Journal of Experimental Social Psychology 105 (2023) 104437

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

journal homepage: www.elsevier.com/locate/jesp

Desired attitudes guide actual attitude change☆

Thomas I. Vaughan-Johnston [a,*], Leandre R. Fabrigar [b], Ji Xia [c], Kenneth G. DeMarree [c], Jason K. Clark [d]

[a] Durham University, Department of Psychology, United Kingdom
[b] Queen's University, Department of Psychology, Canada
[c] University at Buffalo, Department of Psychology, United States of America
[d] University of Virginia, Department of Psychology, United States of America

ARTICLE INFO

Editor: Dr Evava Pietri

Keywords:
Attitudes
Attitude change
Desired attitudes
Mere thought
Motivation

ABSTRACT

Whereas *actual attitudes* represent people's evaluations of specific objects as being good or bad, *desired attitudes* represent the attitude positions that people wish they held. Previous work has established that desired attitudes can have psychological consequences, but has not yet tested the extent to which desired attitudes can predict actual attitude change. In five datasets involving a variety of populations and procedural variations, we explored how political and moral factors motivate people to form desired attitudes distinct from their actual attitudes. These desired attitudes then predicted actual attitude change occurring after the formation of the desired attitudes, even though participants received no new information about the object. The present work demonstrates an additional value of the desired attitude construct: its ability to anticipate how people's attitudes will shift in the future.

In 2020, J. K. Rowling wrote a series of tweets expressing her viewpoints about sex and gender, resulting in substantial public controversy (Duggan, 2022; Rowling, 2020). Rowling is the bestselling author of the Harry Potter fantasy novel series, and doubtlessly many people held positive views of her books when this controversy arose. However, many fans evidently saw Rowling's comments as expressing problematic values and/or beliefs. We suspect that simultaneously liking the books and hating the author produced intense psychological conflict for many people. Some initially Potter-liking people might have felt confident "separating the art from the artist," and for that reason might not desire to change their opinion of her books, but others may have felt an intense public and/or private pressure to revisit their opinions of Rowling's novels. We are interested in whether people who form such desired attitudes might attain these goals. Liking something while wanting to dislike it (or vice versa), as the present work argues, may have important attitudinal implications.

The above is hardly an isolated example. People often fall in love with art or other products, only to become alienated from the producers of those things, potentially creating discrepancies between the opinions people actually hold (versus wish to hold) about objects. More broadly, people often experience conflicts between their current opinions and their identity-laden moral positions or broader goals. People may want to like a co-worker more than they actually do, or desire to enjoy the flavor of healthy foods like kale or sardines. Researchers have examined this interesting and commonplace phenomenon. Whereas *actual attitudes*, people's evaluations of things as being good and/or bad, have long been recognized as being a central construct in social psychology (Eagly & Chaiken, 1993; Maio & Haddock, 2009; Ostrom, 1989; Petty & Cacioppo, 1981; Petty, Wegener, & Fabrigar, 1997; Thurstone & Chave, 1929), *desired attitudes* are a relatively novel construct that captures the attitudes that people wish they held (DeMarree, Wheeler, Briñol, & Petty, 2014). A growing literature has explored desired attitudes' conceptual distinctiveness from existing constructs, their frequency, antecedents and consequences, and moderators (Carrera, Caballero, Muñoz, & Fernández, 2017; DeMarree et al., 2014; DeMarree, Clark, Wheeler, Briñol, & Petty, 2017; DeMarree & Rios, 2014).

Although desired attitudes relate to a range of important consequences, researchers have yet to examine the extent to which they can influence a key concept in the attitudes literature: attitude change. Given that desired attitudes are the opinions that people want to hold, this absence of a direct connection to actual attitude change is surprising. We propose that people adopt desired attitudes as internal goals that

☆ This paper has been recommended for acceptance by Dr Evava Pietri.

* Corresponding author.
E-mail address: thomasvaughanjohnston@gmail.com (T.I. Vaughan-Johnston).

https://doi.org/10.1016/j.jesp.2022.104437
Received 28 July 2022; Received in revised form 28 November 2022; Accepted 12 December 2022
Available online 24 December 2022

then can effectively predict the direction of their subsequent attitude regulation efforts. These internal goals are important because self-persuasion is often a cognitively demanding task (Maio & Thomas, 2007); therefore, people may benefit from forming a target of what attitude they want (i.e., whether they want a more positive/negative attitude, and how extremely the discrepancy from their present attitude is). Previous work establishes that desired attitudes are related to a biased preference for, and processing of novel, external information (DeMarree et al., 2017). However, we propose that desired attitudes can shift people's opinions even when they lack the opportunity to facilitate attitude change with new information. Thus, we examine an impoverished situation in which people cannot receive any novel information between their forming a desired attitude and their subsequent actual attitude change. Nonetheless, we anticipate that even without new information, people will self-persuade towards their desired attitude goal, by reflecting on and reorganizing information that they already possess about an object (Briñol, McCaslin, & Petty, 2012; Clarkson, Tormala, & Leone, 2011; Maio & Thomas, 2007; Tesser, 1978; Tesser & Conlee, 1975; Tesser & Leone, 1977). No previous work has examined whether these attitude regulation efforts guided by desired attitudes effectively produce attitude change.

1. Desired attitudes and actual attitude change

The beliefs we wish we held often differ markedly from those that we actually hold. This is recognized in the self literature, which has examined *self-discrepancies* as gaps between the views people actually hold about themselves, versus the views people wish they held about themselves (e.g., Higgins, 1987, 1989; also see Carver & Scheier, 1998). Past work has often suggested that actual/desired self-discrepancies are aversive, and that people may effortfully pursue these desired "self-guides." However, discrepancies need not apply to self-evaluations. DeMarree et al. (2014, 2017) described desired attitudes as evaluations that people are motivated to hold about any discrete attitude object (i.e., person, thing, or concept; also see Maio & Thomas, 2007). For instance, people may like Walmart (actual attitude) because it is cheap, and has a wide variety of products, but may wish to dislike Walmart (desired attitude) because of its associations with worker exploitation and deceiving the public (Benoit & Dorries, 1996). Desired attitudes follow from people's motivational drives, helping them to accomplish goals, claim important identities, or maintain consistency with higher order values and ideologies (Wheeler & DeMarree, 2019). However, desired attitudes are distinct from the drives that create them, representing desires to hold particular opinions about particular attitude objects.

Desired attitudes are consequential for how people pursue attitude positions. For instance, just as people's actual attitudes can predict increased seeking and processing of attitude-congenial information (e.g., Houston & Fazio, 1989; Lord, Ross, & Lepper, 1979), people's desired attitudes similarly predict evaluatively congruent information seeking and processing, over and above the influence of actual attitudes (DeMarree et al., 2017). Additionally, people may modify the properties of attitude objects to reach a desired attitude goal. For example, people who anticipated drinking some coffee used more additives (i.e., sweeteners, creamers) as their desired attitudes towards coffee became more positive (DeMarree et al., 2017, Study 4). In contrast, those with positive actual attitudes towards coffee used less additives. Together, these findings suggest that people are motivated to pursue their desired attitudes through various strategies, dedicating significant effort to reconcile their actual attitudes with their desired attitude goals. However, no existing empirical work has connected desired attitudes to actual attitude change.

Connecting desired attitudes to actual attitude change is important not only because such attitude change effects are theoretically interesting, but also because extant research might challenge the likelihood of such attitude change unfolding. For instance, people effortfully pursue both their desired and their actual attitudes simultaneously (e.g., DeMarree et al., 2017). Consequently, it is not obvious that people will shift their actual attitudes to match their desired attitudes, rather than the reverse. Consider that when people are ambivalent about a topic, they generally resolve this ambivalence by strengthening the relatively dominant (stronger) attitude component rather than by bolstering the conflicting (weaker) attitude component (e.g., Clark, Wegener, & Fabrigar, 2008). Insofar as desired attitudes are a hypothetical or imagined state of mind (e.g., "it *would be* good *if* I evaluated X positively"), it might be easier for people to change their desired attitudes to match actual attitudes, rather than changing their actual ("real") attitudes to match desired states. However, past work indicates that people are often highly committed to their 'hypothetical' (desired) attitudes (DeMarree et al., 2017) so people may shift their actual attitudes to match their desired attitudes, provided that the motivation that promotes the desired attitudes is sufficiently powerful (e.g., driven by one's political identity, fundamental moral principles; Wheeler & DeMarree, 2019).

2. Desired attitudes as self-persuasive goals

For the above reasons, we propose that desired attitudes will often have consequences for people's actual attitude change. This idea is also consistent with *cognitive dissonance theory*, which is based on the premise that people often experience tension when incongruency exists between elements of their cognition and thus people work to minimize the resulting aversive state of dissonance through several tactics including changing one or more of the discrepant cognitive elements (Cooper, 2007; Festinger, 1957; Harmon-Jones & Mills, 2019; Simon, Greenberg, & Brehm, 1995; see also Carver & Scheier, 1998; Heider, 1946). Thus, people could shift their opinions over time to reconcile their actual and desired attitudes, thereby minimizing conflict between their attitude positions. Given that a desired attitude is, by definition, an attitude position that people want to hold, people might engage in considerable cognitive efforts to attain that attitude. Desired attitudes may be initially stimulated by external information, but they may lead people to feel an internal pressure to regain cognitive consonance between their actual and desired attitudes. For instance, learning that a painting you initially liked was created by a reviled person (e.g., a Nazi) doubtlessly will lead you to immediately like the painting less. But the reasons for the originally positive evaluation (e.g., its pleasant aesthetic qualities) are not strictly contradicted by this new information, and substantial cognitive effort may be required to talk yourself into a satisfactorily negative opinion of the painting. Thus, the Nazi-related information might also create a desire to dislike the painting, and that desire might be felt as an internally driven pressure that shapes subsequent thinking and thus attitude change (i.e., self-persuasion; Aronson, 1999; Maio & Thomas, 2007). The desired attitude is not just memory of the novel information because it relies on an evaluation of the information (i.e., how much one hates Nazis) and judgments of relevance and importance (e.g., should one judge a painting according to its creator's moral characteristics, and to what degree?), integrated into an estimate of "taking it all in, what attitude do I want to obtain?" (i.e., a desired attitude).

