Running Head: COUNTERFACTUALS AND POLITICAL FALSEHOODS

It Could Have Been True:
How Counterfactual Thoughts Reduce Condemnation of Falsehoods and Increase Political Polarization

Daniel A. Effron

London Business School

Article in press at *Personality and Social Psychology Bulletin*

Word count: 9,961

Abstract

This research demonstrates how counterfactual thoughts can lead people to excuse others for telling falsehoods. When a falsehood aligned with participants' political preferences, reflecting on how it *could have been true* led them to judge it as less unethical to tell, which in turn led them to judge a politician who told it as having a more moral character and deserving less punishment. When a falsehood did not align with political preferences, this effect was significantly smaller and less reliable, in part because people doubted the plausibility of the relevant counterfactual thoughts. These results emerged independently in three studies (two pre-registered; total $N = 2{,}783$) and in meta- and Bayesian analyses, regardless of whether participants considered the same counterfactuals or generated their own. The results reveal how counterfactual thoughts can amplify partisan differences in judgments of alleged dishonesty. I discuss implications for theories of counterfactual thinking and motivated moral reasoning.

Keywords: counterfactual thinking, ethics/morality, social judgment, dishonesty, lies, political psychology, Trump, Clinton

It Could Have Been True:
How Counterfactual Thoughts Reduce Condemnation of Falsehoods and Increase Political Polarization

No one has ever doubted that truth and politics are on rather bad terms with each other.

- Hannah Arendt (1961)

The political events of 2016 did little to challenge Arendt's observation. Amid swirls of "fake news" and "alternative facts," British politicians spread misinformation during the Brexit referendum (The Economist, 2016), and the American Presidential candidates made dozens of false claims (Politifact, n.d.). Once the public recognizes a falsehood, how do they judge the ethicality of telling it? These judgments have the potential to mobilize voters and shape elections by sparking moral outrage (cf. Skitka, Bauman, & Sargis, 2005; Pagano & Huo, 2007), but not all falsehoods receive the same harsh judgments. I argue that imagining alternatives to reality powerfully shapes such judgments. For example, to defend the false claim that more people attended Donald Trump's inauguration than Barack Obama's, spokespeople proposed Trump's crowd would have been larger if the weather had been better (Rossoll, 2017). Or some people might justify Hillary Clinton's false claim that no Trump-brand products are made in the U.S. (Dent, 2016) by imagining they would have been made abroad if it had been cheaper to do so. These conditional propositions – *if* circumstances had been different, *then* an event would have occurred – are called *counterfactuals* (see Byrne, 2016). Logically, they do not render falsehoods true, but psychologically, they may make falsehoods seem less unethical. The present research tests whether making such counterfactuals salient can reduce the moral condemnation falsehoods receive. Minimizing the fallout from telling a falsehood may not require convincing people the falsehood is literally true; convincing them it *could have been* true may be sufficient.

Why would salient counterfactuals reduce moral condemnation of falsehoods? Even falsehoods that people explicitly recognize as false can still *feel* more or less truthful (Shidlovski, Schul, & Mayo, 2014). Moral judgments of such falsehoods depend not only on their literal truthfulness (Levine & Schweitzer, 2014; Rogers, Zeckhauser, Gino, Norton, & Schweitzer, 2017) but also on how close to the truth they feel. Falsehoods that feel close to reality may seem more justified and less dishonest (Schweitzer & Hsee, 2002). For example, participants were more likely to lie about achieving a prize-winning die-roll when their actual roll was numerically close to (vs. far from) the winning number (Hilbig & Hessler, 2013). Despite the lie being just as false in all cases, participants may have felt it was less unethical to tell when it involved misreporting their roll by a small margin than a large margin (see also Briazu, Walsh, Deeprose, & Ganis, 2017 Study 4).

Mental simulation is likely to play an important role in how close a falsehood feels to reality. In general, mentally simulating events bring them closer to reality (Koehler, 1991). People judge future events as more likely to occur when they have been instructed to imagine them (Anderson, 1983; Carroll, 1978; Gregory, Cialdini, & Carpenter, 1982) – and the easier it is to imagine them, the more likely they seem (Sherman, Cialdini, Schwartzman, & Reynolds, 1985). Similarly, when an event did not occur, people judge whether it "almost" occurred by mentally simulating counterfactual scenarios in which it *could have* occurred (Kahneman & Tversky, 1982). The more easily such scenarios come to mind, the closer to reality the event will feel (Miller, Turnbull, & McFarland, 1989; Kahneman & Miller, 1986). For example, a traveler will feel she "nearly" made her flight when she missed it by 5 minutes versus 30 minutes, because it is easier to imagine scenarios in which she arrived 5 minutes earlier (Kahneman & Tversky, 1982). In this way, salient counterfactuals can bring an event closer to reality (Medvec, Madey, & Gilovich, 1995; Markman & Tetlock, 2000). Extending this reasoning, falsehoods should feel closer to the truth, and thus seem less

unethical to tell, when evaluated in the presence of salient counterfactuals about how they could have been true.

Not everyone will be equally swayed by salient counterfactuals, however, particularly in political contexts. Reflecting on how a falsehood could have been true should affect moral judgments more powerfully when the falsehood aligns with one's political views, for two reasons. First, for a salient counterfactual to affect judgments, it must be subjectively plausible (Petrocelli, Percy, Sherman, & Tormala, 2011). For example, suggesting "if the weather had been nicer, then Trump's inauguration would have been larger" should have little effect on someone who believes (a) the weather had little chance of being nicer, or (b) even if the weather had been nicer, turnout would have remained the same. Second, people will find a counterfactual more plausible when it aligns with beliefs and ideologies they are motivated to defend (Tetlock, 1998; Tetlock & Henik, 2005; cf. Markman & Hirt, 2002), perhaps because people are unwilling or unable to imagine a counterfactual occurring if it challenges their views (Tetlock, Kristel, Elson, Green, & Lerner, 2000). For example, the Trump-inauguration counterfactual should seem more plausible to Trump supporters, who should be inclined to believe in Trump's popularity, than to Clinton supporters.

In this way, counterfactuals may help people reach motivated moral conclusions. Partisans should want to excuse falsehoods aligned with their political views (cf. Abrams, Randsley de Moura, & Travaglino, 2013; Mueller & Skitka, in press), but be reluctant to do so without justification (Kunda, 1990). Salient counterfactuals provide justification to the extent they are plausible, and the same counterfactual will seem more plausible when it aligns with one's political preferences (Tetlock, 1998). Making a counterfactual salient should therefore have a larger effect on moral judgments of falsehoods that are aligned (vs. misaligned) with one's political preferences, amplifying partisan differences in moral judgments of falsehoods. For example, Trump supporters should already express less

condemnation of the inauguration falsehood than his opponents do; imagining how his inauguration *could have been* bigger should increase this difference by mitigating supporters' (more than opponents') condemnation.

Based on this logic, I tested three main hypotheses. First, reflecting on how a falsehood could have been true should lead people to judge it as less unethical to tell. Second, this effect should be larger when the falsehood aligns with their political preferences, resulting in larger partisan differences in judgments (i.e., increasing political polarization). Third, this increase in political polarization should be mediated by perceptions of the counterfactual's plausibility. Evidence for the hypotheses would suggest leaders can reduce the negative consequences of telling a falsehood merely by convincing their supporters it could have been true.

These hypotheses advance work suggesting counterfactuals can facilitate dishonest behavior. In one study, participants who spontaneously generated more counterfactuals about how a car accident could have been avoided also generated more ways of lying to the police about how the accident occurred (Briazu et al., 2017). These results suggest lying and counterfactual thinking may share an underlying process. In another study, people were more likely to falsely claim they achieved an outcome when they had seen it occur than when they had not (Shalvi, Dana, Handgraaf, & De Dreu, 2011). Participants were asked to report a private die-roll honestly, but could earn more money for reporting higher numbers. They (dishonestly) reported higher numbers when they first completed two unpaid rolls than when they did not. Rather than maximizing their payout by reporting the highest roll possible, they reported the highest outcome they had observed on the unpaid rolls. Apparently, observing a "desired counterfactual" made lying feel more justified (Shalvi et al., 2011). However, because participants physically saw the counterfactual outcome, it is unclear whether mental simulation produced this effect. Going beyond this prior work, the present research is the first

to examine how mentally simulating counterfactuals can mitigate moral condemnation of others' falsehoods, help people reach motivated moral conclusions, and magnify political polarization.

The Present Research

Three experiments tested these ideas in the context of the 2016 U.S. Presidential election. Trump and Clinton supporters judged falsehoods that were aligned or misaligned with their political preferences. In Study 1, half the participants were randomly assigned to read a counterfactual proposition about how the falsehoods could have been true. I expected that this salient counterfactual would reduce how unethical they judged the falsehoods as being, and that this effect would be more pronounced when the falsehoods aligned with their political preferences. Study 2 sought to replicate these effects and rule out the possibility that the act of mental simulation in general (rather than the act of mentally simulating the counterfactual in particular) explained the results. Study 2 also examined whether salient counterfactuals would make politician who told the falsehoods seem more moral and less deserving of punishment. Study 3 sought to generalize these effects and address an alternative explanation by manipulating whether participants generated their own counterfactuals. All studies included measures to ensure people knew the falsehoods were false. I determined the targeted sample sizes in advance, and I report all conditions, dependent measures, and data exclusions. Studies 2 and 3 were pre-registered.

Study 1

Method

Design. Participants evaluated six falsehoods in a 2 (condition: counterfactual vs. control; between-subjects) X 2 (falsehood: aligned vs. misaligned with participants' political preferences; within-subjects) factorial design.

