3), 393–413. https://doi.org/10.1037/a0037207

Baron, J., & Spranca, M. (1997). Protected values. *Organizational Behavior and Human Decision Processes*, *70*(1), 1-16. https://doi.org/10.1006/obhd.1997.2690

Bench, S. W., Schlegel, R. J., Davis, W. E., & Vess, M. (2015). Thinking about change in the self and others: The role of self-discovery metaphors and the true self. *Social Cognition*, *33*(3), 169–185. https://doi.org/10.1521/soco.2015.33.3.2

Boiney, L. G., Kennedy, J., & Nye, P. (1997). Instrumental bias in motivated reasoning: More when more is needed. *Organizational Behavior and Human Decision Processes*, *72*(1), 1-24. https://doi.org/10.1006/obhd.1997.2729

Böttigheimer, C., & Widenka, W. M. (2023). *The concept of soul in Judaism, Christianity and Islam*. De Gruyter.

Brenner, L., & Bilgin, B. (2011). Preference, projection, and packing: Support theory models of judgments of others' preferences. *Organizational Behavior and Human Decision Processes*, *115*(1), 121-132. https://doi.org/10.1016/j.obhdp.2010.11.007

Chen, S. Y., Urminsky, O., & Bartels, D. M. (2016). Beliefs about the causal structure of the self-concept determine which changes disrupt personal identity. *Psychological Science*, *27*(10), 1398–1406. https://doi.org/10.1177/0956797616656800

Christy, A. G., & Schlegel, R. J. (2024). Perceiving persons and their purposes: Teleology, normativity, and personal identity. *Self and Identity*, *23*(1-2), 23-48. https://doi.org/10.1080/15298868.2024.2314921

Christy, A. G., Schlegel, R. J., & Cimpian, A. (2019). Why do people believe in a "true self"? The role of essentialist reasoning about personal identity and the self. *Journal of Personality and Social Psychology*, *117*(2), 386-416. https://doi.org/10.1037/pspp0000254

Collins, S. E., Malone, D. K., & Larimer, M. E. (2012). Motivation to change and treatment attendance as predictors of alcohol-use outcomes among project-based Housing First residents. *Addictive Behaviors*, *37*(8), 931-939. https://doi.org/10.1016/j.addbeh.2012.03.029

Critcher, C. R., Inbar, Y., & Pizarro, D. A. (2013). How quick decisions illuminate moral character. *Social Psychological and Personality Science*, *4*(3), 308-315. https://doi.org/10.1177/1948550612457688

Cusimano, C., & Lombrozo, T. (2021). Morality justifies motivated reasoning in the folk ethics

of belief. *Cognition*, *209*, 104513. https://doi.org/10.1016/j.cognition.2020.104513

Dane, E., Rockmann, K. W., & Pratt, M. G. (2012). When should I trust my gut? Linking domain expertise to intuitive decision-making effectiveness. *Organizational Behavior and Human Decision Processes*, *119*(2), 187–194. https://doi.org/10.1016/j. obhdp.2012.07.009

De Freitas, J., & Cikara, M. (2018). Deep down my enemy is good: Thinking about the true self reduces intergroup bias. *Journal of Experimental Social Psychology*, *74*, 307-316. https://doi.org/10.1016/j.jesp.2017.10.006

De Freitas, J., Sarkissian, H., Newman, G. E., Grossmann, I., De Brigard, F., Luco, A., & Knobe, J. (2018). Consistent belief in a good true self in misanthropes and three interdependent cultures. *Cognitive science*, *42*, 134-160. https://doi.org/10.1111/cogs.12505

De Freitas, J., Tobia, K. P., Newman, G. E., & Knobe, J. (2017). Normative judgments and individual essence. *Cognitive Science*, *41*, 382-402. https://doi.org/10.1111/cogs.12364

Frankfurt, H. G. (1971). Freedom of the will and the concept of a person. *The Journal of Philosophy*, *68*(1), 5–20. https://doi.org/10.2307/2024717

Furnham, A., & Henley, S. (1988). Lay beliefs about overcoming psychological problems. *Journal of Social and Clinical Psychology*, *6*(3-4), 423-438.