If desired attitudes represent the motivation people have to like/dislike something, how might they convince themselves to actually adopt their desired attitude positions? Psychologists have often demonstrated strategies that people use to self-persuade, shaping their own opinions towards an object (Briñol et al., 2012). Additionally, much of this past work supports that people can change their minds without having to collect new information to justify attitude change. Even without changing the content of thoughts, metacognitive processes (i.e., thoughts about one's thoughts) can change the impact of thoughts on attitudes (Petty, Briñol, & Tormala, 2002). Validated thoughts have larger influence over people's attitudes, whereas invalidated thoughts have reduced influence over attitudes (Briñol et al., 2018; Requero, Briñol, & Petty, 2021). Additionally, Maio and Thomas (2007) pointed out a variety of metacognitive strategies people may use to self-persuade

towards situationally important goals by reorganizing or reweighting information (e.g., changing attributions, reweighting the importance of information). Thus interestingly, people may show substantial attitude change towards an object even without learning anything new about that object, and thus propose that people may be able to shift their actual attitudes to match their desired attitudes even without any opportunity to modify the object or collect a biased set of new information to facilitate that actual attitude change (i.e., the processes previously established for desired attitudes). We see desired attitudes as likely to guide this process by providing people with a target to direct their cognitive and metacognitive strategies. To maximize the likelihood of self-persuasion, we provided a series of metacognitive prompts in Experiments 1–2. However, we suspected that participants also would engage in metacognitive reflection without explicit prompting, and we test this in Experiment 3.

3. Overview and hypotheses

We investigated these questions with five experiments. Our main goal was to examine if desired attitudes shape the direction of people's actual attitude change. Because desired attitudes indicate the actual attitude position that people are motivated to pursue, we reasoned that people could persuade themselves into the attitude position indicated by their desired attitude. Specifically, we hypothesized that the desired attitudes we created in these paradigms would predict subsequent actual attitude change, even across a period when participants would have no opportunity to change the object's objective properties or acquire new information, thus creating circumstances where participants could only self-persuade to reach their desired attitude position.

4. Experiment 1

The central goal in Experiment 1 was to provide an initial test of our hypothesis that desired attitudes can predict subsequent actual attitude change, adjusting for the immediate attitude change produced by information, and even in a context where established attitude-relevant processes (physically modifying the object, biased information search) would be impossible. To do this, we had people form an actual attitude about a fictitious target individual, and then observed how desired attitudes formed when learning of the target's political affiliation (which matched or mismatched the participants' own political stance). We reasoned that a matching (mismatching) target would create positive (negative) desired attitudes, motivated by participants' fidelity to their social/political group (i.e., their social identity; Hogg & Ridgeway, 2003; Tajfel, 1974), and/or as an expression of participants' core values (i.e., moral principles leading them to identify as liberal or conservative in the first place; Graham, Haidt, & Nosek, 2009). Because our predictions were novel, we collected two samples in order to self-replicate, labeled as Experiment 1a and 1b.

4.1. Methods

4.1.1. Participants

For Experiment 1a, we recruited 324 Canadian university students (M_{age} = 20.2 years, SD_{age} = 6.8). We neglected to include demographic items, but this sample was drawn from a population that is 78.9% women, 20.7% men, and 0.4% non-binary; and majority white (73.3%; 15.5% East Asian, 5.3% South Asian, 2.7% Black, 1.8% Hispanic), according to a pre-screening survey. For Experiment 1b, we recruited 700 students from an American university (M_{age} = 19.1 years, SD_{age} = 1.3; 62.6% women, 37.4% men, 0% non-binary). We used time-based stopping rules, obtaining 80% power to identify small effects, $r > 0.15$ ($r > 0.11$), for individual regression parameters in Experiment 1a (Experiment 1b; G*Power v. 3.1.9.4; Faul, Erdfelder, Buchner, & Lang, 2009). This compares favorably with the results of an internal meta-analysis of desired attitudes predicting "pursuit-oriented outcomes" of desired attitudes; r = 0.19 (DeMarree et al., 2017). Sample size was always determined before any data analysis. Experiment 1b was preregistered, although the preregistration covers additional analyses not reported here (see footnote 1), and some alterations were made compared to the preregistration (see "Changes from the preregistration" section, below).

4.1.2. Procedure

The procedure is depicted in Fig. 1. Participants answered demographic questions including a political orientation scale. Next, participants read a transcript from an ostensibly real interview between a politician and an interviewer. The interview highlighted positive and negative attributes of the politician (e.g., political experience, cigarette-smoking habit) but not her political stance. To create moderate actual attitudes towards the politician (so that matching/mismatching information could drive subsequent attitude change towards her), participants listed three positive and three negative thoughts about the politician in counter-balanced order, and rated their Time 1 attitude towards the politician. Because order had no effects across our experiments, we do not discuss it further.

Next, to manipulate desired attitudes, half of our participants learned that the politicians' political views *matched* their own political views (i.e., she was liberal if the participant was liberal, conservative if the participant was conservative), whereas the other half were given information about the politician's views that *mismatched* their own political views (i.e., she was conservative for liberals; liberal for conservatives). Participants then rated their Time 2 desired attitudes towards the politician. Participants also rated their actual attitudes at this time, so that we could empirically distinguish between the immediate attitude change cultivated by the matching/mismatching information, and participants' later attitude changes. The format of our Time 1, Time 2, and Time 3 actual attitude measures were distinct (having two, one, and eight items, respectively; different endpoint values; different endpoint labels). These different formats helped to avoid consistency effects that might obscure real attitude change (see Blankenship, Wegener, & Murray, 2012, for a similar method). That is, if formatting was kept invariant, participants might have felt "locked in" to earlier attitude questions when answering later attitude questions about the same object (Downing, Judd, & Brauer, 1992; Petty & Cacioppo, 1981).

We next stimulated participants to reflect on their past thoughts with several metacognitive prompts intended to give them an opportunity to self-persuade (Petty et al., 2002). Participants read their positive and negative thoughts from earlier in the procedure and we prompted them to consider how likeable, certain, and clear each thought was.[1] Importantly, these prompts only encouraged participants to think further, and did not provide any new information about the politician. Participants then rated their Time 3 actual attitudes.[2]

4.2. Materials

For this and future studies, additional information on the measures including descriptive statistics and correlation tables between the

[1] In the present study, these three prompts were included to encourage participants to engage in meta-cognitive reappraisals of their evaluative responses to the attitude object. Thus, they were not intended as measures to test the current hypothesis, although responses to these prompts were recorded. Analyses of these measures are pertinent to a separate research project examining different types of thought perceptions, and will be reported in a future manuscript.

[2] We also assessed Time 3 behavioral intentions towards the politician with six items, e.g., "I would vote for [the politician] to be my political representative." Results for behavioral intentions converged very closely with results reported for actual attitudes, so to avoid redundancy while maintaining full transparency we report details for these analyses in the Supplementary Online Materials (SOM-3).

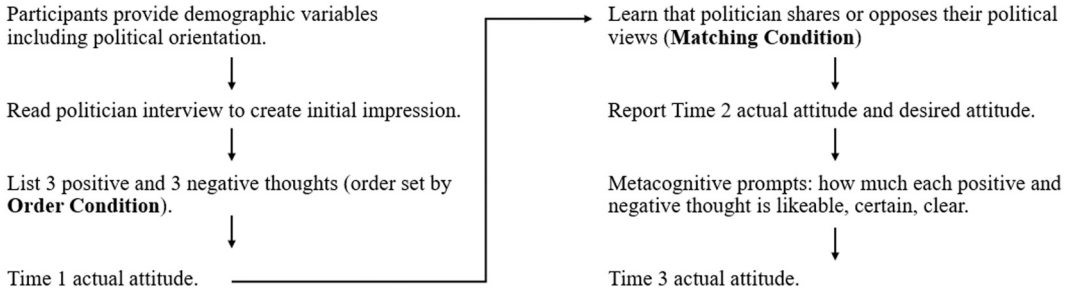

Fig. 1. Flow of the Experimental Procedure (Exp. 1).

variables appear in SOM-2.

4.2.1. Political orientation

We used a "forced-choice" single-item, six-point bipolar scale (1 = *very conservative*, 6 = *very liberal*). In Experiment 1a, 25.4% of participants were conservative, and 74.6% were liberal. In Experiment 1b, all but 11 participants provided their political orientation. In total, 42.8% of participants were conservative, and 57.2% were liberal. Experiment 1b's American sample was significantly less liberal ($M = 3.7$, $SD = 1.3$) than Experiment 1a's Canadian sample ($M = 4.3$, $SD = 1.2$), $t(1022) = -7.04$, $p < .001$, $d = -0.47$.

Because our politician's political beliefs were very liberal / very conservative, we examined political extremity as a potential moderator of the matching condition's effect on desired attitudes. We calculated political extremity as deviation from neutrality (i.e., 3.5), with a + 0.5 adjustment to create whole number scores, such that a 1 on the extremity scale reflected participants were "slightly" liberal/conservative, 2 reflected participants who were "moderately" liberal/conservative, and 3 reflected participants who were "strongly" liberal/conservative. In Experiment 1a, political extremity was significantly correlated with political ideology, $r(321) = 0.62$, $p < .001$, unsurprisingly given that the sample was mostly liberal. In Experiment 1b, political extremity was more weakly correlated with political ideology, $r(687) = 0.14$, $p < .001$; the correlation differed across samples; *r*-to-*z* transformation test, $z = 8.68$, $p < .001$.

4.2.2. Time 1 actual attitude

Participants provided their actual attitude towards the politician after reading the interview document. We used two items (1 = *very unlikeable* to 7 = *very likeable*; 1 = *dislike her* to 7 = *like her*); these items were highly inter-correlated in each sample, $r_{Study1a}(322) = 0.80$, $p < .001$, $r_{Study1b}(695) = 0.77$, $p < .001$, so we averaged these items in each dataset, higher scores indicating more positive attitudes.

4.2.3. Time 2 actual attitude

Participants provided their actual attitude towards the politician after learning about her political views on one item ranging from 1 (*strongly dislike*) to 9 (*strongly like*).

4.2.4. Time 2 desired attitude

Consistent with past research (e.g, DeMarree & Rios, 2014, p. 203), we briefly explained desired attitudes by stating, "Sometimes the attitudes we have are different from the attitudes we ideally would like to have or the attitudes we feel we should hold, and sometimes these attitudes are the same." Participants then indicated the attitude they "IDEALLY would like to have" (*ideal attitude*) and "the attitude [they] SHOULD or OUGHT to have" (*ought attitude*) from 1 (would IDEALLY / OUGHT TO strong dislike) to 9 (would IDEALLY / OUGHT TO strong like). Because past literature finds that effects of ideal and ought attitude effects converge (e.g., DeMarree et al., 2014), our interest was in desired attitudes broadly, and ideal/ought attitudes correlated highly, we averaged ideal and ought attitudes into a composite desired attitudes measure, only reporting ideal and ought attitude effects separately if results differed meaningfully for these items.