Participants. In March, 2017, I requested 1,000 U.S.-based participants from Amazon Mechanical Turk (MTurk) who supported either Donald Trump or Hillary Clinton, rivals in the 2016 Presidential election. I selected this large sample size because a pilot study suggested a small main effect of the counterfactual manipulation, and because I wanted sufficient statistical power to examine a potential moderator. After applying *a priori* exclusion criteria to promote data quality (e.g., discarding data from duplicates IP address), 1,030 participants were reduced to 1,019. I classified people as supporting a candidate if they had voted or had intended to vote for him or her. Programmed quotas ensured a similar number of participants supported each candidate (507 Trump supporters and 512 Clinton supporters). (See Online Supplement for details about eligibility criteria, exclusions, and sample characteristics).

Procedure. Participants viewed a fact related to the 2016 Presidential candidates, randomly selected from a bank of six (e.g., "It's a proven fact that Donald Trump won the electoral vote, but lost the popular vote to Hillary Clinton"; see Table 1). Participants randomly assigned to the *counterfactual condition* read a conditional proposition about how the opposite of the fact could have been true (e.g., "If only Trump had tried to win the popular vote, then he would have won the popular vote"), and rated the proposition's plausibility (see below). Participants in the *control condition* neither read nor rated the proposition. For the dependent measure (described below), all participants rated how unethical it would be to tell a falsehood that contradicted the fact (e.g., "Trump won the popular vote;" see Table 1).

Staying in the same condition, participants then repeated this procedure for the remaining facts (and corresponding falsehoods) from the bank, in randomized order (see Table 1). Of the six falsehoods, three aligned with Trump supporters' political views (i.e., two painted Trump positively and one painted Clinton negatively), and three aligned with

Clinton supporters' political views (i.e., two painted Trump negatively and one painted Clinton positively). The falsehoods were taken from or inspired by claims actually made by or about the candidates. Finally, I tested participants' memory for the facts' truthfulness (see below).

Measures.

Dependent measure: Unethicality of telling falsehoods. Six items assessed the unethicality of telling each falsehood (αs > .84 for each falsehood): how dishonest, justified*, unethical, acceptable*, and problematic it would be to make the statement, and how much of a lie it was (starred items were reverse-coded). To respond, participants moved a slider on 100-point scales anchored at *Not at all* and *Extremely*, with the slider initially appearing at 50 (αs > .84 for each falsehood).

Counterfactual plausibility/potency. Would participants find a counterfactual more plausible when it aligned with their political preferences? When a counterfactual is expressed as a conditional ("if only X had occurred, then Y would have occurred"), a person could discount the plausibility of either the *if* component (i.e., "X could not have occurred") or the *then* component (i.e., "even if X had occurred, Y would not have occurred"; Tetlock & Henik, 2005). Participants in the counterfactual condition rated the subjective plausibility of both components, for each counterfactual they considered, using items by Petrocelli and colleagues (2011). For example, after considering the counterfactual "*If investigators had been able to see [Clinton's] deleted emails, then the FBI would have brought charges against Hillary Clinton*," participants were asked, "What do you perceive was the likelihood of investigators actually being able to see the deleted emails?" and "Suppose that investigators had actually seen the deleted emails. Given that, what do you perceive was the likelihood that the FBI would have brought charges agains Clinton?" (1 = *Not at all likely;* 11 = *Extremely likely*). A counterfactual's overall plausibility – termed *counterfactual potency* – is

operationalized as the multiplicative effect of the *if* and *then* likelihood ratings (Petrocelli et al., 2011; Petrocelli, Kammrath, Brinton, Uy, & Cowens, 2015).

Memory. Would salient counterfactuals shape how people judge a falsehood simply by making them forget it is false (Gerlach, Dornblaser, & Schacter, 2014; Petrocelli & Crysel, 2009)? To rule out this possibility, I asked participants to categorize six statements as true/false (see Online Supplement). Three statements referred to the facts and three referred to the falsehoods participants had seen earlier. I predicted the manipulation would not affect this memory measure.

Additional measures. All studies also included exploratory measures of political knowledge and approval of Donald Trump, discussed in the Online Supplement.

Results

All studies report results collapsed across the six falsehoods; the Online Supplement reports results for each individual falsehood. The analyses in all studies produced identical conclusions when they included fixed effects for individual falsehoods.

Unethicality of telling falsehoods. Overall, people thought it would be unethical to tell the falsehoods, shown by mean ratings exceeding the scale midpoint. As predicted, however, they thought it would be less unethical when they had considered counterfactuals (M = 77.49, SD = 16.23) than when they had not (M = 82.17, SD = 17.65), d = .28, b = –4.70, z = 4.46, $p < .001$ in a mixed-effect regression analysis, with a dummy code for condition (1 = counterfactual, 0 = control), and a random effect for participant (see Table 2, step 1). This effect was significantly larger when participants judged falsehoods that were aligned (vs. misaligned) with their political preferences (see Figure 1). That is, when I added a dummy code for alignment (1 = aligned, 0 = misaligned) and its interaction with condition to the regression, the interaction was negative and significant, b = -1.93, z = 2.14, p = .032 (see

Table 2, step 2). At this point in the research process, I considered this interaction test exploratory; Studies 2 and 3 provide confirmatory tests.

To better understand the interaction, I computed simple slopes. The manipulation had a significant effect regardless of alignment, but as noted this effect was bigger for statements that were aligned with political preferences ($M_{counterfactual}$ = 72.83, $M_{control}$ = 78.46; *SD*s = 19.77 and 20.80, respectively), d = .28, z = 4.92, p < .001, versus misaligned ($M_{counterfactual}$ = 82.33, $M_{control}$ = 85.98, *SD*s = 16.51 and 17.70, respectively), d = .21, z = 3.24, p = .001. Decomposing the interaction the other way shows that participants thought it was less unethical to make a false statement that aligned (vs. did not align) with their views in the control condition, d = .39, z = 11.96, p < .001, and the counterfactual condition amplified this effect, d = .52, z = 14.89, p < .001. In other words, considering counterfactuals increased political polarization in moral judgments.

Counterfactual potency. Why did considering counterfactuals increase political polarization? I speculated it was because people found a counterfactual more potent (i.e., plausible) if it aligned with their political views, and the more potent they found the counterfactual, the less unethical they would judge the falsehood. Consistent with this idea, the data showed a significant negative indirect effect of alignment on unethicality judgments via potency ratings in the counterfactual condition (see Figure 2), b = -3.00 [-3.68, -2.29]. This analysis omits the control condition because its participants neither saw nor rated counterfactuals. As noted, potency was computed as the product of likelihood ratings for the *if* and *then* part of each counterfactual (Petrocelli et al., 2011). The 95% CI around the indirect effect was calculated with 5,000 resamples and bias-corrected with Stat's *ml_mediation* function, which accounted for the multi-level nature of the data. (The path from potency to unethicality remained significant when controlling for the simple effects of both *if* and *then* likelihoods).

Memory. Did considering counterfactuals make people forget the falsities were false? I expected not, and to find out, I coded whether responses to each memory item were correct (1) or incorrect (0). People correctly distinguished fact from fiction a high percentage of the time, and this percentage was virtually identical in both conditions (control: 91.81%; counterfactual: 92.26%), $b = .08$, $z = .63$, $p = .53$ in a mixed-effect logistic regression with participant as a random effect. This result did not depend on whether the falsehood aligned with political preferences, $b = -.25$, $z = 1.16$, $p = .245$ for the interaction when added with the simple effect of alignment to the regression. Thus, considering counterfactuals reduced condemnation of falsehoods even though participants acknowledged their falsity.

Discussion

Supporting the hypotheses, salient counterfactuals made falsehoods seem less unethical to tell. This effect was larger when the falsehoods aligned with one's political preferences, and the results were consistent with the idea that this was because alignment increased the counterfactuals' perceived plausibility. There was no support for the possibility that considering counterfactuals simply made people forget the falsehoods were false.

Study 2

Study 1 manipulated whether people imagined counterfactuals or did nothing. To address whether the results were due to the mere act of imagination in general, rather than imagining counterfactuals in particular, Study 2 manipulated whether people imagined how a falsehood might have been true (counterfactual) or how an event might occur (control). Although both conditions require imagination, I expected the falsehood to seem less unethical in the counterfactual condition.

Study 2 also examined potential downstream consequences of salient counterfactuals. Participants indicated how they would react if a politician they supported told the relevant falsehood. I predicted that, by diminishing the falsehoods' subjective unethicality, the

counterfactual manipulation would lead people to judge the candidate as having better moral character and deserving less punishment.

Finally, Study 2 included a more sensitive memory measure to ensure the results were not due to forgetting the falsehoods were false.

Method

Design. I preregistered the hypotheses, target sample size, methods, and analytic strategy (see http://aspredicted.org/blind.php?x=4qd5jw). As in Study 1, participants evaluated six falsehoods in a 2 (condition: counterfactual vs. control; between-subjects) X 2 (falsehood: aligned vs. misaligned with participants' political preferences; within-subjects) factorial design.

Participants. I posted slots for 800 MTurk participants in April, 2017. I targeted this sample size because a power simulation showed it provided 85% power to detect Study 1's main effect at $\alpha = .05$, one-tailed. Of the 906 people who began the study, 884 remained after applying pre-registered exclusion criteria. With no quotas programmed, the sample contained more Clinton supporters (468) than Trump supporters (224), and 192 people who supported neither candidate. (See Online Supplement for more details about the power simulation, eligibility criteria, exclusions, and sample characteristics).