Garrison, K. E., Rivera, G. N., Schlegel, R. J., Hicks, J. A., & Schmeichel, B. J. (2023). Authentic for thee but not for me: Perceived authenticity in self-control conflicts. *Personality and Social Psychology Bulletin*, *49*(12), 1646-1662. https://doi.org/10.1177/01461672221118187

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. Oxford University Press, USA.

Gooding, D. (2024, October 24). The polls could be wrong—And still benefit Harris. *Newsweek*. https://www.newsweek.com/polling-errors-harris-advantage-trump-strategist-1974317

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

Haslam, N., Bastian, B., & Bissett, M. (2004). Essentialist beliefs about personality and their implications. *Personality and Social Psychology Bulletin*, *30*(12), 1661-1673. https://doi.org/10.1177/0146167204271182

Heiphetz, L., Strohminger, N., & Young, L. L. (2017). The role of moral beliefs, memories, and preferences in representations of identity. *Cognitive Science, 41*(3), 744–767. https://doi.org/10.1111/cogs.12354

Hofmann, W., Meindl, P., Mooijman, M., & Graham, J. (2018). Morality and self-control: How they are intertwined and where they differ. *Current Directions in Psychological Science*, *27*(4), 286-291. https://doi.org/10.1177/0963721418759317

Inbar, Y., Cone, J., & Gilovich, T. (2010). People's intuitions about intuitive insight and intuitive choice. *Journal of Personality and Social Psychology*, *99*(2), 232–247. https://doi.org/10.1037/a0020215

Johnson, J. T., & Boyd, K. R. (1995). Dispositional traits versus the content of experience:

Actor/observer differences in judgments of the "Authentic Self". *Personality and Social Psychology Bulletin*, *21*(4), 375–383. https://doi.org/10.1177/0146167295214008

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

Jost, J. T., Banaji, M. R., & Nosek, B. A. (2004). A decade of system justification theory: Accumulated evidence of conscious and unconscious bolstering of the status quo. *Political Psychology*, *25*(6), 881-919. https://doi.org/10.1111/j.1467-9221.2004.00402.x

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Kahneman, D., & Frederick, S. (2005). A model of heuristic judgment. In K. J. Holyoak & R. G. Morrison (Eds.), T*he Cambridge handbook of thinking and reasoning* (pp. 267-294). Cambridge Univervisty Press.

Keil, F. C. (1989). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.

Kidder, R. M. (2009). *How good people make tough choices: Resolving the dilemmas of ethical living*. HarperCollins.

Koole, S. L., & Kuhl, J. (2003). In search of the real self: A functional perspective on optimal self-esteem and authenticity. *Psychological Inquiry*, *14*(1), 43–48.

Kruger, J., Wirtz, D., & Miller, D. T. (2005). Counterfactual thinking and the first instinct fallacy. *Journal of Personality and Social Psychology*, *88*(5), 725-735. https://doi.org/10.1037/0022-3514.88.5.725

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

Kunda, Z., & Sanitioso, R. (1989). Motivated changes in the self-concept. *Journal of Experimental Social Psychology*, *25*(3), 272-285. https://doi.org/10.1016/0022-1031(89)90023-1

Lee, S. C. & Feldman, G. (2022). Revisiting the link between true-self and morality: Replication and extension of Newman, Bloom, and Knobe (2014) Studies 1 and 2, in principle acceptance of Version 2 by Peer Community in Registered Reports. https://osf.io/v2tpf

Maglio, S. J., & Reich, T. (2018). Feeling certain: Gut choice, the true self, and attitude certainty. *Emotion*, *19*(5), 876-888. https://doi.org/10.1037/emo0000490

Maglio, S. J., & Reich, T. (2020). Choice protection for feeling-focused decisions. *Journal of Experimental Psychology: General*, *149*(9), 1704–1718. https://doi.org/10.1037/xge0000735

Mallon, E. (2024, October 4). *Harris struggles with undecided voters as she flees making promises about policy*. Washington Examiner. https://www.washingtonexaminer.com/news/campaigns/presidential/3177276/kamala-harris-struggles-undecided-voters/

Marks, G., & Miller, N. (1985). The effect of certainty on consensus judgments. *Personality and Social Psychology Bulletin*, *11*(2), 165-177. https://doi.org/10.1177/0146167285112005

Miller, D. T., & Taylor, B. R. (2002). Counterfactual thought, regret, and superstition: How to

avoid kicking yourself. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 367–378). Cambridge University Press