Desired attitudes have often been captured either with this composite (i.e., averaging ideal/ought attitudes) or through a subjective discrepancies measure (i.e., asking participants to directly estimate how much more/less they want to like an attitude object). In the present experiments we measured both. Results were similar for the two measures, and the two measures were highly correlated, $r_{Study1a}(322) = 0.69$, $p < .001$, $r_{Study1b}(694) = 0.65$, $p < .001$. We therefore report the subjective discrepancy measure's results in SOM-4.

4.2.5. Time 3 actual attitude

Eight items (Crites Jr., S. L., Fabrigar, & Petty, 1994) captured attitudes, each rated from 1 (*not at all*) to 9 (*definitely*): good, bad (R), like, dislike (R), desirable, undesirable (R), positive, and negative (R). After reversing (R) items, the scale was reliable ($\alpha = 0.97$), so we averaged items, higher scores indicating favourable attitudes.

4.3. Results

SOM-8 provides means and standard deviations for each dependent variable at all levels of the experimental variable(s) for all experiments.

4.3.1. Manipulation check: Desired attitudes

Our first analysis examined if desired attitudes formed based on identity-related pressures (the matching manipulation). Effects of matching / mismatching condition on desired attitudes would demonstrate that desired attitudes can form based on social/political goals, since the political identity of the target was the only variable manipulated. Additionally, we included Time 2 actual attitude (i.e., the degree of actual attitude change induced post-manipulation) as a covariate in this analysis. Obviously, people (actually) prefer members of their political ingroups over political outgroups (Bruchmann, Koopmann-Holm, & Scherer, 2018; Byrne, 1961), so by statistically controlling for this effect we isolated the unique effect that the manipulation had on desired attitudes specifically. In all experiments, the desired attitude manipulation also affected Time 2 actual attitudes; $t_{study1a}(322) = 12.06$, $p < .001$, $d = 1.34$; $t_{study1b}(694) = 10.22$, $p < .001$, $d = 0.78$; $t_{study2a}(115) = 9.04$, $p < .001$, $d = 1.67$; $t_{study2b}(512) = 7.65$, $p < .001$, $d = 0.67$; $t_{study3}(199) = 9.79$, $p < .001$, $d = 1.38$. We also considered if the influence of matching condition on desired attitude differed across levels of participant political extremity. Because our conservative and liberal politician profiles were relatively extreme, we expected an interaction of matching condition X political extremity. In other words, a "matching" (versus "mismatching") politician should match/mismatch most strongly for extreme-political participants, and thus the politician's ideology would dictate desired attitudes more powerfully for people at the extremes of the political spectrum (versus moderates).

To test these ideas, we regressed Time 2 desired attitudes onto a contrast-coded matching condition variable (−0.5 = mismatching, +0.5 = matching), Time 1 actual attitude, Time 2 actual attitude, centered political extremity, and a matching condition X political extremity

interaction term. Table 1 shows the results for each of our two samples. Matching condition (i.e., the first data row) had a large first-order effect on desired attitudes for people of average political extremity in Experiment 1a. The unstandardized *B* indicates that at average political extremity, the mean desired attitudes for people reading the matching condition passage was about one scale unit more positive than the mean for their mismatching condition counterparts. Thus, for people of average political extremity, learning that the politician matched (vs mismatched) their political views led participants in Experiment 1a (but not Experiment 1b) to desire to like her. (See Fig. 2)

We found no effect of Time 1 actual attitude on desired attitudes in either dataset. This null result may reflect that participants' Time 1 attitudes were captured before participants learned the politician's political orientation. This information might have been of reduced relevance after learning that politician's strong conservative or liberal views. However, Time 2 actual attitude was strongly related to Time 2 desired attitudes, consistent with past work that identifies large correlations between actual/desired attitudes (DeMarree et al., 2014, 2017).

Most critically, we also found a significant interaction of matching condition X political extremity in each dataset. We analysed these interactions by testing the effect of matching condition across levels of political extremity. In Experiment 1a, matching condition had a relatively modest effect on desired attitudes among the least extreme participants, but a substantial effect among very extreme political participants. In Experiment 1b, similarly, matching condition had a negligible effect on desired attitudes among less extreme participants, but a significantly positive effect among more politically extreme participants. This moderation qualifies the mixed main effect of manipulation on desired attitudes by highlighting a boundary condition. That is, political moderates' desired attitudes were minimally affected by the politician's political identity, but politically polarized participants' desired attitudes were significantly shaped by the politician's political identity.

This effect also is depicted in Fig. 2, the two panels displaying the two samples separately. Both figures are very similar. Following political extremity from low (left part of figure) to high (right part of figure), we see that the difference in desired attitudes (and subjective discrepancies) prompted by the matching (versus mismatching) politician passage grows from virtually non-existent into a quite large difference. This pattern reflects the same breakdown narrated above: as political extremity increases, the desired attitudes expressed by participants are increasingly dictated by the political stance of the politician.[3]

Table 1
Moderation of Matching Condition on Time 2 Desired Attitude Formation by Participant Political Extremity. (Exp. 1).

Predictor Variable	Parameter Coefficients (Exp. 1a)	Parameter Coefficients (Exp. 1b)
Matching Condition	$b = 1.00$ [0.65, 1.34], $t(317) = 5.65$, $p < .001$, $r_{sp} = 0.17$	$b = 0.12$ [−0.06, 0.31], $t(682) = 1.29$, $p = .198$, $r_{sp} = 0.03$
Time 1 Actual Attitude	$b = -0.04$ [−0.18, 0.10], $t(317) = -0.54$, $p = .593$, $r_{sp} = -0.02$	$b = -0.004$ [−0.09, 0.08], $t(682) = -0.09$, $p = .932$, $r_{sp} = 0.00$
Time 2 Actual Attitude	$b = 0.64$ [0.57, 0.72], $t(317) = 16.69$, $p < .001$, $r_{sp} = 0.49$	$b = 0.71$ [0.66, 0.77], $t(682) = 26.31$, $p < .001$, $r_{sp} = 0.63$
Participant Political Extremity	$b = -0.01$ [−0.21, 0.19], $t(317) = -0.12$, $p = .903$, $r_{sp} = 0.00$	$b = -0.08$ [−0.20, 0.05], $t(682) = -1.20$, $p = .231$, $r_{sp} = -0.03$
Matching Condition X Participant Political Extremity	$b = 0.82$ [0.38, 1.25], $t(317) = 3.70$, $p < .001$, $r_{sp} = 0.11$	$b = 0.29$ [0.03, 0.55], $t(682) = 2.15$, $p = .032$, $r_{sp} = 0.05$
Matching Condition Simple Slope: Low (−1 SD) Political Extremity	$b = 0.42$ [0.02, 0.82], $t(317) = 2.08$, $p = .039$, $r_{sp} = 0.06$	$b = -0.05$ [−0.28, 0.18], $t(682) = -0.41$, $p = .683$, $r_{sp} = -0.01$
Matching Condition Simple Slope: High (+1 SD) Political Extremity	$b = 1.57$ [1.05, 2.08], $t(317) = 6.00$, $p < .001$, $r_{sp} = 0.18$	$b = 0.32$ [0.05, 0.60], $t(682) = 2.29$, $p = .022$, $r_{sp} = 0.05$
Model Statistics	$F(4, 317) = 172.9$, $p < .001$, $R^2 = 0.73$	$F(5, 682) = 216.6$, $p < .001$, $R^2 = 0.61$

Note. Each row displays the statistics associated with a single independent variable. The left data column displays results for Experiment 1a, the right data column for Experiment 1b. r_{sp} indicates the semi-partial *r* effect size.

4.3.2. Desired attitudes predicting subsequent actual attitude change

We next tested if desired attitudes predicted the actual attitude change that people next experienced. Recall that participants received no new information between reporting their Time 2 and Time 3 actual attitude. This implies that actual attitude change predicted by desired attitudes represents people's self-persuasion, that is, people shifting their actual attitudes to closer match their desired attitudes merely by thinking further about the topic. This would demonstrate that desired attitudes shape people's actual attitude change, even in the absence of new information.

We regressed Time 3 actual attitudes onto Time 1 and Time 2 actual attitudes, and Time 2 desired attitudes. Results for both datasets are displayed in Table 2. The actual attitude measures in these models were included as covariates so that Time 3 actual attitudes captured attitude change. That is, the Time 3 actual attitude dependent variable here indicates the favourability of people's opinions towards the politician controlling for their opinion before engaging in additional reflection. In both experiments (left and right data columns), Time 1 and Time 2 actual attitudes related positively to Time 3 actual attitudes, suggesting that people's earlier attitudes influenced their later attitudes even after engaging in metacognitive reflection. Critically, Time 2 desired attitudes positively predicted Time 3 actual attitudes in both samples. In addition to how much people actually liked the politician, then, people's desire to like her accounted for additional attitude change. People thus pursued their desired attitudes by merely reflecting on their own previous thoughts with a new metacognitive goal.[4]

4.3.3. Quality checks (Exp. 1b only)

In our preregistration, we planned quality checks, re-testing results but removing participants who (i) did not provide enough thoughts ($n = 43$), (ii) revealed suspicion on our probes ($n = 12$), (iii) mentioned having technical issues ($n = 5$), or (iv) did not correctly answer our attention check ($n = 251$). The first three re-analyses did not affect our results, and the analyses remained unchanged: political extremity still moderated passage's effect on desired attitude; desired attitude predicted attitude change. Removing participants who failed the attention check rendered both key analyses non-significant (*p*s = .416, .247, respectively). However, this is likely attributable to losing >1/3 of the sample, likely undermining statistical power to detect our effect. Importantly, of the 251 removed participants, 195 (78%) failed because they left the attention check blank.[5] If we re-run the analyses only removing participants who actively answered the attention check

[3] Separating ideal and ought attitudes slightly affected results. In Experiment 1a, the interaction effect emerged both for ideal attitudes assessed individually, $b = 0.90$ [0.38, 1.41], $t(317) = 3.41$, $p < .001$, and for ought attitudes, $b = 0.74$ [0.24, 1.25], $t(317) = 2.89$, $p = .004$. In Experiment 1b, the interaction effect emerged for ideal, $b = 0.35$ [0.04, 0.67], $t(682) = 2.18$, $p = .030$, but not for ought attitudes, $b = 0.22$ [−0.09, 0.52], $t(682) = 1.41$, $p = .160$

[4] We also examined if participants' behavioral intentions towards the politician showed similar patterns to the attitude change phenomena explained above. Results were very similar to the attitude change analysis both in structure and results, so are detailed in SOM-3.

[5] We are not certain why so many participants left this attention check item blank. One possibility is that participants felt they had already demonstrated adequate attention by answering the suspicion item, and felt this question was therefore redundant.