Procedure. The counterfactual condition was identical to Study 1, with slight changes to the counterfactuals' wording (see Table 1): Participants read one of the six political facts (e.g., *It's a proven fact that Trump-brand hats, wine, and water are made in the USA*), considered an if-then statement about how its opposite could have been true (*If Trump had been able to make those products more cheaply in a different country, then he would have made them outside of the USA*), and rated the if-then statement's potency.

In a new control condition, participants read the same fact, considered an if-then statement about how a related event could occur in the future (*If an American car*

manufacturer moves a factory outside the USA, then the quality of the cars will decrease; see Table 1), and rated the if-then statement's potency (Petrocelli, Seta, & Seta, 2012). Thus, both conditions involved rating an if-then statement and mentally simulating an event that had not occurred, but only the counterfactual condition required imagining how the falsehood could have been true. Only in the counterfactual condition did the potency ratings represent *counterfactual* potency, so I did not analyze them in the control condition. After the manipulation, participants completed the dependent measures (see below), repeated the procedure for the remaining five political facts (see Table 1), and then completed the memory measure (see below).

Measures.

Main dependent measure: Unethicality of telling falsehoods. Participants rated the unethicality of telling each falsehood using the six-item scale from Study 1 (αs > .82 for each falsehood).

Moral character and punitive sentiment. Two new measures assessed potential consequences of perceiving falsehoods as unethical to tell. For each false statement, participants imagined that a Congressional candidate from their district, for whom they are considering voting, publicly insists that the statement is true. Then they rated the candidate's *moral character* on three items (honest, trustworthy, and principled; αs > .92 for each falsehood; 0 = *Not at all*, 100 = *Extremely*)(Goodwin, Piazza, & Rozin, 2014), and indicated their agreement with four *punitive sentiments*: "The candidate should drop out of the race for making that statement," "The candidate deserves to be publicly criticized for making that statement," "Because of that statement, I would think twice before voting for this candidate," and "That statement would make me withdraw my support from the candidate" (-50 = *Strongly disagree,* 50 = *Strongly agree*; αs > .87). I predicted the manipulation would increase moral character ratings and decrease punitive sentiment by making the falsehoods

seem less unethical to tell (an indirect effect). I also predicted a total effect of the manipulation on these measures, particularly when the falsehoods were aligned (vs. misaligned) with participants' political views.

Memory and confidence. Participants answered the six true-false questions from Study 1. I expected the manipulation not to affect this measure. To assess memory more precisely, I also asked participants to indicate their confidence in each true/false answer on a 0%-to-100% scale (Koch & Forgas, 2012). I expected all predicted effects to remain reliable when controlling for confidence.

Results

As I had strong directional predictions, I pre-registered and report one-tailed significance tests for all confirmatory analyses. As indicated below, I report two-tailed tests for exploratory analyses.

Unethicality of telling falsehoods. Replicating Study 1 with a more stringent control condition, people rated the falsehoods as significantly less unethical when they imagined how the falsehoods could have been true ($M_{\text{counterfactual}}$ = 76.09, SD = 16.07) than when they imagined how an event could occur in the future (M_{control} = 78.07, SD = 15.08), d = .13, b = –2.14, z = 2.08, p = .019 in a mixed regression model with participant as a random effect (see Table 3, Step 1). Also replicating Study 1, and as expected, this effect was significantly larger for statements that were aligned versus misaligned with participants' political views (see Figure 3), b = -2.80, z = 2.35, p = .0095 for the condition X alignment interaction (see Table 3, Step 2). (Because alignment could not be coded for participants who supported neither Trump nor Clinton, I had to omit these participants from the interaction test, leaving 692).

To better understand the interaction, I computed simple slopes. For falsehoods that aligned with political preferences, considering counterfactuals significantly reduced unethicality ratings, as in Study 1 and as expected ($M_{\text{counterfactual}}$ = 71.70, M_{control} = 75.53; SDs

= 19.47 and 17.26, respectively), d = .21, b = -3.64, z = 2.80, p = .003. For falsehoods that were misaligned with political preferences, considering counterfactuals had no significant effect ($M_{counterfactual}$ = 80.69, $M_{control}$ = 81.68; SDs = 17.32 and 15.79, respectively), d = .06, b = -.84, z = .64, p = .260 (see Figure 3). Based on Study 1's results, I had expected the effect for misaligned falsehoods to be significant (which it was not), though smaller than the effect for aligned falsehoods (which it was).

Decomposing the interaction the other way with exploratory (two-tailed) tests revealed that people rated the falsehoods that were aligned with their political preferences as less unethical than those that were misaligned in both conditions, but that this difference was larger in the counterfactual condition, d = .49, b = -9.05, z = 10.39, p < .001, than in the control condition, d = .37, b = -6.25, z = 7.43, p < .001. Thus, as in Study 1, the counterfactuals exacerbated existing political polarization in moral judgments.

Counterfactual potency. Why did the counterfactual-thinking manipulation increase political polarization? As in Study 1, I tested the role of counterfactual potency. A mediation analysis replicated Study 1's results: a significant, negative indirect effect from alignment to potency ratings to dishonesty judgments in the counterfactual condition, b = -3.87 [-4.98, -2.91] (see Figure 4). This analysis omits the control condition because it did not measure counterfactual potency. These results are consistent with the idea that counterfactuals increased political polarization because they seemed more plausible – i.e., were more potent – when they supported desired political conclusions. (I pre-registered the prediction that greater potency would predict lower dishonesty ratings, but did not consider the full mediation model until after data were collected).

Moral character and punitive sentiment. I next tested the prediction that because counterfactual thinking reduced the perceived unethicality of telling a falsehood, it would lead people to rate a candidate who told the falsehood as higher in moral character and less

deserving of punishment. Consistent with this prediction, there was a significant positive indirect effect from condition (counterfactual vs. control), to the unethicality of telling the falsehood, to the moral character of the candidate, b = 1.29 [.16, 2.65], and a negative indirect effect to punitive sentiment, b = -1.29 [-2.59, -.19], computed as above using Stata's *ml_mediation* command.

Additional analyses (see Online Supplement) did not support the prediction that considering counterfactuals would raise moral character ratings or lower punitive sentiment overall. Thus, there was a reliable and theoretically meaningful indirect effect on these measures without a reliable total effect (see Rucker, Preacher, Tormala, & Petty, 2011). As predicted, the manipulation also had a larger effect on moral character judgments and punitive sentiment when falsehoods aligned (vs. misaligned) with participants' political preferences

Memory. I next tested whether memory distortion could explain the effects of considering counterfactuals. Once again, people correctly distinguished fact from fiction a high percentage of the time, and as predicted, it was statistically indistinguishable across conditions (control: 91.45%; counterfactual: 91.24%), b = .04, z = .21, p = .832 (two-tailed) in a mixed logistic regression. An exploratory test showed no significant moderation by whether the falsehood aligned with political preferences, b = .04, z = .16, p = .873 (two-tailed).

Further exploration showed that people were marginally less confident in their judgments about the statements' truth in the counterfactual condition than in the prefactual condition (Ms = 90.16 and 91.66, SDs = 12.11 and 10.11, respectively), b = -1.50, z = 1.95, p = .051 (two-tailed), providing some evidence that the manipulation affected memory. However, this effect did not depend on whether the falsehood aligned with political preferences, b = .04, z = .16, p = .873 (two-tailed) – and crucially, as predicted, the direction

and significant of the results for all dependent measures (statement dishonesty, candidate morality, and negative consequences) were identical after statistically controlling for confidence ratings. Thus, counterfactual thinking may reduce people's confidence that falsehoods are false (cf. Petrocelli & Crysel, 2009), but this effect was insufficient to explain why considering counterfactuals reduced perceptions of falsehoods' dishonesty.

Discussion

Study 2 replicated Study 1 and suggested that the general act of imagination did not produce the results. Imagining *counterfactuals* in particular – compared to imagining the future – made falsehoods seem less unethical to tell. Again, the effect was larger when the falsehoods aligned with participants' political preferences, and the data were consistent with the idea that this was because alignment increased the counterfactuals' perceived plausibility. A salient counterfactual also helped insulate a favored politician from the negative consequences of telling falsehoods. Finally, using a more sensitive measure than Study 1, Study 2 found no evidence that memory distortion could explain the results.

Study 3

A potential alternative explanation is that factual information embedded in the counterfactuals, rather than the counterfactuals themselves, mitigated condemnation of falsehoods. For example, asking whether *Trump's inauguration would have been bigger if security had not been so strict* reveals the fact security was strict. To address this possibility, Study 3 tested whether the effects would replicate if participants generated their own counterfactuals. This manipulation ensures all participants receive the same factual information, and tests generalizability beyond the specific counterfactuals used in the previous studies.

Method

Design. I preregistered the hypotheses, sample size, methods, and analytic strategy (see http://aspredicted.org/blind.php?x=ym5ba9). As in Studies 1 and 2, participants evaluated six falsehoods in a 2 (condition: counterfactual vs. control; between-subjects) X 2 (false statement: aligned vs. misaligned with participants' political preferences; within-subjects) factorial.

Participants. Study 3 ran in May, 2017, on MTurk. I posted 800 slots to target Study 2's sample size. Of the 891 people who began the study, 871 remained after applying the pre-registered exclusion criteria used in Study 2. The final sample contained 457 Clinton supporters, 244 Trump supporters, and 170 who supported neither candidate. (See Online Supplement for additional details).

Procedure. The procedure was identical to Study 2, except participants randomly assigned to the counterfactual condition were asked to generate their own counterfactual thoughts about each of the six facts in Table 1, whereas those in the control condition were not. For example, participants in both conditions read, "It's a proven fact that fewer people attended Donald Trump's inauguration than Barack Obama's." Only people in the counterfactual condition then read:

> However, it's possible to imagine that more people would have attended Trump's inauguration than Obama's if circumstances had been different. Please complete the statement below by filling in the blank.
>
> *More people would have attended Trump's inauguration than Obama's, if* _____.