Molouki, S., & Bartels, D. M. (2017). Personal change and the continuity of the self. *Cognitive Psychology*, *93*, 1-17. https://doi.org/10.1016/j.cogpsych.2016.11.006

Morewedge, C. K., Giblin, C. E., & Norton, M. I. (2014). The (perceived) meaning of spontaneous thoughts. *Journal of Experimental Psychology: General*, *143*(4), 1742-1754. http://dx.doi.org/10.1037/a0036775

Navarick, D. J. (2013). Moral ambivalence: Modeling and measuring bivariate evaluative processes in moral judgment. *Review of General Psychology*, *17*(4), 443-452. https://doi.org/10.1037/a0034527

Newman, G. E., & Keil, F. C. (2008). Where is the essence? Developmental shifts in children's beliefs about internal features. *Child development*, *79*(5), 1344-1356. https://doi.org/10.1111/j.1467-8624.2008.01192.x

Newman, G. E., & Knobe, J. (2019). The essence of essentialism. *Mind & Language*, *34*(5), 585-605. https://doi.org/10.1111/mila.12226

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

Newman, G. E., De Freitas, J., & Knobe, J. (2015). Beliefs about the true self explain asymmetries based on moral judgment. *Cognitive Science*, *39*(1), 96–125.

https://doi.org/10.1111/cogs.12134

Oktar, K., & Lombrozo, T. (2022). Deciding to be authentic: Intuition is favored over deliberation when authenticity matters. *Cognition, 223,* 1-21. https://doi.org/10.1016/j.cognition.2022.105021

Pachur, T., & Spaar, M. (2015). Domain-specific preferences for intuition and deliberation in decision making. *Journal of Applied Research in Memory and Cognition*, *4*(3), 303-311. https://doi.org/10.1016/j.jarmac.2015.07.006

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(3), 534-552. https://doi.org/10.1037/0278-7393.14.3.534

Pew Research Center. (2006, March 28). *A barometer of modern morals*. https://www.pewresearch.org/social-trends/2006/03/28/a-barometer-of-modern-morals/

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. (2024a, 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. (2024b, October 10). *The Harris-Trump matchup*. https://www.pewresearch.org/politics/2024/10/10/the-harris-trump-matchup/

Posit Team (2024). RStudio: Integrated Development Environment for R. Posit Software, PBC, Boston, MA. http://www.posit.co/

Richert, R. A., & Smith, E. (2012). The essence of soul concepts: How soul concepts influence ethical reasoning across religious affiliation. *Religion, Brain & Behavior*, *2*(2), 161-176. https://doi.org/10.1080/2153599X.2012.683702

Rogers, T., Moore, D. A., & Norton, M. I. (2017). The belief in a favorable future. *Psychological Science*, *28*(9), 1290-1301. https://doi.org/10.1177/0956797617706706

Ross, L. D., Lelkes, Y., & Russell, A. G. (2012). How Christians reconcile their personal political views and the teachings of their faith: Projection as a means of dissonance reduction. *Proceedings of the National Academy of Sciences*, *109*(10), 3616-3622. https://doi.org/10.1073/pnas.1117557109

Ross, L., & Ward, A. (1996). Naive realism in everyday life: Implications for social conflict and misunderstanding. In E. S. Reed, E. Turiel, & T. Brown (Eds.), *Values and knowledge* (pp. 103–135). Lawrence Erlbaum Associates, Inc.

Rossignac-Milon, M., Pillemer, J., Bailey, E. R., Horton Jr, C. B., & Iyengar, S. S. (2024). Just be real with me: Perceived partner authenticity promotes relationship initiation via shared reality. *Organizational Behavior and Human Decision Processes*, *180*, 104306. https://doi.org/10.1016/j.obhdp.2023.104306

Schlegel, R. J., Hicks, J. A., Davis, W. E., Hirsch, K. A., & Smith, C. M. (2013). The dynamic interplay between perceived true self-knowledge and decision satisfaction. *Journal of Personality and Social Psychology*, *104*(3), 542–558. https://doi.org/10.1037/a0031183

Schrift, R. Y., Netzer, O., & Kivetz, R. (2011). Complicating choice. *Journal of Marketing Research*, *48*(2), 308–326. https://doi.org/10.1509/jmkr.48.2.308

Schwartz, B. (2004). *The paradox of choice: Why more is less*. Harper Perennial.