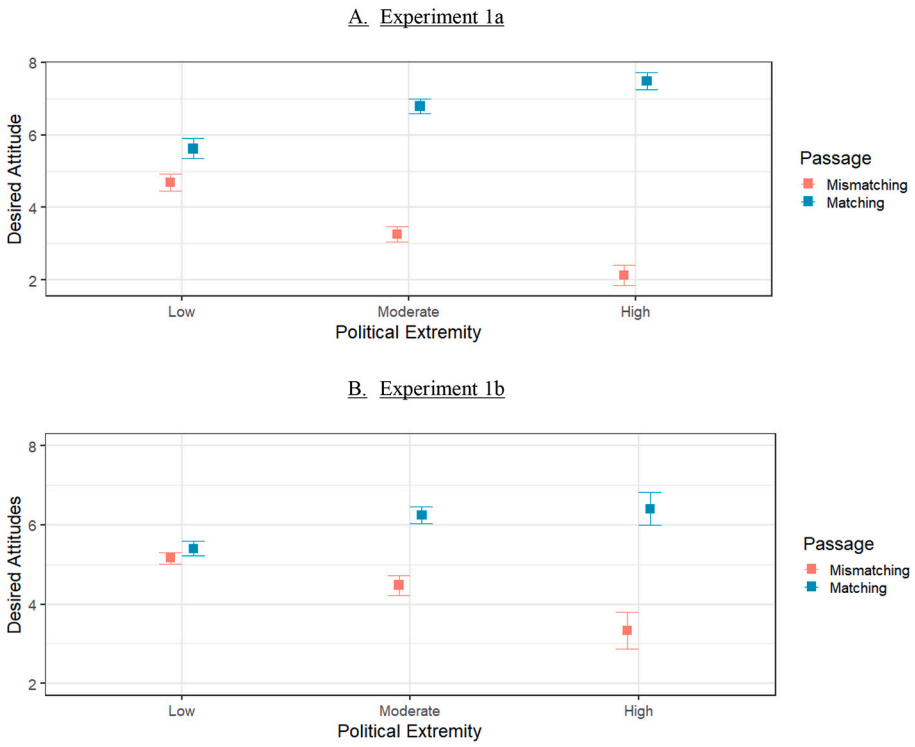

Fig. 2. Desired Attitudes Influenced by Levels of Matching (of Participant with Target Political Beliefs) and Participant Political Extremity.
Note. Errors bars reflect standard error values. Higher scores reflect more positive desired attitudes. Time 1 and Time 2 actual attitudes were controlled in these analyses.

Table 2
Effects of Desired Attitudes on Changes in Actual Attitudes at Time 3. (Experiment 1).

Predictor Variables	Parameter Coefficients (Exp. 1a)	Parameter Coefficients (Exp. 1b)
Time 1 Actual Attitude	$b = 0.22$ [0.14, 0.30], $t(320) = 5.42$, $p < .001$, $r_{sp} = 0.14$	$b = 0.08$ [0.02, 0.13], $t(692) = 2.71$, $p = .007$, $r_{sp} = 0.06$
Time 2 Actual Attitude	$b = 0.40$ [0.34, 0.46], $t(320) = 13.06$, $p < .001$, $r_{sp} = 0.34$	$b = 0.48$ [0.44, 0.53], $t(692) = 20.66$, $p < .001$, $r_{sp} = 0.44$
Time 2 Desired Attitude	$b = 0.15$ [0.09, 0.21], $t(320) = 4.65$, $p < .001$, $r_{sp} = 0.12$	$b = 0.08$ [0.04, 0.13], $t(692) = 3.46$, $p = .001$, $r_{sp} = 0.07$
Model Statistics	$F(3,320) = 391.21$, $p < .001$, $R^2 = 0.79$	$F(3, 692) = 515.01$, $p < .001$, $R^2 = 0.69$

Note. r_{sp} indicates the semi-partial r effect size.

incorrectly, both analyses become significant again ($ps < 0.037$). A full breakdown of results is available in SOM-7.

4.3.4. Changes from the preregistration (Exp. 1b only)

First, in the preregistration, we stated that *t*-tests would be used for some analyses (e.g., effect of matching condition on Time 2 desired attitude), but also that political extremity would be tested as a moderator of these analyses. Ultimately, we foregrounded the moderated version of these analyses (i.e., in Table 1) because lower-order effects such as the main effects of matching condition on desired attitudes are already expressed in those regression models, and are qualified by the significant interaction with political extremity. Second, in the preregistration we stated that the effects of our desired attitudes manipulation (matching condition) would be used to predict Time-3 actual attitudes. Indeed, this test supported our hypothesis, $t(694) = 9.83$, $p < .001$, $d = 0.75$ [0.59, 0.90], with more positive attitudes given a matching ($M = 4.65$, $SD = 1.16$) versus mismatching politician ($M = 3.73$, $SD = 1.30$). However, we instead foregrounded regression models in which Time-2 desired attitudes and Time-1/Time-2 actual attitudes predicted Time-3 actual attitudes simultaneously (i.e., Table 2). Our reasoning was that this analytic approach is more precise because it reveals the unique effects attributable to the manipulation's effect through desired attitudes specifically. But either analysis supports our contentions.

4.4. Discussion

Our first experiments provided an initial empirical demonstration that desired attitudes can shape subsequent attitude change, supporting our claim that people are not only motivated but also capable of altering their opinions in a direction indicated by their desired attitude standards. In essence, participants expressed a desire to hold a particular attitude position, and their subsequent attitude change aligned with that desire. This shows that people at Time 2 had some awareness of the direction that their actual attitudes would shift in by Time 3. Critically, this actual attitude change then occurred even though participants learned nothing new about the object between Time 2 and Time 3, suggesting a surprising self-awareness of participants. It is noteworthy that actual attitude change occurred despite the likely consistency pressure that participants may have felt to not contradict themselves by changing attitudes from Time 2 to Time 3 with no new information provided between these intervals.

Furthermore, we created desired attitudes in these experiments in a manner different from past research: by associating an attitude object with information relevant to people's political identities. Past theorizing suggests that identities can be powerful origins of desired attitudes (Wheeler & DeMarree, 2019), but this idea had not been explicitly tested. Indeed, past research that has manipulated desired attitudes has used manipulations that emphasize the utility of particular viewpoints (DeMarree et al., 2014; DeMarree & Rios, 2014). We build on this past work by showing desired attitudes forming based on broader factors that

relate to people's ideological positions. This broadens our understanding of why people form desired attitudes.

However, as some exploratory follow-up analyses revealed, Experiments 1a/1b were imperfect in that our matching/mismatching paradigm worked more effectively for liberals (who almost universally preferred the liberal over the conservative candidate) than for conservatives (who evaluated the liberal and the conservative candidate more equally). One explanation is that our conservative profile was too extreme given that few of our conservatives (4%/11% in Experiment 1a/1b) self-described as "extreme" conservatives. This asymmetry might have worked against the construct validity of the political matching/mismatching manipulation. Poignantly, this issue would in principle work against our hypothesized pattern. Regardless, subsequent experiments avoid this issue because they expose participants to intensely moralized stimuli that clearly match/mismatch the moral sensibilities of nearly any university student.

A second limitation is that the politician paradigm in Experiment 1 would ideally be extended to other persuasion contexts. We wanted to enhance the construct validity and generalizability of our findings beyond politicians, and therefore introduced a new paradigm in Experiment 2–3. To increase the chance that all participants would be powerfully affected by our manipulation, we employed a more radical approach involving artwork being associated with extreme value-relevant information: association with Nazi victims versus Nazi sympathizers.

5. Experiment 2

In Experiment 2 we continued exploring the attitude change consequences of desired attitudes in a new context involving extreme moral motivations. Whereas Experiment 1a/1b related an attitude object with political beliefs directly related to the attitude object, from Experiment 2 onwards we connected an attitude object (i.e., a painting) with morally provocative information (i.e., that the painter had anti-fascist or pro-fascist sensibilities). Thus, Experiment 2–3's art paradigm is distinct from the political paradigm in two key regards. First, the information manipulating desired attitudes is somewhat less directly connected with the object itself (information about the object's creator, rather than information about the attitude object itself). Second, the information is about highly moralized associations of the object (rather than a commonplace political group identity). Once again, we directly replicated our experiment, reporting the results of these datasets together as Experiments 2a/2b.

5.1. Methods

5.1.1. Participants

For Experiment 2a, we recruited 119 undergraduates from a Canadian University to complete the study online for partial course credit (79.8% women, 20.2% men, 0% non-binary or prefer not to answer; $M_{age} = 19.5$ years, $SD_{age} = 5.2$).[6] For Experiment 2b, we recruited 514 online volunteer American participants from ResearchMatch (74.0% women, 24.0% men, 0.8% non-binary, 1.2% prefer not to answer; 86.7% Caucasian/White, 3.5% African American/Black, 3.1% "mixed," 1.8% Latinx, 1.2% East Asian, 1.4% "other," and 0.6% East Indian; 2.2% did not answer). The latter group was substantially older than all previous samples, $M_{age} = 52.2$ years, and with much greater heterogeneity in age, $SD = 16.4$. In each case, sample sizes were determined by time-based stopping rules, and they permitted us with 80% statistical power to detect effect sizes of $r > 0.25$ and 0.13, respectively. We reasoned that smaller samples compared to Experiment 1 were reasonable given that (i) the manipulation is probably stronger than the political paradigm because it involved highly moralized content (hatred of Nazis) versus mere political preferences (e.g., liberal/conservative dislike of the other party), and because (ii) our design did not require interaction tests, unlike Experiment 1a/1b.

5.1.2. Procedure

The structure of Experiment 2 was highly similar to the politician paradigm. Participants first viewed an abstract-expressionist painting by Arshile Gorky (*Agony*). As before, participants then listed three positive and three negative thoughts, in counterbalanced order, and rated their Time 1 actual attitude regarding the painting. Next, participants were informed that the artist who had created the painting was Bruno Steingard, who ostensibly worked in 1940s Munich, Germany. In the *moral painter* condition Steingard was described as a Nazi resistor, whereas in the *immoral painter* condition Steingard was described as a Nazi sympathizer. We reasoned the painter resisting (supporting) Nazis would provide an intense, morally-grounded desire to form positive (negative) desired attitudes, despite people already having adopted their initial (Time 1) opinions based on the aesthetic merits of the painting.

Having learned this information, participants reported their post-manipulation attitudes towards the painting (1 = *strongly dislike* to 7 = *strongly like*), desired attitude (ideal, ought), and subjective discrepancies. Finally, participants engaged in the same metacognitive prompts that we used in Experiment 1, and then we assessed their Time 3 attitudes towards the painting. Because behavioral intentions showed very similar results to Time 3 actual attitudes in both Experiment 1 samples (see SOM-3), we dropped it to avoid redundancy. We did not ask about participants' political orientations because there is strong evidence that Nazis are roundly despised by most North Americans (e.g., McCarthy, 2019; Vonasch, Reynolds, Winegard, & Baumeister, 2018).