There was space to complete the sentence in up to three different ways.

Measures. As in Study 2, participants rated the unethicality of telling the six falsehoods shown in Table 1 (αs > .81), the moral character of a politician who asserted the falsehoods (αs > .91), and their punitive sentiment toward the politician (αs > .86). At the end of the study, they completed Study 2's memory check items and confidence ratings. Study 3 did not administer the counterfactual-potency measure due to concern that the wide

range of counterfactuals generated would lend too much variance to this measure for meaningful interpretation.

Results

As I had strong directional predictions, I pre-registered and report one-tailed significance tests for all confirmatory analyses. Where indicated, I report two-tailed tests for exploratory analyses. Showing that participants complied with the manipulation, 94% of responses in the counterfactual condition provided at least one counterfactual thought (M = 1.78 counterfactuals listed, SD = .96).

Unethicality of telling falsehoods. Replicating Studies 1 and 2's results with a new manipulation, Study 3 participants rated the falsehoods as significantly less dishonest when prompted to generate counterfactuals (M = 78.43, SD = 14.90) than when not prompted (M = 80.50, SD = 16.94), b = –2.12, z = 2.00, p = .023, d = .13 (see Table 4, Step 1) in a mixed regression model with condition (1 = counterfactual, 0 = control) as a fixed effect and participant as a random effect. Also replicating Studies 1 and 2, the manipulation effect was significantly larger for statements that were aligned versus misaligned with participants' political views (see Figure 5), b = -4.37, z = 3.98, p < .001 for the condition X alignment interaction (see Table 4, Step 2). (Alignment could not be coded for participants who supported neither Clinton nor Trump, so I had to omit their data from the interaction test, leaving 701 people).

Next, I decomposed the interaction by computing simple slopes. For falsehoods that aligned with political preferences, the new counterfactual manipulation significantly decreased unethicality ratings, as predicted ($M_{control}$ = 77.81, $M_{counterfactual}$ = 74.04; SDs = 19.26 and 18.64, respectively), b = -3.81, z = 2.92, p =.002, d = .20. Based on Study 1's results, I had predicted that the smaller manipulation effect for misaligned statements would still be significant, but as in Study 2 it was not ($M_{control}$ = 83.01, $M_{counterfactual}$ = 83.46; SDs =

14.96 and 17.78, respectively), $b = .56$, $z = .43$, $p = .334$, $d = -.03$. Decomposing the interaction the other way with (two-tailed) exploratory analyses shows that participants thought it was less unethical to make a false statement when it aligned (vs. did not align) with their views in the control condition, $d = -.28$, $b = -5.31$, $z = 6.97$, $p < .001$, and the counterfactual condition amplified this effect, $d = -.56$, $b = -9.68$, $z = 12.26$, $p < .001$. In this way, generating counterfactuals exacerbated existing political polarization.

Moral character and punitive sentiment. I next tested the prediction that because counterfactual thinking reduced the perceived unethicality of telling a falsehood, it would lead people to rate a candidate who told the falsehood as higher in moral character and less deserving of punishment. Like Study 2, Study 3 supported this prediction with a significant indirect effect from condition, to the unethicality of telling the falsehood, to the moral character of the candidate, $b = 1.14$ [.053, 2.33], and (in a separate analysis) from condition to falsehood unethicality to punitive sentiment, $b = -1.17$ [-2.40, -.054], computed as in Study 2 using Stata's *ml_mediation* command.

Additional analyses (see Online Supplement) confirmed the prediction that generating counterfactuals would raise moral character ratings overall, but not the prediction that it would lower punitive sentiment overall. These analyses also showed, as predicted, a significantly larger manipulation effect on these variables when falsehoods were aligned (vs. misaligned) with participants' political preferences.

Memory. Did the counterfactual manipulation affect people's memory for the statements' truth value? Some evidence suggests so. Unlike Studies 1 and 2, and contrary to predictions, participants at the end of the study were significantly less likely to correctly identify whether the statements were true (90.37%) than in the control condition (92.56%), $b = -.58$, $z = 2.41$, $p = .016$ (two-tailed). Also, people were marginally less confident about their

memories in the counterfactual condition (M = 91.97, SD = 11.60) than in the control condition (M = 93.37, SD = 9.68), b = -1.40, z = 1.89, p = .059 (two-tailed), d = .13.

Importantly, though, these memory effects are insufficient to explain why generating counterfactuals affected unethicality judgments. First, neither memory effect depended on whether the falsehoods aligned with political preferences, b = -.04, z = .14, p = .892 and b = -.04, z = .06, p = .954, respectively. Second, all results above remained significant when I reran the analyses including only observations where people correctly identified the statement's falsity. Third, regardless of whether only these or all observations were retained, the results also remained significant when controlling for confidence judgments, as predicted. Thus, even among people who accurately remembered the statements as false, and even holding constant people's confidence in their falsity across conditions, generating counterfactuals still reduced how dishonest people thought it was to make those statements.

Discussion

Prompting people to generate their own counterfactuals about how a falsehood *could have been true* reduced their condemnation of it, as long as the falsehood aligned with their political preferences. Study 3 thus attests to the effects' generalizability and robustness by demonstrating they do not depend on providing people with specific counterfactual thoughts.

Meta-Analysis and Bayesian Analysis

To better estimate the effect sizes, I meta-analyzed Studies 1-3. Given the similarity of paradigms and participant populations, I adopted a fixed-effects approach, but a random-effects approach produced the same conclusions, except where indicated.

Across studies, considering counterfactuals decreased unethicality judgments (see Figure 6). Across all falsehoods, this effect was small but significant, as indicated by a 95% CI that excluded 0, d = .18 [0.11, 0.26]. The effect was significant regardless of whether falsehoods aligned with political preferences, but was larger for those that were aligned than

for those that were misaligned, *d*s = .23 [.15, .03] and .010 [.02, .18], respectively. Also, the effect for misaligned falsehoods was not very robust; it only emerged in Study 1, and was not significant in the meta-analysis when specifying random effects, 95% CI = [-.06, .23].

Figure 7 shows how considering counterfactuals increased political polarization in unethicality judgments across studies. Here, the effect sizes represent the difference between judgments of falsehoods that were aligned versus misaligned with political preferences. Larger effect sizes thus represent greater political polarization in judgments of the same falsehoods. The results showed significant and modestly-sized political polarization regardless of condition, but the effect size was larger in the counterfactual conditions than in the control conditions, *d*s = .52 [.44, .60] and .35 [.027, .043].

To quantify the evidence in favor of the hypothesis, I performed Bayesian *t*-tests (two-tailed) on data collapsed across all studies, using the JASP statistical software and its default Cauchy prior width of $r = .707$ (Wagenmakers et al., 2016). The conclusions complemented the meta-analysis. Across all falsehoods, the Bayes factor (BF_{10}) was 5,536, indicating the data were over 5,000 times more likely to have been observed under the hypothesis that the manipulation affected ethicality judgments than under the null hypothesis that it did not. Falsehoods that aligned with political preferences showed even stronger evidence favoring the hypothesis; the data were over 600,000 times more likely under the hypothesis than under the null ($BF_{10} = 613{,}655$). By contrast, falsehoods that were misaligned with political preferences did not show Bayesian evidence of the effect; the likelihood of observing the data was about equal under the hypothesis and the null ($BF_{10} = .93$). One-tailed tests produced similar conclusions.

Together, these results provide robust meta-analytic and Bayesian evidence that salient counterfactuals can decrease unethicality judgments of falsehoods. The effect sizes were small but robust when falsehoods aligned with political preferences (e.g., average *d* =

.23), but there was not consistent evidence for an effect when falsehoods were misaligned with such preferences. Theoretically, it is important that brief counterfactual reflection can shift moral judgments of falsehoods at all (see Prentice & Miller, 1992). Practically, small effects on judgments of political falsehoods are important when elections, like the 2016 US Presidential race, are won by slim margins.

General Discussion

"Truth. It's grounded in facts," reminds a New York Times advertisement. The present research suggests that judgments of falsehoods are grounded in counterfactuals. Reflecting on how a falsehood could have been true led people to judge it as less unethical to tell, despite acknowledging its falsity. This effect emerged in three studies (two pre-registered; total $N = 2{,}783$), and it was associated with greater moral condemnation and punitive sentiment towards a politician who told a falsehood. The effect was driven by falsehoods aligned with one's political views (see Figure 6), and it exacerbated a tendency to condemn such falsehoods less harshly than falsehoods that were misaligned with one's politics (see Figure 7). When judging alleged dishonesty, Americans already show dramatic partisan disagreement (Washington Post-ABC News National Poll, 2015); the present results show how counterfactual thinking can increase this political polarization.

What explains these effects? People may evaluate a falsehood's ethicality based on its closeness to reality, and an event that did not occur will feel closer to reality when a counterfactual about how it *could have* occurred is both salient and plausible (Kahneman & Miller, 1986; Petrocelli et al., 2011). The studies' manipulations made the counterfactual salient, and participants' political preferences shaped perceptions of its plausibility. In support of this mechanism, mediation analysis showed that an identical counterfactual seemed more plausible (potent) when it aligned with participants' political preferences, and the more plausible participants found it, the more willing they were to excuse the

corresponding falsehood. No mediation analysis can demonstrate a *causal* mechanism because the mediator is measured, not manipulated (see Fiedler, Schott, & Meiser, 2011). However, the results fit with the idea that partisan differences in plausibility judgments explain why reflecting on a counterfactual increased political polarization in moral judgments.