Sedikides, C., & Schlegel, R. J. (2024). Distilling the concept of authenticity. *Nature Reviews Psychology*, *3*(8), 509-523. https://doi.org/10.1038/s44159-024-00323-y

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Snyder, M. L., Kleck, R. E., Strenta, A., & Mentzer, S. J. (1979). Avoidance of the handicapped: An attributional ambiguity analysis. *Journal of Personality and Social Psychology, 37*(12), 2297–2306. https://doi.org/10.1037/0022-3514.37.12.2297

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, *23*(5), 645–665. https://doi.org/10.1017/S0140525X00003435

Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. In L. Berkowitz (Ed.), *Advances in experimental social psychology, Vol. 21. Social psychological studies of the self: Perspectives and programs* (pp. 261–302). Academic Press.

Strohminger, N., & Nichols, S. (2014). The essential moral self. *Cognition*, *131*(1), 159-171. https://doi.org/10.1016/j.cognition.2013.12.005

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

Sun, J., Wilt, J., Meindl, P., Watkins, H. M., & Goodwin, G. P. (2024). How and why people want to be more moral. *Journal of Personality*, *92*(3), 907-925. https://doi.org/10.1111/jopy.12812

Tetlock, P. E., Kristel, O. V., Elson, S. B., Green, M. C., & Lerner, J. S. (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology, 78*(5), 853–870. https://doi.org/10.1037/0022-3514.78.5.853

Todd, A. R., & Tamir, D. I. (2024). Factors that amplify and attenuate egocentric mentalizing. *Nature Reviews Psychology*, *3*(3), 164-180. https://doi.org/10.1038/s44159-024-00277-1

Turiel, E., Hildebrandt, C., Wainryb, C., & Saltzstein, H. D. (1991). Judging social issues: Difficulties, inconsistencies, and consistencies. *Monographs of the Society for Research in Child Development*, 1-116. https://doi.org/10.2307/1166056

Wang, Y. A., Simpson, A. J., & Todd, A. R. (2023). Egocentric anchoring-and-adjustment underlies social inferences about known others varying in similarity and familiarity. *Journal*

of Experimental Psychology: General, *152*(4), 1011–1029. https://doi.org/10.1037/xge0001313

Waytz, A., Dungan, J., & Young, L. (2013). The whistleblower's dilemma and the fairness–loyalty tradeoff. *Journal of Experimental Social Psychology*, *49*(6), 1027-1033. https://doi.org/10.1016/j.jesp.2013.07.002

Woolley, K., & Risen, J. L. (2018). Closing your eyes to follow your heart: Avoiding information to protect a strong intuitive preference. *Journal of Personality and Social Psychology*, *114*(2), 230–245. https://doi.org/10.1037/pspa0000100

Woolley, K., & Risen, J. L. (2021). Hiding from the truth: When and how cover enables information avoidance. *Journal of Consumer Research*, *47*(5), 675-697. https://doi.org/10.1093/jcr/ucaa030