5.1.2.1. Creating indices. The Time 1 actual attitude items were highly inter-correlated in each sample, $r_{Study2a}(115) = 0.73$, $p < .001$, $r_{Study2b}(512) = 0.87$, $p < .001$, so we averaged them into Time 1 attitude indices. Time 2 ideal and ought attitudes were highly correlated in each sample, $r_{Study2a}(115) = 0.82$, $p < .001$, $r_{Study2b}(512) = 0.79$, $p < .001$, so we averaged these to represent desired attitudes.

5.2. Results

5.2.1. Manipulation check: Desired attitudes

In Table 3, we examine several factors related to the formation of desired attitudes towards the painting across the two experimental samples. In Experiment 2a, Time 1 actual attitudes were not significantly related to desired attitudes at Time 2. However, Time 2 actual attitudes were significantly and substantially related to desired attitudes at Time 2. This finding is unsurprising because people likely use their actual attitudes as an anchor for determining their desired attitude although their pre-manipulation attitudes presumably do not serve this same

Table 3
Painter Identity Effects on Desired Attitude Formation. (Experiment 2).

Predictor Variable	Parameter Coefficients (Exp. 2a)	Parameter Coefficients (Exp. 2b)
Painter Identity	$b = 2.37$ [1.68, 3.06], $t(113) = 6.77$, $p < .001$, $r_{sp} = 0.34$	$b = 2.59$ [2.28, 2.89], $t(510) = 16.71$, $p < .001$, $r_{sp} = 0.44$
Time 1 Actual Attitude	$b = -0.21$ [−0.44, 0.03], $t(113) = -1.75$, $p = .083$, $r_{sp} = -0.09$	$b = -0.29$ [−0.43, −0.14], $t(510) = -3.95$, $p < .001$, $r_{sp} = -0.10$
Time 2 Actual Attitude	$b = 0.54$ [0.37, 0.72], $t(113) = 6.24$, $p < .001$, $r_{sp} = 0.32$	$b = 0.63$ [0.53, 0.73], $t(510) = 12.18$, $p < .001$, $r_{sp} = 0.32$
Model Statistics	$F(3,113) = 92.08$, $p < .001$, $R^2 = 0.71$	$F(3, 510) = 319.15$, $p < .001$, $R^2 = 0.65$

Note. r_{sp} indicates the semi-partial r effect size.

[6] We did not include a race item for Experiment 2b, but participants were drawn from a pool of participants with 79.7% White; 13.9% East Asian, 4.7% South Asian, 1.7% Black.

purpose. Examining the effect of painter's identity on desired attitudes, people wanted to like the painting more than two full scale units more when it was made by a Nazi victim rather than a Nazi sympathizer - after adjusting for the actual attitude change caused by this same information. Thus, although people immediately shifted their actual opinions of the painting based on its maker's moral identity, people also desired to like/dislike a painting made by a Nazi victim/sympathizer. Further, these patterns were empirically distinct in that the desired attitude difference held despite controlling for contemporaneous actual attitudes.

We also briefly consider the subjective discrepancy measure to further describe reactions to the manipulation. In Experiment 2a (2b), in the immoral painter condition, 79.3% (66.0%) wanted a more negative opinion than they held, 10.3% (28.2%) wanted to hold their current attitude, and 10.3% (5.8%) wished to hold a more positive attitude. In the moral painter condition, by contrast, 61.0% (60.4%) wanted a more *positive* opinion, 32.2% (31.4%) wanted to hold their current attitude, and only 6.8% (8.2%) wanted a more negative opinion.

5.2.2. Desired attitudes predicting subsequent actual attitude change

We once again regressed Time 3 attitudes onto Time 1 and Time 2 actual attitudes, and (Time 2) desired attitudes (see Table 4). Both Time 1 and Time 2 actual attitudes related positively to Time 3 actual attitudes, although the standardized effect size of Time 2 actual attitudes was double that of Time 1 actual attitudes. As in Experiments 1a-1b, this makes sense primarily because Time 2 actual attitudes reflect opinions at a moment closer to the Time 3 measurement, and here they reflect opinions after (versus before) learning the painter's identity.

Most importantly to our research questions, we examined whether desired attitudes were able to predict actual attitude change. Indeed, desired attitudes were strongly related to actual attitude change after participants had a chance to query their previous thoughts based on their desired attitude goal. Compared to an average participant, each additional unit change in desired attitude corresponded to a 0.3 (0.2) unit change in actual attitudes in Experiment 2a (Experiment 2b). Thus, people formed particular opinions about a painting initially based on its aesthetic attributes. But then they self-persuaded to an actual attitude more congruent with their wished-for actual attitude, even without any access to novel information about the painting itself. Thus, the actual attitude change they eventually engaged in (by Time 3) was predictable based on the attitude they considered desirable by Time 2. This finding is consistent with Experiments 1a/1b but in a substantially different context. Specifically, the moral identity of a painter impacted views of a painting, rather than the political position of a politician impacting views of the politician herself.

5.3. Discussion

Experiment 2 provided novel support for our hypotheses about the formation of desired attitudes, and the effects of desired attitudes on self-persuasion, but in a very different, heavily moralized, context. Once again, associating an attitude object with identity-relevant information with minimal utilitarian implications prompted people to form desired attitudes towards that object, above and beyond the manipulation's immediate effect on their actual attitudes. Interestingly, although the balance of desired/actual attitude discrepancies fell in the expected directions, some participants (10.3%–32.2% depending on sample and condition) felt no desire to hold a different attitude than their current one after exposure to the identity manipulation. However, just as in the prior studies, desired attitudes were predictive of subsequent attitude change towards the object after a period of cognitive reflection. Thus, although not everyone desired to hold a different attitude than their actual one, the direction and magnitude of their desired attitudes were predictive of their subsequent actual attitude change. Experiment 2 also enhances the generalizability of our findings by extending them to a substantively different context.

Table 4
Effects of Desired Attitudes on Actual Attitude Change (Experiment 2).

Predictor Variables	Parameter Coefficients (Exp. 2a)	Parameter Coefficients (Exp. 2b)
Time 1 Actual Attitude	$b = 0.24$ [0.04, 0.44], $t(113) = 2.39$, $p = .019$, $r_{sp} = 0.12$	$b = 0.27$ [0.15, 0.38], $t(510) = 4.59$, $p < .001$, $r_{sp} = 0.12$
Time 2 Actual Attitude	$b = 0.41$ [0.25, 0.57], $t(113) = 4.95$, $p < .001$, $r_{sp} = 0.25$	$b = 0.46$ [0.37, 0.55], $t(510) = 9.70$, $p < .001$, $r_{sp} = 0.25$
Time 2 Desired Attitude	$b = 0.31$ [0.18, 0.45], $t(113) = 4.74$, $p < .001$, $r_{sp} = 0.24$	$b = 0.19$ [0.13, 0.24], $t(510) = 6.46$, $p < .001$, $r_{sp} = 0.16$
Model Statistics	$F(3, 116) = 95.86$, $p < .001$, $R^2 = 0.72$	$F(3, 510) = 342.43$, $p < .001$, $R^2 = 0.67$

Note. r_{sp} indicates the semi-partial r effect size.

The two datasets provide a strong demonstration of our key claims and to some extent address one another's limitations. For example, Experiment 2a had a relatively small sample size, making our statistical tests less precise than they would preferably be, but Experiment 2b provided a much larger sample size for highly precise estimates. Furthermore, the two samples were comprised of almost exclusively young adults (Experiment 2a) versus a broader sampling of participants (Experiment 2b). Thus, replicating the findings very consistently in Experiment 2b's larger, older sample enhanced certainty in our core findings.

6. Experiment 3

The first four datasets provide consistent support for our core hypothesis that desired attitudes shape actual attitude change even without requiring additional information. However, all studies encouraged people to reflect on their past thoughts through several metacognitive prompts. The prompts were included to maximize the likelihood of attitude change occurring after formation of desired attitudes. However, such direct external prompting may not be necessary for desired attitudes to produce change presuming that the motivational drives underpinning people's desired attitudes are sufficiently strong. Indeed, the heavily moralized context used in Experiments 2a/2b was created to produce just such a situation. Thus, it would also be useful to know if the attitude change effects depended on participants being led to engage in metacognitive reflection. Our primary goal in Experiment 3 was to provide conditions in which prompts were not provided, to test if desired attitudes could still prompt subsequent attitude change without any such encouragement. This would help to generalize the phenomenon by demonstrating that people's desired attitudes can still predict their subsequent attitude change without the somewhat contrived stimulation of additional thinking.

6.1. Methods

6.1.1. Participants

We recruited 210 participants from a Canadian university (9 removed for incomplete data). Of the remaining 201, most were young adults ($M_{age} = 18.7$, $SD_{age} = 1.6$), and most were women (72.1%, with 25.9% men, 0% non-binary, 2.0% prefer not to answer). Of these, 78.6% were Caucasian/White, 11.9% East Asian, 3.0% African American/Black, 3.5% East Indian, 2.5% Indigenous, 0.5% Latinx; 5.5% "other." Sample size followed a time-based stopping rule, giving us with 80% statistical power to detect effect sizes of $r \geq 0.19$.

6.1.2. Procedure

The procedure matched Experiment 2a/2b, with one exception: half of participants were given the metacognitive prompts and asked to review their thoughts (as described in the Experiment 1 procedure), whereas the other half were not given these prompts. Thus, Experiment 3 had a 2 (Painter Identity: Moral vs Immoral) X 2 (Order: Positive

thoughts first vs Negative thoughts first) X 2 (Metacognitive Prompts: Present vs Absent) between-participants design. Thus, those in the Metacognitive Prompts Absent condition learned the painter's identity, recorded attitudes and desired attitudes at Time-2, considered attitude/desired attitude discrepancies (see SOM-4), and then reported Time-3 actual attitudes. Order had no main effects or interactive effects with other study variables, so we do not discuss it further.

6.1.2.1. Creating indices. The Time 1 actual attitude items were intercorrelated, $r(199) = 0.72$, $p < .001$, so we averaged them into a Time 1 attitude index as in prior experiments. Time 2 ideal and ought attitudes correlated, $r(199) = 0.75$, $p < .001$; their average represented desired attitudes.