I have argued that mentally undoing a falsehood makes it seem closer to reality and thus less unethical to tell. A potential alternative interpretation of the results is that reflecting on a counterfactual puts people in a *mental-simulation mindset* (Galinsky & Moskowitz, 2000; Hirt, Kardes, & Markman, 2004) that changes the way they evaluate falsehoods. These two interpretations represent, respectively, a *content-specific* pathway, in which the counterfactual itself provides insights specifically related to the falsehood, and a *content-neutral* pathway, in which the counterfactual affects information processing in general (Epstude & Roese, 2008; Roese & Epstude, in press). The present results are more consistent with the content-specific pathway. Study 2 found that imagining a counterfactual reduced unethicality judgments more than imagining a future event – even though both tasks required mental simulation and therefore could be expected to foster a mental-simulation mindset. Additionally, Studies 1 and 2 found the effect emerged most strongly when people perceived the specific counterfactuals as plausible. An implausible counterfactual should not provide new insight into the falsehood, but can still be expected to foster a mental-simulation mindset.

The results make several theoretical contributions. First, they advance understanding of how motivated reasoning affects moral judgments. People are reluctant to jump to desired moral conclusions without evidence, but will strategically construe ambiguous information as supporting these conclusions (Ames & Fiske, 2015; Ditto, Pizarro, & Tannenbaum, 2009; Effron, 2014). Counterfactuals may represent one source of ambiguous information. Even

when people are motivated to excuse a falsehood, it can be difficult to convince themselves it is literally true, because facts can be checked. It is easier to convince themselves it *could have been* true, because counterfactuals cannot be falsified; history cannot be rerun to test what would have occurred in alternative circumstances. Thus, counterfactuals provide a degree of freedom people can exploit to make motivated moral judgments.

Second, the results contribute to a growing understanding of counterfactual thinking's role in people's moral lives. Previous work focused on the moral consequences of imagining counterfactual actions. For example, people feel licensed to act less virtuously when they reflect counterfactually on the transgressions they *declined* to perform (Effron, Miller, & Monin, 2012; Effron, Monin, & Miller, 2013), they feel guilty when they think counterfactually about the transgressions they *did* perform (Mandel & Dhami, 2005; Niedenthal, Tangney, & Gavanski, 1994; Gaspar, Seabright, Reynolds, & Yam, 2015), and they judge others' moral character based on counterfactual actions they imagine the others *could have* or *would have* performed (Miller, Visser, & Staub, 2005; Newman & Cain, 2014). By contrast, the present research focused on the moral consequences of imagining how a falsehood could have been true. Mentally undoing a falsehood reduces the moral condemnation it receives.

Third, the studies contribute to a debate over whether counterfactual thinking has mainly functional or dysfunctional consequences (see Roese & Epstude, in press). On the functional side, counterfactuals prepare people for effective goal pursuit following failure (Epstude & Roese, 2008; Roese, 1997). On the dysfunctional side, counterfactuals can interfere with learning and memory (Petrocelli & Harris, 2011; Petrocelli, Rubin, & Stevens, 2016; Petrocelli, Seta, & Seta, 2013). By examining counterfactual thinking's moral and political consequences, the present research adds arguments to both sides. To the extent counterfactuals shield people from the discomfort of viewing admired leaders in a negative

light, they could be considered functional for individuals. But to the extent counterfactuals increase political polarization in judgments of falsehoods, and perhaps even license dishonesty by shielding these leaders from reproach, the consequences could be considered dysfunctional for society.

Future research should examine whether the effect ever reverses. The Reflection and Evaluation Model of Comparative Thinking (Markman & McMullen, 2003, 2005) holds that merely imagining counterfactual outcomes – i.e., *counterfactual reflection* – will make people feel as if the outcomes *did* occur. By contrast, directly comparing counterfactual outcomes to reality – i.e., *counterfactual evaluation* – will highlight they did *not* occur. Situations fostering counterfactual evaluation could make falsehoods seem further from the truth and thus *more* unethical to tell. Research suggests that directly instructing people to engage in reflection versus evaluation can moderate the effect of counterfactual thinking on overconfidence (Petrocelli & Crysel, 2009). Perhaps such manipulations could also moderate the effect of counterfactuals on moral judgments.

Future research should also examine whether this effect depends on the number of exposures to a particular counterfactual. In contrast to the present studies, real political contexts provide multiple opportunities to encounter the same counterfactual about how a falsehood could have been true. Such counterfactuals may get repeated in talking points, echoed throughout the news cycle, and shared on social media. On one hand, repeated exposure makes a counterfactual easier to imagine (De Brigard, Szpunar, & Schacter, 2013), which could amplify its exculpating effect. On the other hand, repeated exposure can also reduce a counterfactual's subjective plausibility (De Brigard et al., 2013), which should diminish its exculpating effect. Testing these competing predictions would have both theoretical and practical value.

The present studies prompted participants to consider counterfactuals that "undid" falsehoods, but people are likely to generate such counterfactuals even without prompting. For example, negative affect causes counterfactual thought (Roese, 1997); perhaps the cognitive dissonance experienced when facts contradict an appealing falsehood triggers thoughts about how the falsehood could have been true. People may also strategically generate and publicize counterfactuals when motivated to excuse themselves or others for advancing a patently false claim. Thus, both dissonance and self-presentational concerns could motivate counterfactual thought about falsehoods.

The rise of social media has allowed misinformation about everything from politics to health to history to spread at unprecedented speed, raising concerns about the difficulty of distinguishing fact from fiction. Even when people initially recognize information as false, they may later mistake their familiarity with it as a sign of truthfulness (Fazio, Brashier, Payne, & Marsh, 2015; Garry & Polaschek, 2000; Skurnik, Yoon, Park, & Schwarz, 2005). The present research raises a different concern. Mere exposure to misinformation, perhaps especially in the form of vivid narratives like "fake news" stories, may encourage people to imagine how it could have been true. As a result, even when people remember the misinformation as false, they may be more inclined to let public figures off the hook for asserting it as true.

How can one blunt this effect? Warning about persuasion attempts and presenting weak arguments can "inoculate" against persuasion by subsequent, stronger arguments (Banas & Rains, 2010; McGuire, 1964). Perhaps warning about attempts to use counterfactuals to excuse dishonesty, and presenting implausible counterfactuals (e.g., *Trump's inauguration would have been bigger if Clinton had revealed she had voted for him*), could similarly inoculate against the subsequent influence of more potent counterfactuals. Another strategy could be to encourage reflection on how the falsehood

would not have been true *even if* circumstances had been different (i.e., *semi-factual thought*; Byrne, 2016). For example, the proposition *Trump's inauguration would have been bigger if the weather had been better* might seem less compelling when one considers whether *it would have been the same size even if it had been an hour shorter.* This strategy should be most effective when the *even-if* thought seems highly plausible (cf. Hirt & Markman, 1995). Future research should test these strategies' effectiveness.

Conclusion

Pundits claim people privilege ideology over facts in our "post-truth" world (The Economist, 2016), but the present research suggests more nuance. Regardless of political views, participants condemned falsehoods. However, falsehoods supporting their views received less condemnation – and merely considering a counterfactual magnified this effect. Thus, partisans may not ignore facts, but readily excuse falsehoods based on weak justifications. We should thus be wary of our ability to imagine alternatives to reality. When leaders we support encourage us to consider how their lies could have been true, we may hold them to laxer ethical standards.

References

Abrams, D., Randsley de Moura, G., & Travaglino, G. A. (2013). A double standard when group members behave badly: Transgression credit to ingroup leaders. *Journal of Personality and Social Psychology, 105*(5), 799. doi: 10.1037/a0033600

Ames, D. L., & Fiske, S. T. (2015). Perceived intent motivates people to magnify observed harms. *Proceedings of the National Academy of Sciences, 112*(12), 3599-3605.

Anderson, C. A. (1983). Imagination and expectation: The effect of imagining behavioral scripts on personal influences. *Journal of Personality and Social Psychology, 45*(2), 293.

Arendt, H. (1961). *Between past and future: Six exercises in political thought*. New York: Viking.

Banas, J. A., & Rains, S. A. (2010). A meta-analysis of research on inoculation theory. *Communication Monographs, 77*(3), 281-311.

Briazu, R. A., Walsh, C. R., Deeprose, C., & Ganis, G. (2017). Undoing the past in order to lie in the present: Counterfactual thinking and deceptive communication. *Cognition, 161*, 66-73.

Byrne, R. M. J. (2016). Counterfactual thought. *Annual Review of Psychology, 67*, 135-157.

Carroll, J. S. (1978). The effect of imagining an event on expectations for the event: An interpretation in terms of the availability heuristic. *Journal of Experimental Social Psychology, 14*(1), 88-96.

De Brigard, F., Szpunar, K. K., & Schacter, D. L. (2013). Coming to grips with the past: Effect of repeated simulation on the perceived plausibility of episodic counterfactual thoughts. *Psychological Science, 24*(7), 1329-1334. doi: 10.1177/0956797612468163

Dent, M. (2016). Sorry, Hillary: Trump actually has made a few things in the USA. *Politifact*. Retrieved from

http://www.politifact.com/pennsylvania/statements/2016/aug/01/hillary-clinton/hillary-clinton-went-overboard-calling-out-donald-/

Ditto, P. H., Pizarro, D. A., & Tannenbaum, D. (2009). Motivated moral reasoning. In H. R. Brian (Ed.), *Psychology of Learning and Motivation* (Vol. 50, pp. 307-338): Academic Press.