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

Appendix A. Vignettes used in Study 1

Target:	***Self***		***Other***	
Scenario Type:	***Normative***	***Neutral***	***Normative***	***Neutral***
Vignette 1	Suppose that you are feeling fatigued at work. You are having conflicting feelings about taking an illegal drug that increases alertness. Your "gut" feeling is to **not** take the drug. When you think for a long time, your feeling is **to take** the drug.	Suppose that you are feeling fatigued at work. You are having conflicting feelings about having another cup of coffee. Your "gut" feeling is to **not** have another cup of coffee. When you think for a long time, your feeling is **to have** another cup of coffee.	Suppose that someone is feeling fatigued at work. They are having conflicting feelings about taking an illegal drug that increases alertness. Their "gut" feeling is to **not** take the drug. When they think for a long time, their feeling is **to take** the drug.	Suppose that someone is feeling fatigued at work. They are having conflicting feelings about having another cup of coffee. Their "gut" feeling is to **not** have another cup of coffee. When they think for a long time, their feeling is **to have** another cup of coffee.
Vignette 2	Suppose that you are getting lunch at your workplace cafeteria. You are having conflicting feelings about stealing a sandwich. Your "gut" feeling is to **not** steal the sandwich. When you think for a long time, your feeling is **to steal** the sandwich.	Suppose that you are getting lunch at your workplace cafeteria. You are having conflicting feelings about buying a sandwich. Your "gut" feeling is to **not** buy the sandwich. When you think for a long time, your feeling is **to buy** the sandwich.	Suppose that someone is getting lunch at their workplace cafeteria. They are having conflicting feelings about stealing a sandwich. Their "gut" feeling is to **not** steal the sandwich. When they think for a long time, their feeling is **to steal** the sandwich.	Suppose that someone is getting lunch at their workplace cafeteria. They are having conflicting feelings about buying a sandwich. Their "gut" feeling is to **not** buy the sandwich. When they think for a long time, their feeling is **to buy** the sandwich.
Vignette 3	Suppose that you are responsible for choosing a supplier for office supplies for the coming year. You are having conflicting feelings about switching to a new supplier that is known to use child labor. Your "gut" feeling is to **not** switch to the new supplier that uses child labor When you think for a long time, your feeling is **to switch** to the new supplier that uses child labor.	Suppose that you are responsible for choosing a supplier for office supplies for the coming year. You are having conflicting feelings about switching to a new supplier. Your "gut" feeling is to **not** switch to the new supplier When you think for a long time, your feeling is **to switch** to the new supplier.	Suppose that someone is responsible for choosing a supplier for office supplies for the coming year. They are having conflicting feelings about switching to a new supplier that is known to use child labor. Their "gut" feeling is to **not** switch to the new supplier that uses child labor When they think for a long time, their feeling is **to switch** to the new supplier that uses child labor.	Suppose that someone is responsible for choosing a supplier for office supplies for the coming year. They are having conflicting feelings about switching to a new supplier. Their "gut" feeling is to **not** switch to the new supplier When they think for a long time, their feeling is **to switch** to the new supplier.
Vignette 4	Suppose that you are a sales representative. You are having conflicting feelings about recommending a product that is harmful to the environment. Your "gut" feeling is to **not** recommend the environmentally harmful product. When you think for a long time, your feeling is **to recommend** the environmentally harmful product.	Suppose that someone is a sales representative. They are having conflicting feelings about recommending a product. Their "gut" feeling is to **not** recommend the product. When they think for a long time, their feeling is **to recommend** the product.	Suppose that someone is a sales representative. They are having conflicting feelings about recommending a product that is harmful to the environment. Their "gut" feeling is to **not** recommend the environmentally harmful product. When they think for a long time, their feeling is **to recommend** the environmentally harmful product.	Suppose that someone is a sales representative. They are having conflicting feelings about recommending a product. Their "gut" feeling is to **not** recommend the product. When they think for a long time, their feeling is **to recommend** the product.

Appendix B.

Vignettes used in Study 2

Neutral Scenarios

Suppose that someone is deciding to buy their first computer. They are deciding between a Mac or PC.

- Their "gut" feeling is to buy a Mac.
- When they think for a long time, their feeling is to buy a PC.

Suppose that someone is deciding where to sit on an upcoming flight. They are deciding between an aisle seat or a window seat.

- Their "gut" feeling is to select the aisle seat.
- When they think for a long time, their feeling is to select the window seat.

Suppose that someone is deciding to adopt their first pet. They are deciding between a cat or a dog.

- Their "gut" feeling is to get a cat.
- When they think for a long time, their feeling is to get a dog.

Suppose that someone is deciding to go watch a movie. They are deciding between a horror film or a romantic comedy.

- Their "gut" feeling is to watch the horror film.
- When they think for a long time, their feeling is to watch the romantic comedy.

Normative Scenarios

Suppose that someone is deciding what their stance on gun control is. They are deciding between favoring strict regulations or favoring more lenient policies.

- Their "gut" feeling is to support strict gun regulations.
- When they think for a long time, their feeling is to support more lenient gun regulations.

Suppose that someone is deciding what their stance on immigration is. They are deciding between being an advocate for open borders or an advocate for closed borders.

- Their "gut" feeling is to advocate for open borders.
- When they think for a long time, their feeling is to advocate for closed borders.

Suppose that someone is deciding what their stance on climate change is. They are deciding between viewing it as a top priority or considering it a low priority.

- Their "gut" feeling is to view climate change as a top priority.
- When they think for a long time, their feeling is to view climate change as a low priority.

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).

- Their "gut" feeling is to be pro-life.
- When they think for a long time, their feeling is to be pro-choice.