6.2. Results

6.2.1. Manipulation check: Desired attitudes

As shown in Table 5, Time 1 actual attitudes were negatively related to desired attitudes; a likely explanation is that this pattern is due to Time 2 actual attitudes also being in the model and correlating moderately with Time 1 actual attitudes, $r(199) = 0.34$, $p < .001$. Indeed, this association vanishes when examined as a zero-order correlation, $r(199) = 0.06$, $p = .422$. Thus, the 'negative' effect of Time 1 actual attitudes on (Time 2) desired attitudes reflects only Time 1 actual attitudes' residual variance after accounting for the effect of the more proximal Time 2 actual attitudes. Replicating all past studies, more positive Time 2 actual attitudes were substantially related to positive desired attitudes at Time 2. A substantial main effect of painter identity directly replicates Experiment 2. Like our prior experiments, more positive desired attitudes were formed given a moral versus immoral artist.

6.2.2. Actual attitude change

We next turned to the critical test of Experiment 3: an interaction effect between desired attitudes and the presence/absence of metacognitive prompts. We were primarily interested in whether the simple slope of desired attitudes on Time 3 actual attitudes remained significant in the prompt-absent condition specifically. If so, this would suggest that prompting participants was not necessary for actual attitude change. Results are noted in Table 6. First, positive main effects of Time 1 and Time 2 actual attitude were detected, which simply indicate that people who initially liked the painting also liked it later. We also replicated the main effect of desired attitudes, suggesting that aggregating across metacognitive prompt conditions, a typical participant ended up liking (disliking) the painting more when they had earlier desired to like (dislike) it more. Most critically, the interaction effect is non-significant. To probe our main question, we broke down the slopes of desired attitudes predicting actual attitude change at each level of metacognitive prompting. Because the interaction test is non-significant, there is not sufficient evidence to conclude that these slopes are different from one another. Desired attitudes predicted actual attitude change whether metacognitive prompts were present, $b = 0.42$ [0.30, 0.53], $t(195) = 7.29$, $p < .001$, $r_{sp} = 0.29$, or absent, $b = 0.33$ [0.21, 0.44], $t(195) = 5.38$, $p < .001$, $r_{sp} = 0.18$. This strongly supports our contention that our decision to include metacognitive prompts was not a necessary condition for people to self-persuade in a direction prognosticated by their desired attitudes.[7]

Table 5
Painter Identity Effects on Desired Attitude Formation Controlling for Prior Actual Attitude. (Experiment 3).

Predictor Variable	Parameter Coefficients
Painter Identity	$b = 2.31$ [1.82, 2.81], $t(197) = 9.23$, $p < .001$, $r_{sp} = 0.37$
Time 1 Actual Attitude	$b = -0.20$ [−0.38, −0.03], $t(197) = -2.29$, $p = .023$, $r_{sp} = -0.09$
Time 2 Actual Attitude	$b = 0.59$ [0.46, 0.71], $t(197) = 9.07$, $p < .001$, $r_{sp} = 0.36$
Model Statistics	$F(3, 197) = 145.30$, $p < .001$, $R^2 = 0.69$

Note. r_{sp} indicates the semi-partial r effect size.

Table 6
Effects of Desired Attitudes on Actual Attitude Change (Experiment 3).

Independent Variables	Parameter Coefficients
Time 1 Actual Attitude	$b = 0.21$ [0.06, 0.35], $t(195) = 2.85$, $p = .005$, $r_{sp} = 0.12$
Time 2 Actual Attitude	$b = 0.35$ [0.22, 0.47], $t(195) = 5.51$, $p < .001$, $r_{sp} = 0.23$
Time 2 Desired Attitude	$b = 0.37$ [0.27, 0.47], $t(195) = 7.57$, $p < .001$, $r_{sp} = 0.31$
Metacognitive Prompts	$b = -0.01$ [−0.33, 0.31], $t(195) = -0.05$, $p = .957$, $r_{sp} = 0.00$
Time 2 Desired Attitude X Metacognitive Prompts	$b = 0.09$ [−0.04, 0.22], $t(195) = 1.39$, $p = .165$, $r_{sp} = 0.06$
Model Statistics	$F(5, 195) = 78.42$, $p < .001$, $R^2 = 0.67$

Note. r_{sp} indicates the semi-partial r effect size.

6.3. Discussion

Experiment 3 examined one objection that might be raised regarding the prior experiments through a single procedural variation: half of participants undertook the metacognitive prompts (as in earlier experiments) whereas another half did not complete these prompts. This addresses the possibility that the self-persuasion effects documented in Experiments 1–2 (people's actual attitude change tending to follow their desired attitude position) occurred only because we facilitated these effects with our metacognitive prompts. Strongly contradicting this counter-explanation, our effect emerged significantly ($ps < 0.001$) even when all metacognitive prompting was removed. These findings might suggest that people engage in metacognitive reflections without prompting, or might suggest that people in this condition used tactics other than metacognitive reflection to self-persuade. Maio and Thomas (2007) provide examples of such processes, of which some (e.g., teleologic strategies) might be more easy to complete in the narrow window of opportunity provided in the Metacognitive Prompts Absent condition (e.g., concentrating on information facilitating pursuit of the desired attitude, e.g., Robbins, 1991) without using more difficult, epistemically-driven self-justifications.

What is less clear is whether metacognitive prompting at least contributed to more attitude change, given that our focal interaction of prompting X desired attitudes on actual attitudes was non-significant, but a significant interaction emerged for prompting X ideal attitude. Although not a central issue for our present case, a larger sample size might have generated a significant interaction, which would be interesting from the perspective of understanding how these self-persuasion effects emerge most strongly. In sum, Experiment 3 suggests that given a sufficiently strong motivational foundation for desired attitudes, people may self-persuade towards the position of their desired attitudes even without any prompting.

[7] Ideal and ought attitudes produced somewhat different results when analysed separately. Ideal attitudes X metacognitive prompts produced an interaction effect, $b = 0.14$ [0.01, 0.26], $t(195) = 2.18$, $p = .030$, $r_{sp} = 0.09$, such that ideal attitudes predicted greater attitude change when prompts were included, $b = 0.37$ [0.26, 0.47], $t(195) = 6.94$, $p < .001$, $r_{sp} = 0.29$, versus when they were absent, $b = 0.23$ [0.13, 0.33], $t(195) = 4.41$, $p < .001$, $r_{sp} = 0.19$. Ought attitudes X metacognitive prompts led to no such interaction, $b = 0.06$ [−0.07, 0.18], $t(195) = 0.88$, $p = .382$, $r_{sp} = 0.04$. Supporting our main contention, however, desired attitudes, whether measured as "ideal" or "ought," robustly predicted attitude change in all analyses/breakdowns.

7. General discussion

Five datasets provided consistent evidence that desired attitudes have important consequences for the actual attitude change that people undergo. Desired attitudes can be understood as attitude "guides." That is, people seem able to persuade themselves towards their desired attitude positions, even in the absence of compelling new information about the object (DeMarree et al., 2017, Study 2), and even without the ability to modify the object to align it with their desires (DeMarree et al., 2017, Study 4). Indeed, although most of our experiments allowed participants to engage in metacognitive reflection about their thoughts (Experiments 1–2), Experiment 3 revealed that such encouragement at most helped but was certainly not necessary for participants to self-persuade towards their desired attitudes.

7.1. Theoretical implications

7.1.1. Desired attitudes' consequences for actual attitude change

The present experiments shed light on several gaps in the growing literature on desired attitudes (for a review, see Wheeler & DeMarree, 2019). Most crucially, our work addresses whether desired attitudes can direct actual attitude change. The experiments provided consistent evidence that desired attitudes indeed predict actual attitude change. Thus, in response to a situational pressure to like or dislike an object, people immediately change their actual attitudes. However, the information may also induce a separate *desire* to like/dislike the object, which prognosticates the actual attitude they will reach after a few minutes of deliberation. Previous models have posited that people direct their own attitude change towards desired attitudes (e.g., Maio & Thomas, 2007) but here we directly demonstrate this effect. We show this result in two distinct ways: (1) in the main manuscript, we show that Time-2 desired attitudes predict Time-3 actual attitudes adjusting for Time-2 actual attitudes; and (2) in the supplementary SOM-4, people's explicit comparisons (how much more/less they want to like the object, compared to how much they actually do) replicate these patterns.

Reviews often conclude that variables often have only small and unreliable effects in shaping attitude change, even in laboratory contexts that afford substantial experimental control ($r = 0.10$–0.11; Albarracin & Shavitt, 2018; Tannenbaum et al., 2015). Despite this, we found that desired attitudes consistently could predict subsequent attitude change. Furthermore, this actual attitude change was detected in contexts where people were not provided any additional opportunity to attain actual attitude change by obtaining a biased collection of new information or to engage in biased processing of additional information (see DeMarree et al., 2017); and participants could not modify objective properties of the object to improve it (see Wheeler & DeMarree, 2019). A skeptical reader may reason that between Time 2 and Time 3, in the present paradigms, participants were continuing to react to the identity-relevant information from just before the Time 2 attitude ratings. However, the consistent pattern of desired attitudes predicting the direction and magnitude of that delayed reaction nonetheless remains interesting, suggesting that those delayed reactions to the information are motivated, in that they literally are predictable from participants' stated desires. This extended self-persuasion even occurred when we provided them with no help at all, that is, when we ceased to even ask them to reflect on their thoughts (prompt-absent condition in Experiment 3). Despite all these barriers, people were robustly able to self-persuade in a direction forecasted by their desired attitudes ($rs = 0.07$–0.31 across five datasets). In short, beyond the previously captured consequences of desired attitudes (that they shape people's pursuit of new information, or physical modifications of objects), mere thinking can facilitate people's changing their opinions towards their desired attitude positions.

Interestingly, then, our results suggest an important degree of self-awareness among participants. Whereas past scholars have been clear that people are not aware of their self-persuasion processes (e.g., that they will become more attitudinally extreme merely by thinking more; Tesser, 1978; also see Dijksterhuis, 2004), we show that people can be consciously aware of the direction that their future actual attitude change will move in. This is not as paradoxical as it may sound. It has long been recognized that people being unaware of their attitude change processes/tendencies (i.e., towards polarization; Tesser, 1978) does not entail that they are unaware of what attitudes they hold (Nisbett & Wilson, 1977). Similarly, people may be quite aware of their attitudinal goals (i.e., desired attitudes), which in fact may effectively predict the direction of their subsequent attitude change (shown in all five present experiments), even if they could not explain how they talked themselves into that attitude change. Indeed, this identification of desired attitude goals may be quite instrumental to self-persuasion. Many self-directed attitude change strategies (Maio & Thomas, 2007) and thought control strategies (Wenzlaff & Wegner, 2000) are cognitively demanding, so it makes sense that people's organization of effective and proportionate self-persuasion (e.g., to avoid "over-correcting," e.g., Sommers & Kassin, 2001) would benefit from a clear goal-state (i.e., "I want to like the painting *slightly* [versus *moderately*; *extremely*] more.").