Effron, D. A. (2014). Making mountains of morality from molehills of virtue: Threat causes people to overestimate their moral credentials. *Personality and Social Psychology Bulletin, 40*(8), 972-985. doi: 10.1177/0146167214533131

Effron, D. A., Miller, D. T., & Monin, B. (2012). Inventing racist roads not taken: The licensing effect of immoral counterfactual behaviors. *Journal of Personality and Social Psychology, 103*, 916-932. doi: 10.1037/a0030008

Effron, D. A., Monin, B., & Miller, D. T. (2013). The unhealthy road not taken: Licensing indulgence by exaggerating counterfactual sins. *Journal of Experimental Social Psychology, 49*, 573-578. doi: 10.1016/j.jesp.2012.08.012

Epstude, K., & Roese, N. J. (2008). The functional theory of counterfactual thinking. *Personality and Social Psychology Review, 12*(2), 168-192.

Fazio, L. K., Brashier, N. M., Payne, B. K., & Marsh, E. J. (2015). Knowledge does not protect against illusory truth. *Journal of Experimental Psychology: General, 144*(5), 993.

Fiedler, K., Schott, M., & Meiser, T. (2011). What mediation analysis can (not) do. *Journal of Experimental Social Psychology, 47*(6), 1231-1236.

Galinsky, A. D., & Moskowitz, G. B. (2000). Counterfactuals as behavioral primes: Priming the simulation heuristic and consideration of alternatives. *Journal of Experimental Social Psychology, 36*(4), 384-409.

Garry, M., & Polaschek, D. L. (2000). Imagination and memory. *Current Directions in Psychological Science, 9*(1), 6-10.

Gaspar, J. P., Seabright, M. A., Reynolds, S. J., & Yam, K. C. (2015). Counterfactual and factual reflection: The influence of past misdeeds on future immoral behavior. *The Journal of Social Psychology, 155*(4), 370-380.

Gerlach, K. D., Dornblaser, D. W., & Schacter, D. L. (2014). Adaptive constructive processes and memory accuracy: Consequences of counterfactual simulations in young and older adults. *Memory, 22*(1), 145-162. doi: 10.1080/09658211.2013.779381

Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology, 106*(1), 148-168. doi: 10.1037/a0034726

Gregory, W. L., Cialdini, R. B., & Carpenter, K. M. (1982). Self-relevant scenarios as mediators of likelihood estimates and compliance: Does imagining make it so? *Journal of Personality and Social Psychology, 43*(1), 89.

Hilbig, B. E., & Hessler, C. M. (2013). What lies beneath: How the distance between truth and lie drives dishonesty. *Journal of Experimental Social Psychology, 49*(2), 263-266.

Hirt, E. R., Kardes, F. R., & Markman, K. D. (2004). Activating a mental simulation mind-set through generation of alternatives: Implications for debiasing in related and unrelated domains. *Journal of Experimental Social Psychology, 40*(3), 374-383.

Hirt, E. R., & Markman, K. D. (1995). Multiple explanation: A consider-an-alternative strategy for debiasing judgments. *Journal of Personality and Social Psychology, 69*(6), 1069-1086. doi: 10.1037/0022-3514.69.6.1069

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review, 93*(2), 136-153.

Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In D. Kahneman, P. Slovic & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 201-208). New York: Cambridge University Press.

Koch, A. S., & Forgas, J. P. (2012). Feeling good and feeling truth: The interactive effects of mood and processing fluency on truth judgments. *Journal of Experimental Social Psychology, 48*(2), 481-485.

Koehler, D. J. (1991). Explanation, imagination, and confidence in judgment. *Psychological Bulletin, 110*(3), 499.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108*(3), 480-498.

Levine, E. E., & Schweitzer, M. E. (2014). Are liars ethical? On the tension between benevolence and honesty. *Journal of Experimental Social Psychology, 53*, 107-117.

Mandel, D. R., & Dhami, M. K. (2005). "What I did" versus "what I might have done": Effect of factual versus counterfactual thinking on blame, guilt, and shame in prisoners. *Journal of Experimental Social Psychology, 41*(6), 627-635. doi: 10.1016/j.jesp.2004.08.009

Markman, K. D., & Hirt, E. R. (2002). Social prediction and the “allegiance bias”. *Social Cognition, 20*(1), 58-86. doi: 10.1521/soco.20.1.58.20943

Markman, K. D., & McMullen, M. N. (2003). A reflection and evaluation model of comparative thinking. *Personality and Social Psychology Review, 7*(3), 244-267. doi: 10.1207/s15327957pspr0703_04

Markman, K. D., & McMullen, M. N. (2005). Reflective and evaluative modes of mental simulation. In D. R. Mandel, D. J. Hilton & P. Catellani (Eds.), *The psychology of counterfactual thinking*. Abington, Oxfordshire, UK: Routledge.

Markman, K. D., & Tetlock, P. E. (2000). Accountability and close-call counterfactuals: The loser who nearly won and the winner who nearly lost. *Personality and Social Psychology Bulletin, 26*(10), 1213-1224.

McGuire, W. J. (1964). Inducing resistence to persuasion: Some contemporary approaches. *Advances in Experimental Social Psychology, 1*, 191-229.

Medvec, V. H., Madey, S. F., & Gilovich, T. (1995). When less is more: Counterfactual thinking and satisfaction among Olympic medalists. *Journal of Personality and Social Psychology, 69*(4), 603-610. doi: 10.1037/0022-3514.69.4.603

Miller, D. T., Turnbull, W., & McFarland, C. (1989). When a coincidence is suspicious: The role of mental simulation. *Journal of Personality and Social Psychology, 57*(4), 581-589.

Miller, D. T., Visser, P. S., & Staub, B. D. (2005). How surveillance begets perceptions of dishonesty: The case of the counterfactual sinner. *Journal of Personality and Social Psychology, 89*(2), 117-128.

Mueller, A. B., & Skitka, L. J. (in press). Liars, damned liars, and zealots: The effect of moral mandates on transgressive advocacy acceptance. *Social Psychological and Personality Science*. doi: 10.1177/1948550617720272

Newman, G. E., & Cain, D. M. (2014). Tainted altruism: When doing some good is evaluated as worse than doing no good at all. *Psychological Science, 25*(3), 648-655.

Niedenthal, P. M., Tangney, J. P., & Gavanski, I. (1994). "If only I weren't" versus "If only I hadn't": Distinguishing shame and guilt in counterfactual thinking. *Journal of Personality and Social Psychology, 67*, 585-595.

Pagano, S. J., & Huo, Y. J. (2007). The role of moral emotions in predicting support for political actions in post-war Iraq. *Political Psychology, 28*(2), 227-255.

Petrocelli, J. V., & Crysel, L. C. (2009). Counterfactual thinking and confidence in blackjack: A test of the counterfactual inflation hypothesis. *Journal of Experimental Social Psychology, 45*(6), 1312-1315. doi: 10.1016/j.jesp.2009.08.004

Petrocelli, J. V., & Harris, A. K. (2011). Learning inhibition in the Monty Hall Problem: The role of dysfunctional counterfactual prescriptions. *Personality and Social Psychology Bulletin, 37*(10), 1297-1311.

Petrocelli, J. V., Kammrath, L. K., Brinton, J. E., Uy, M. R. Y., & Cowens, D. F. (2015). Holding on to what might have been may loosen (or tighten) the ties that bind us: A counterfactual potency analysis of previous dating alternatives. *Journal of Experimental Social Psychology, 56*, 50-59.

Petrocelli, J. V., Percy, E. J., Sherman, S. J., & Tormala, Z. L. (2011). Counterfactual potency. *Journal of Personality and Social Psychology, 100*(1), 30-46.

Petrocelli, J. V., Rubin, A. L., & Stevens, R. L. (2016). The sin of prediction: When mentally simulated alternatives compete with reality. *Personality and Social Psychology Bulletin, 42*(12), 1635-1652. doi: 10.1177/0146167216669122

Petrocelli, J. V., Seta, C. E., & Seta, J. J. (2012). Prefactual potency: The perceived likelihood of alternatives to anticipated realities. *Personality and Social Psychology Bulletin, 38*(11), 1467-1479.

Petrocelli, J. V., Seta, C. E., & Seta, J. J. (2013). Dysfunctional counterfactual thinking: When simulating alternatives to reality impedes experiential learning. *Thinking & Reasoning, 19*(2), 205-230.

Politifact. (n.d.). Comparing Hillary Clinton, Donald Trump on the Truth-O-Meter. Retrieved August 11, 2017, from http://www.politifact.com/truth-o-meter/lists/people/comparing-hillary-clinton-donald-trump-truth-o-met/

Prentice, D. A., & Miller, D. T. (1992). When small effects are impressive. *Psychological Bulletin, 112*(1), 160.

Roese, N. J. (1997). Counterfactual thinking. *Psychological Bulletin, 121*(1997), 133-148.

Roese, N. J., & Epstude, K. (in press). The functional theory of counterfactual thinking: New evidence, new challenges, new insights. *Advances in Experimental Social Psychology*.

Rogers, T., Zeckhauser, R., Gino, F., Norton, M. I., & Schweitzer, M. E. (2017). Artful paltering: The risks and rewards of using truthful statements to mislead others. *Journal of Personality and Social Psychology, 112*(3), 456.

Rossoll, N. (2017, January 25). Conway: Crowd size at Trump's inauguration 'was historic,' considering projections and rain, *ABC News*. Retrieved from http://abcnews.go.com/ThisWeek/crowd-size-trumps-inauguration-historic-projections-rain/story?id=44968944

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. doi: 10.1111/j.1751-9004.2011.00355.x

Schweitzer, M. E., & Hsee, C. K. (2002). Stretching the truth: Elastic justification and motivated communication of uncertain information. *The Journal of Risk and Uncertainty, 25*(2), 185-201.

Shalvi, S., Dana, J., Handgraaf, M. J. J., & De Dreu, C. K. W. (2011). Justified ethicality: Observing desired counterfactuals modifies ethical perceptions and behavior. *Organizational Behavior and Human Decision Processes, 115*(2), 181-190.

Sherman, S. J., Cialdini, R. B., Schwartzman, D. F., & Reynolds, K. D. (1985). Imagining can heighten or lower the perceived likelihood of contracting a disease the mediating effect of ease of imagery. *Personality and Social Psychology Bulletin, 11*(1), 118-127.

Shidlovski, D., Schul, Y., & Mayo, R. (2014). If I imagine it, then it happened: The implicit truth value of imaginary representations. *Cognition, 133*(3), 517-529.

Skitka, L. J., Bauman, C. W., & Sargis, E. G. (2005). Moral conviction: Another contributor to attitude strength or something more? *Journal of Personality and Social Psychology, 88*(6), 895-917.

Skurnik, I., Yoon, C., Park, D. C., & Schwarz, N. (2005). How warnings about false claims become recommendations. *Journal of Consumer Research, 31*(4), 713-724.

Tetlock, P. E. (1998). Close-call counterfactuals and belief-system defenses: I was not almost wrong but I was almost right. *Journal of Personality and Social Psychology, 75*(3), 639-652. doi: 10.1037/0022-3514.75.3.639

Tetlock, P. E., & Henik, E. (2005). Theory- versus imagination-driven thinking about historical counterfactuals: Are we prisoners of our preconceptions? In D. R. Mandel, D. J. Hilton & P. Catellani (Eds.), *The psychology of counterfactual thinking* (pp. 199-216). New York: Routledge.

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

The Economist. (2016, September 10). The post-truth world: Yes, I'd lie to you Retrieved June 7, 2017, from http://www.economist.com/news/briefing/21706498-dishonesty-politics-nothing-new-manner-which-some-politicians-now-lie-and

Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., . . . Boutin, B. (2016). Bayesian inference for psychology. Part II: example applications with JASP. *Psychonomic Bulletin & Review*, 1-19.

Washington Post-ABC News National Poll. (2015). 2016 primaries, political dysfunction and Obama. Retrieved June 26, 2017 https://www.washingtonpost.com/page/2010-2019/WashingtonPost/2015/09/14/National-Politics/Polling/release_409.xml

Tables

Table 1

Stimuli in Studies 1-3

Fact	Counterfactual (Studies 1 and 2 only)	Control statement (Study 2 only)	Falsehood	Falsehood aligns with politics of …
It's a proven fact that fewer people attended Donald Trump's inauguration than Barack Obama's.	Some ticketholders were unable to attend Trump's inauguration because the lines were too long at security. Consider the following thought: *If security had been less tight at Trump's inauguration, then more people would have attended it than Obama's inauguration.*	Think about how large the next presidential inauguration might be. Consider the following thought: *If a Republican is elected president in 2020, then at least 100,000 people will attend the inauguration.*	More people attended Trump's inauguration than Obama's	Trump supporters
It's a proven fact that Donald Trump won the electoral vote, but lost the popular vote to Hillary Clinton	Trump did not campaign for the popular vote, because the law says that the winner of the electoral vote wins the presidency. Consider the following thought: *If Trump had tried to win the popular vote, then he would have won the popular vote.*	Senator Mitch McConnell is a Republican from Kentucky. He is currently the Senate Majority Leader. Consider the following thought: *If Mitch McConnell runs for President in 2020, then he will win the popular vote.*	Trump won the popular vote	Trump supporters
During the 2016 election campaign, the FBI investigated Hillary Clinton for improper use of a private email server. It's a proven fact that the FBI never brought charges against her.	Investigators did not see all of the emails stored on the private server because Clinton deleted some of them. Some people wonder whether the deleted emails contained evidence that Clinton broke the law. Consider the following thought: *If investigators had been able to see the deleted emails, then the FBI would have brought charges against Hillary Clinton.*	The current head of the FBI is James Comey, who does not have a military background. Some people wonder who the next FBI director will be. Consider the following thought: *If the next director of the FBI has experience in the military, then more FBI agents will carry firearms.*	The FBI brought charges against Hillary Clinton	Trump supporters

(***table continues on next page)***

Table 1, continued

Fact	Counterfactual (Studies 1 and 2 only)	Control statement (Study 2 only)	Falsehood	Falsehood aligns with politics of …
It's a proven fact that Hillary Clinton lost the electoral vote to Donald Trump.	Some people argue that new state laws passed by Republicans made it more difficult for Clinton supporters to cast their votes. For example, they argue that stricter voter ID laws made voting harder for poor and minority citizens, who tend to support Clinton. Consider the following thought: *If Republicans had not passed stricter voting laws, then Hillary Clinton would have won the electoral vote.*	Some people wonder who will run for president in 2020. For example, some wonder whether Senator Elizabeth Warren, a Democrat from Massachusetts, will choose to run. Consider the following thought: *If Elizabeth Warren runs for president in 2020, then she will win the electoral vote.*	Hillary Clinton won the electoral vote	Clinton supporters
It's a proven fact that Trump-brand hats, wine, and water are made in the USA. `	Consider the following thought: *If Trump had been able to make those products more cheaply in a different country, then he would have made them outside the USA*.	Consider the following thought: *If an American car manufacturer moves a factory outside the USA, then the quality of the cars will decrease.*	No Trump-brand products are made in the USA	Clinton supporters
When Barack Obama was president, he placed a bust of Martin Luther King, Jr., in the oval office in the White House. It's a proven fact that the bust has remained in the oval office since Donald Trump became president, and has never been removed.	Consider the following thought: *If it had been possible for Trump to remove the MLK bust without the public finding out, then Trump would have removed it.*	Consider the following thought: *If a Democrat is President in 50 years, then there will be a bust of Obama in the oval office.*	Trump removed the bust of Martin Luther King, Jr., from the oval office	Clinton supporters

Note. In Study 1 the counterfactuals were worded slightly differently: The phrase “consider the following thought” was replaced with “some people might think” and some of the counterfactuals were phrased “if only … then” instead of “if … then.”

Table 2

Regression Results for Unethicality Ratings in Study 1

DV: Unethicality ratings		*b*	*SE(b)*	*z*	*p*	95% Confidence interval of *b*	
Step 1							
	condition	-4.70	1.05	-4.46	0.000	-6.76	-2.63
	(constant)	82.24	0.74	110.60	0.000	80.79	83.70
Step 2							
	condition	-3.71	1.15	-3.24	0.001	-5.96	-1.46
	alignment	-7.60	0.64	-11.96	0.000	-8.84	-6.35
	condition X alignment	-1.93	0.90	-2.14	0.032	-3.70	-0.16
	(constant)	86.03	0.81	106.40	0.000	84.45	87.62

Note. *Condition* was coded 1 = counterfactual, 0 = control. *Alignment* was coded = 1 aligned, 0 = misaligned. The regression also included a random effect for participant.

Table 3

Regression Results for Study 2

DV: Unethicality ratings	*b*	*SE(b)*	*z*	*p*	95% Confidence interval of *b*	
Unethicality ratings						
Step 1						
condition	-2.14	1.03	-2.08	0.038	-4.16	-0.12
(constant)	78.26	0.73	107.23	0.000	76.83	79.69
Step 2						
condition	-0.83	1.30	-0.64	0.525	-3.38	1.73
alignment	-6.23	0.83	-7.48	0.000	-7.86	-4.60
condition X alignment	-2.80	1.19	-2.35	0.019	-5.14	-0.47
(constant)	81.66	0.91	89.50	0.000	79.88	83.45

Note. Two-tailed p-values are reported for reference; the main text reports 1-tailed values for pre-registered analyses. *Condition* was coded 1 = counterfactual, 0 = control. *Alignment* was coded 1 = aligned, 0 = misaligned. The regression also included a random effect for participant. Participants who did not support either candidate could not be included in Step 2, leaving 692.