One potential objection involves our examination of Time 2 desired attitudes as predictors of Time 3 attitudes (in Experiments 1–3), in which we controlled for Time 2 actual attitudes. Like all measures, attitude scales are prone to error – hence, the Time 2 actual attitude measure did not capture all variance in opinions at Time 2. If we failed to adequately control for Time 2 actual attitudes, Time 2 desired attitudes might have only appeared to predict Time 3 actual attitudes, but the effect might have been driven by Time 2 desired attitudes' overlap with Time 2 actual attitudes (insufficiently controlled because Time 2 actual attitudes were imperfectly measured; see Bollen, 1989). However, several arguments render this inadequate partialling objection less compelling. First, this measurement error criticism is not specific to the present work and is applicable to any analysis that includes a covariate, or partials for variance in one variable when examining the influence of a second variable (ANCOVAs, regressions, etc.). Our studies are no more susceptible to this problem than the large part of the scientific literature that makes claims based on this broad range of analyses. Second, our attitude measures were generally very high in internal consistency (inter-item correlations $rs > 0.70$; $\alpha s > 0.90$), making it less likely that the existence of random error introduced substantial distortions to our estimates of effects and that our attitude change effects were thus purely reflective of incomplete partialling of effects. Finally, we conducted supplementary analyses using structural equation modeling to further assess the potential impact of measurement error. In these analyses, we specified actual and desired attitudes as latent variables so that we could estimate relations among these measures after having removed the effects of random error. These analyses did not produce any substantive differences in our findings, with all critical effects remaining significant (SOM-6).

The above arguments mostly assume random measurement error. One might instead object that our briefer Time 1 and Time 2 actual attitude measures, combined with the longer Time 3 actual attitude measure, artificially inflated actual attitude change at Time 3 through systematic measurement error. For instance, if participants felt relatively constrained to give similar responses to the "like"/"likeable" items at Time 1 and Time 2, the Time 3 measure's novel items (e.g., "positive") might have allowed participants to express attitude change that really occurred earlier. This interpretation implies that the actual attitude change should largely evaporate if actual attitudes at Time 3 are measured with just the "like" item. We re-analysed the self-persuasion with this substitution, using an integrative dataset that combined all of our datasets. As predicted, desired attitudes still significantly predicted actual attitude change. Therefore, the attitude change observed at Time 3 in our studies cannot be explained away by the final attitude measure introducing new items.

One alternative interpretation of our findings involves attitude polarization from mere thought (e.g., Tesser, 1978). After receiving the matching/mismatching information about the politician (Time 2),

people responded by immediately changing their actual attitudes to be positive/negative, respectively. A polarization explanation might argue that later, given more time to think (i.e., by Time 3), people might have polarized these actual attitudes due to the mere thought phenomenon, with matching (mismatching) people becoming more positive (negative) in their views. Desired attitudes might have shifted more quickly to be consistent with participants' cognitive schemas (Tesser, 1978) about political groups, with actual attitudes taking longer to become schema-consistent. Desired attitudes then predicted later attitude change because they changed more extremely in the same direction that actual attitudes would later shift. However, the data is not in line with this explanation. Although Time 2 desired attitudes shifted more strongly than did Time 2 actual attitudes in some data sets (Experiment 2–3), Time 2 desired attitudes changed *less* strongly than did Time 2 actual attitudes in other data (Experiment 1b), and there was no difference in yet others (Experiment 1a). Thus, established processes of attitude polarization cannot consistently account for our effects. We provide complete details in SOM-5.

Furthermore, it would be beneficial to replicate our effects among other populations. Our experiments generalized effectively across Canadian and American University students, as well as a broader (and older) sample of American volunteers on ResearchMatch. Thus, the modest diversity of sampling we employed failed to undermine our effects, and we are optimistic that our effects would emerge in other types of adult populations.

7.1.2. Identity-relevant desired attitudes

Additionally, both the political (Experiment 1) and moral (Experiment 2–3) paradigms supported some conjectures about the origins of desired attitudes. That is, past scholarship raised the idea that consistency with a person's social identities and values might be possible sources of desired attitudes (DeMarree et al., 2014; Wheeler & DeMarree, 2019), but no empirical work has investigated these links directly. The present studies provide experimental evidence that links between an attitude object and a person's social identity or relevant values does, in fact, affect people's desired attitudes. Thus, just as people form actual attitudes for a variety of motivated reasons (Herek, 1986, 1987; Katz, 1960; Petty & Wegener, 1998; Shavitt, 1990; Smith, Bruner, & White, 1956; Snyder & DeBono, 1985), both utilitarian (DeMarree et al., 2017) and social/value relevant bases (the present work) can shape desired attitude formation.

7.2. Limitations and future directions

Much work on desired attitudes (DeMarree et al., 2014, 2017; the present work) focuses on situations where actual and desired attitudes diverge, and how people then reconcile those differences. This research has successfully identified numerous objects (e.g., the self, African-Americans, exercise) for which most people experience discrepancies, suggesting that people cannot always reconcile actual and desired attitudes. Thus, the phenomenon we have identified here clearly is not in universal operation, or else actual/desired discrepancies would always eliminate themselves. The present paradigms were designed specifically to capture cases where desired attitudes have the greatest potential to predict actual attitude change. We did this by ensuring that actual attitudes were formed without deep connections to pre-existing values, identities, or attitudes (e.g., an unfamiliar painting) but created desired attitudes richly connected to existing attitudes (e.g., hatred of Nazis). Future research should address when desired attitudes do (not) predict actual attitude change. Effects like ours should primarily be anticipated when desired attitudes are "stronger" than actual attitudes. Additionally, predictions may be drawn from the cognitive dissonance literature, in which conflicting cognitive elements may be resolved via a range of processes including trivialization (Simon et al., 1995) or bolstering (Sherman & Gorkin, 1980). For example, if one's actual attitude is very strong, and one encounters information entailing an opposite-valence desired attitude, people may trivialize the new information (e.g., form a weak desired attitude) or bolster their original actual attitude, instead of changing their actual attitude to match the desired attitude. Returning this to our opening example, people with especially strong positive opinions of Rowling's books may trivialize her disliked Tweets (avoiding forming a negative desired attitude towards her books).

This mention of process also brings us to a second important limitation: we did not directly provide a mechanism of how actual attitudes are changing in the present work, only that desired attitudes predict such change. Maio and Thomas (2007) work on self-persuasion may provide guidance here. Their epistemic-teleologic model maps two types of process by which people change their own opinions: epistemic (attempting to form new valid attitude) and teleologic processes (attitude change with minimal concerns for validity). For instance, epistemic tactics include motivated interpretation (e.g., shifting judgment of a negative attribute so that it may be viewed as a positive), whereas teleologic tactics include distraction and suppression (i.e., putting undesired information out of awareness). Importantly, Maio and Thomas (2007) posited that deliberate self-persuasion is activated when attitude-holders recognize their own actual/desired attitude discrepancies. Thus, their proposed processes may guide future research that shows how people accomplish the attitudinal shifts that we have documented. Their model also suggests that far from those more specific processes "replacing" desired attitudes as an explanation, these variables should work in tandem. That is, contextual features elicit desired attitudes discrepant from actual attitudes. These desired attitudes then guide the use of epistemic and teleologic processes that permit people to shift their actual attitudes (also see Rucker, Preacher, Tormala, & Petty, 2011, for a discussion of reconciling multiple mediators).

One methodological limitation is that although in Experiment 3 we eliminated all metacognitive prompts for half of participants, these participants nonetheless completed the subjective discrepancy measure (see SOM-5). It is possible that this single item, which asks participants whether (or not) they want to hold a more positive or negative attitude than they currently do, might have made this discrepancy more salient and accessible to participants. This could be important because ambivalent attitudinal elements may be more psychologically impactful for people when they are high (vs low) in accessibility (Newby-Clark, McGregor, & Zanna, 2002).

In many contexts actual and desired attitudes will align, and the present findings represent several circumstances in which people's actual attitudes shift in directions indicated by their desired attitudes. One interesting possibility is that such actual/desired attitude agreement may itself be important. For example, actual attitudes that do shift to match desired attitudes may confer strength properties (Krosnick & Petty, 1995) to the actual attitude, for example by reducing the subjective ambivalence associated with that attitude (DeMarree et al., 2014). Thus, stimulating people to align their actual and desired attitudes, or making salient such existing overlaps, may bolster actual attitude strength. On the other hand, desired attitudes also may vary in strength properties; that is, that some people may have more relatively more important, certain, and/or accessible (strong) desired attitudes, relative to others (also see discussion of desired attitude *commitment*; DeMarree et al., 2017). Potentially, both actual *and* desired attitudes may be weakened (strengthened) when the valence of those attitudes conflict (agree). Past evidence already hints at this phenomenon for actual attitudes at least (DeMarree et al., 2014, 2016), but such effects may be bolstered by making both types of attitudes salient to people, and dampened by not making each attitude salient. Thus, future research on actual and desired attitudes may stimulate richer understanding of both constructs, and how they relate in nuanced and interesting ways.

Open practices

For all experiments, we report how we determined our sample size, all data exclusions, all manipulations, and all measures. Verbatim

materials are included in the Supplementary Online Materials (SOM-1), and data/syntax are available (https://tinyurl.com/desiredattitude). Experiment 1b was preregistered: (https://osf.io/7wgrh/?view_only=e85ef18892d64615afe1b9cf20852bbc).

Data access

All data have been made available to reviewers at https://tinyurl.com/desiredattitude, and will be made publicly available in the event of publication.

Data availability

Data has been provided as part of Open Practices section.

Acknowledgements

This research was supported by a graduate fellowship (767-2018-1484) to the first author and an Insight Grant (435-2015-0114) to the second author provided by the Social Sciences and Humanities Research Council of Canada.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jesp.2022.104437.

References

Albarracin, D., & Shavitt, S. (2018). Attitudes and attitude change. *Annual Review of Psychology, 69*(1), 299–327.
Aronson, E. (1999). The power of self-persuasion. *American Psychologist, 54*(11), 875–884.
Benoit, W. L., & Dorries, B. (1996). Dateline NBC's persuasive attack on Wal-Mart. *Communication Quarterly, 44*(4), 463–477.
Blankenship, K. L., Wegener, D. T., & Murray, R. A. (2012). Circumventing resistance: Using values to indirectly change attitudes. *Journal of Personality and Social Psychology, 103*(4), 606–621.
Bollen, K. A. (1989). A new incremental fit index for general structural equation models. *Sociological Methods & Research, 17*, 303–316.
Briñol, P., McCaslin, M. J., & Petty, R. E. (2012). Self-generated persuasion: Effects of the target and direction of arguments. *Journal of Personality and Social Psychology, 102*(5), 925–940.
Briñol, P., Petty, R. E., Stavraki, M., Lamprinakos, G., Wagner, B., & Díaz, D. (2018). Affective and cognitive validation of thoughts: An appraisal perspective on anger, disgust, surprise, and awe. *Journal of Personality and Social Psychology, 114*(5), 693–718.
Bruchmann, K., Koopmann-Holm, B., & Scherer, A. (2018). Seeing beyond political affiliations: The mediating role of perceived moral foundations on the partisan similarity-liking effect. *PLoS One, 13*(8), Article e0202101.
Byrne, D. (1961). Interpersonal attraction as a function of affiliation need and attitude similarity. *Human Relations, 14*(3), 283–289.
Carrera, P., Caballero, A., Muñoz, D., & Fernández, I. (2017). Abstractness leads people to base their behavioral intentions on desired attitudes. *Journal of Experimental Social Psychology, 70*, 27–33.
Carver, C. S., & Scheier, M. F. (1998). *On the self-regulation of behavior*. Cambridge University Press.
Clark, J. K., Wegener, D. T., & Fabrigar, L. R. (2008). Attitudinal ambivalence and message-based persuasion: Motivated processing of proattitudinal information and avoidance of counterattitudinal information. *Personality and Social Psychology Bulletin, 34*(4), 565–577.
Clarkson, J. J., Tormala, Z. L., & Leone, C. (2011). A self-validation perspective on the mere thought effect. *Journal of Experimental Social Psychology, 47*(2), 449–454.
Cooper, J. (2007). *Cognitive dissonance: Fifty years of a classic theory*. Sage.
Crites Jr., S. L., Fabrigar, L. R., & Petty, R. E. (1994). Measuring the affective and cognitive properties of attitudes: Conceptual and methodological issues. *Personality and Social Psychology Bulletin, 20*(6), 619–634.
DeMarree, K. G., Clark, C. J., Wheeler, S. C., Briñol, P., & Petty, R. E. (2017). On the pursuit of desired attitudes: Wanting a different attitude affects information processing and behavior. *Journal of Experimental Social Psychology, 70*, 129–142.
DeMarree, K. G., & Rios, K. (2014). Understanding the relationship between self-esteem and self-clarity: The role of desired self-esteem. *Journal of Experimental Social Psychology, 50*, 202–209.
DeMarree, K. G., Rios, K., Randell, J. A., Wheeler, S. C., Reich, D. A., & Petty, R. E. (2016). Wanting to be different predicts nonmotivated change: Actual–desired self-discrepancies and susceptibility to subtle change inductions. *Personality and Social Psychology Bulletin, 42*(12), 1709–1722.
DeMarree, K. G., Wheeler, S. C., Briñol, P., & Petty, R. E. (2014). Wanting other attitudes: Actual–desired attitude discrepancies predict feelings of ambivalence and ambivalence consequences. *Journal of Experimental Social Psychology, 53*, 5–18.
Dijksterhuis, A. (2004). Think different: The merits of unconscious thought in preference development and decision making. *Journal of Personality and Social Psychology, 87*(5), 586–598.
Downing, J. W., Judd, C. M., & Brauer, M. (1992). Effects of repeated expressions on attitude extremity. *Journal of Personality and Social Psychology, 63*(1), 17–29.
Duggan, J. (2022). Transformative readings: Harry potter fan fiction, trans/queer reader response, and JK Rowling. *Children's Literature in Education, 53*(2), 147–168.
Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. Harcourt.
Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149–1160.
Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press.
Graham, J., Haidt, J., & Nosek, B. (2009). Liberals and conservatives use different sets of moral foundations. *Journal of Personality and Social Psychology, 96*, 1029–1046.
Harmon-Jones, E., & Mills, J. (2019). An introduction to cognitive dissonance theory and an overview of current perspectives on the theory. In E. Harmon-Jones (Ed.), *Cognitive dissonance: Reexamining a pivotal theory in psychology* (2nd ed., pp. 3–24). American Psychological Association.
Heider, F. (1946). Attitudes and cognitive organization. *Journal of Psychology, 21*, 107–112.
Herek, G. M. (1986). The instrumentality of attitudes: Toward a neofunctional theory. *Journal of Social Issues, 42*(2), 99–114.
Herek, G. M. (1987). Can functions be measured? A new perspective on the functional approach to attitudes. *Social Psychology Quarterly*, 285–303.
Higgins, E. T. (1987). Self-discrepancy: A theory relating self and affect. *Psychological Review, 94*, 319–340.
Higgins, E. T. (1989). Self-discrepancy theory: What patterns of self beliefs cause people to suffer? In L. Berkowitz (Ed.), *Vol. 22. Advances in experimental social psychology* (pp. 93–136). Academic Press.
Hogg, M. A., & Ridgeway, C. L. (2003). Social identity: Sociological and social psychological perspectives. *Social Psychology Quarterly*, 97–100.
Houston, D. A., & Fazio, R. H. (1989). Biased processing as a function of attitude accessibility: Making objective judgments subjectively. *Social Cognition, 7*(1), 51–66.
Katz, D. (1960). The functional approach to the study of attitudes. *Public Opinion Quarterly, 24*(2), 163–204.
Krosnick, J. A., & Petty, R. E. (1995). Attitude strength: An overview. In R. E. Petty, & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 1–24). Erlbaum.
Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology, 37*(11), 2098–2109.
Maio, G. R., & Haddock, G. (2009). *The psychology of attitudes and attitude change*. Sage.
Maio, G. R., & Thomas, G. (2007). The epistemic-teleologic model of deliberate self-persuasion. *Personality and Social Psychology Review, 11*(1), 1–22.
McCarthy, M. (2019). If Nazi = Red, and Canadian = Red, does Red = Good or Bad? Testing the effects of valenced perceptual cues on Implicit Association Test performance. Retrieved from: https://psyarxiv.com/fsjba/.
Newby-Clark, I. R., McGregor, I., & Zanna, M. P. (2002). Thinking and caring about cognitive inconsistency: When and for whom does attitudinal ambivalence feel uncomfortable? *Journal of Personality and Social Psychology, 82*(2), 157–166.
Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*(3), 231–259.
Ostrom, T. M. (1989). Interdependence of attitude theory and measurement. In A. R. Pratkanis, S. J. Breckler, & A. G. Greenwald (Eds.), *Attitude structure and function* (pp. 11–36). Lawrence Erlbaum.
Petty, R. E., Briñol, P., & Tormala, Z. L. (2002). Thought confidence as a determinant of persuasion: The self-validation hypothesis. *Journal of Personality and Social Psychology, 82*(5), 722–741.
Petty, R. E., & Cacioppo, J. T. (1981). *Attitudes and persuasion: Classic and contemporary approaches*. Brown.
Petty, R. E., & Wegener, D. T. (1998). Matching versus mismatching attitude functions: Implications for scrutiny of persuasive messages. *Personality and Social Psychology Bulletin, 24*(3), 227–240.
Petty, R. E., Wegener, D. T., & Fabrigar, L. R. (1997). Attitudes and attitude change. *Annual Review of Psychology, 48*(1), 609–647.
Requero, B., Briñol, P., & Petty, R. E. (2021). The impact of hope and hopelessness on evaluation: A meta-cognitive approach. *European Journal of Social Psychology, 51*(2), 222–238.
Robbins, A. (1991). *Awake the giant within*. Fireside.
Rowling, J. K. (2020). J.K. Rowling writes about her reasons for speaking out on sex and gender issues Accessed September 3, 2020 from https://www.jkrowling.com/opinions/j-k-rowling-writesabout-her-reasons-for-speaking-out-on-sex-and-gender-issues/.
Rucker, D. D., Preacher, K. J., Tormala, Z. L., & Petty, R. E. (2011). Mediation analysis in social psychology: Current practices and new recommendations. *Social and Personality Psychology Compass, 5*(6), 359–371.
Shavitt, S. (1990). The role of attitude objects in attitude functions. *Journal of Experimental Social Psychology, 26*(2), 124–148.
Sherman, S. J., & Gorkin, L. (1980). Attitude bolstering when behavior is inconsistent with central attitudes. *Journal of Experimental Social Psychology, 16*(4), 388–403.
Simon, L., Greenberg, J., & Brehm, J. (1995). Trivialization: The forgotten mode of dissonance reduction. *Journal of Personality and Social Psychology, 68*(2), 247–260.
Smith, M. B., Bruner, J. S., & White, R. W. (1956). *Opinions and personality*. Wiley.

Snyder, M., & DeBono, K. G. (1985). Appeals to image and claims about quality: Understanding the psychology of advertising. *Journal of Personality and Social Psychology, 49*(3), 586–597.
Sommers, S. R., & Kassin, S. M. (2001). On the many impacts of inadmissible testimony: Selective compliance, need for cognition, and the overcorrection bias. *Personality and Social Psychology Bulletin, 27*(10), 1368–1377.
Tajfel, H. (1974). Social identity and intergroup behaviour. *Social Science Information, 13*(2), 65–93.
Tannenbaum, M. B., Hepler, J., Zimmerman, R. S., Saul, L., Jacobs, S., Wilson, K., & Albarracín, D. (2015). Appealing to fear: A meta-analysis of fear appeal effectiveness and theories. *Psychological Bulletin, 141*(6), 1178–1204.
Tesser, A. (1978). Self-generated attitude change. In , *Vol. 11. Advances in experimental social psychology* (pp. 289–338). Academic Press.
Tesser, A., & Conlee, M. C. (1975). Some effects of time and thought on attitude polarization. *Journal of Personality and Social Psychology, 31*(2), 262.
Tesser, A., & Leone, C. (1977). Cognitive schemas and thought as determinants of attitude change. *Journal of Experimental Social Psychology, 13*(4), 340–356.
Thurstone, L. L., & Chave, E. J. (1929). *The measurement of attitude*. University of Illinois Press.
Vonasch, A. J., Reynolds, T., Winegard, B. M., & Baumeister, R. F. (2018). Death before dishonor: Incurring costs to protect moral reputation. *Social Psychological and Personality Science, 9*(5), 604–613.
Wenzlaff, R. M., & Wegner, D. M. (2000). Thought suppression. *Annual Review of Psychology, 51*(1), 59–91.
Wheeler, S. C., & DeMarree, K. G. (2019). Prevalence, antecedents and consequences of actual-desired attitude discrepancies. In A. Reed, II, & M. Forehand (Eds.), *Handbook of research on identity theory in marketing* (pp. 346–359). Edward Elgar Publishing.