Table 4

Regression Results for Study 3

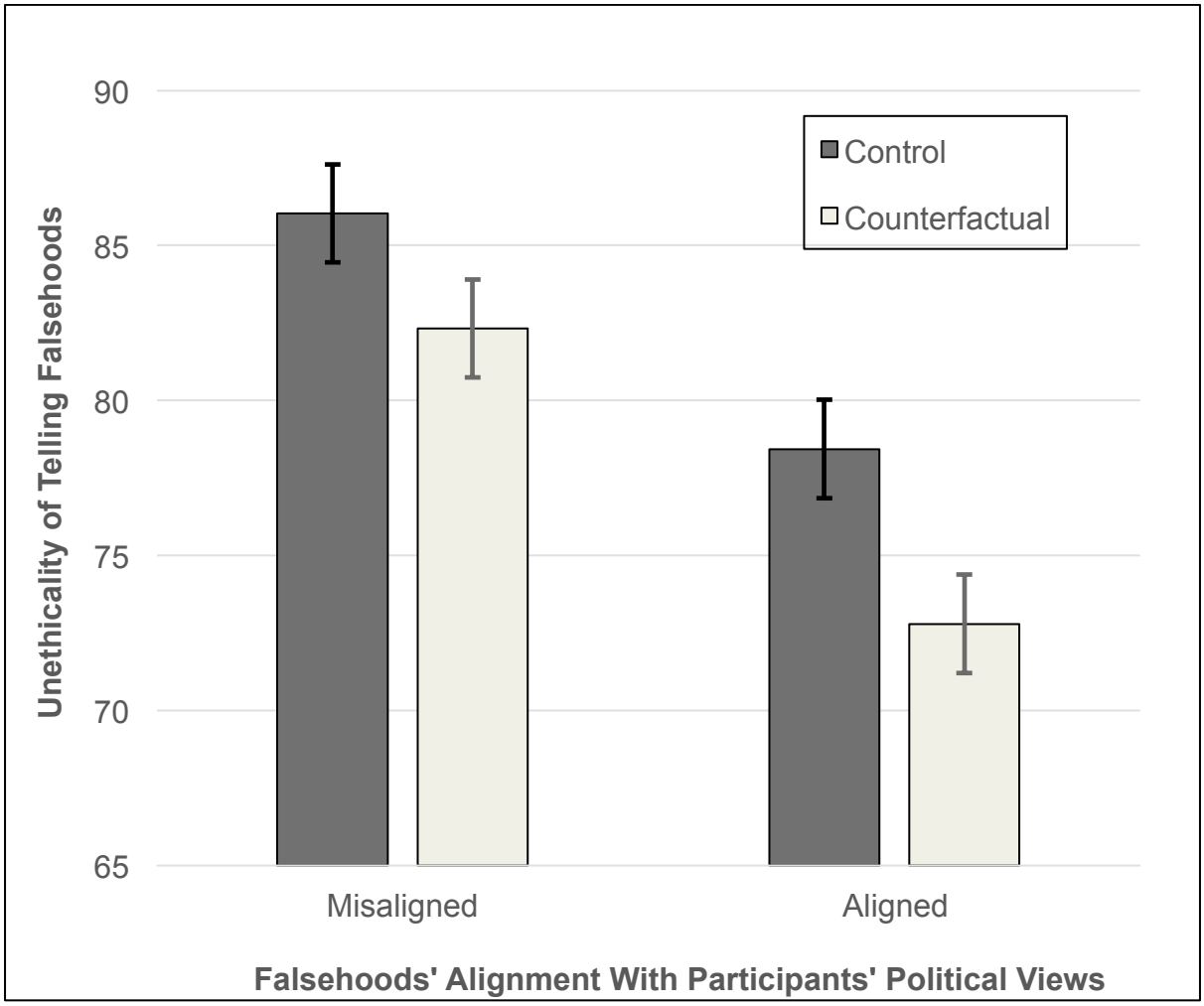
Dependent measure	*b*	*SE(b)*	*z*	*p*	95% Confidence interval of *b*	
Unethicality ratings						
Step 1						
condition	-2.12	1.06	-2.00	0.046	-4.20	-0.04
(constant)	80.69	0.74	108.59	0.000	79.24	82.15
Step 2						
condition	0.56	1.31	0.43	0.668	-2.00	3.12
alignment	-5.31	0.76	-6.97	0.000	-6.81	-3.82
condition X alignment	-4.37	1.10	-3.98	0.000	-6.52	-2.22
(constant)	83.13	0.91	91.34	0.000	81.34	84.91

Note. Two-tailed p-values are reported for reference; the main text reports 1-tailed values for pre-registered analyses. *Condition* was coded 1 = counterfactual, 0 = control. *Alignment* was coded = 1 aligned, 0 = misaligned. The regression also included a random effect for participant. Participants who did not support either candidate could not be included in Step 2, leaving 701.

Figures

Figure 1

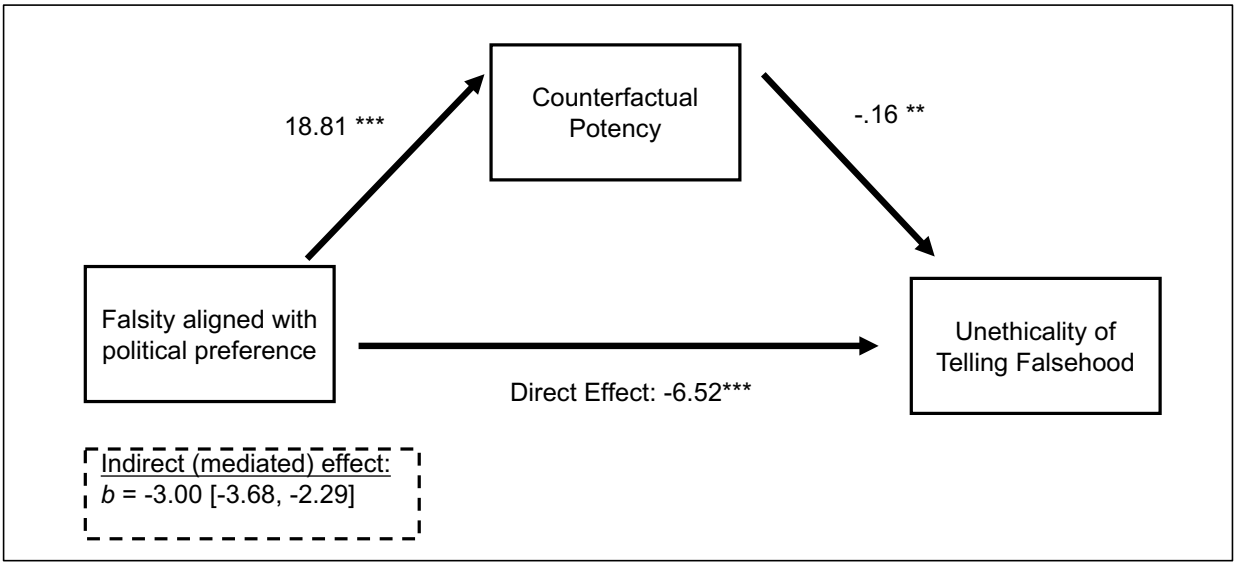
Mean Unethicality Rating, by Condition and Alignment, ±95% CI, in Study 1



Note. Means and 95% CIs computed from mixed regression analysis. Full scale of unethicality ratings is 0-100.

Figure 2

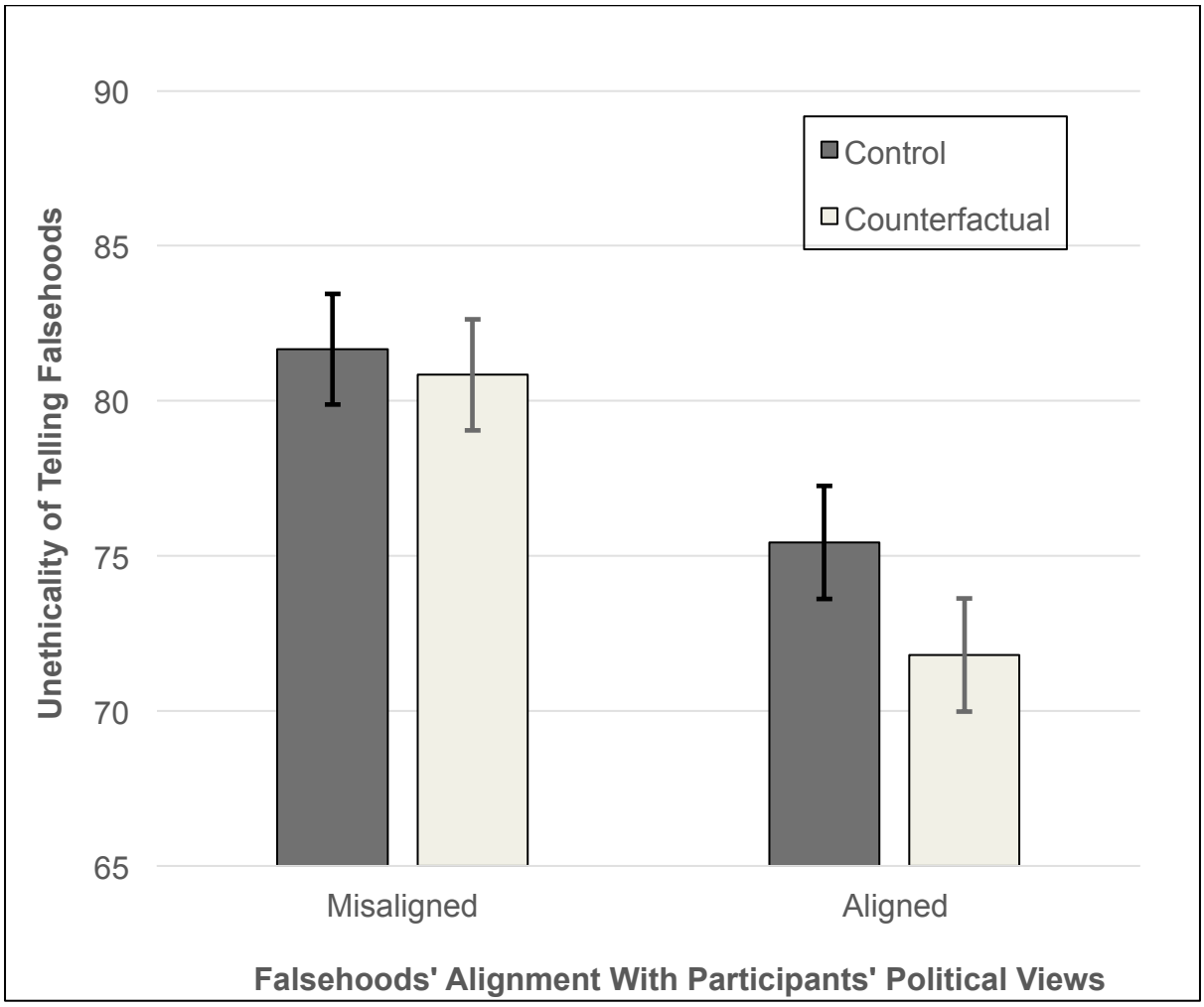
Indirect Effect of Alignment on Unethicality Judgments, Through Counterfactual Potency, in Study 1's Counterfactual Condition



Note. Coefficients are unstandardized

Figure 3

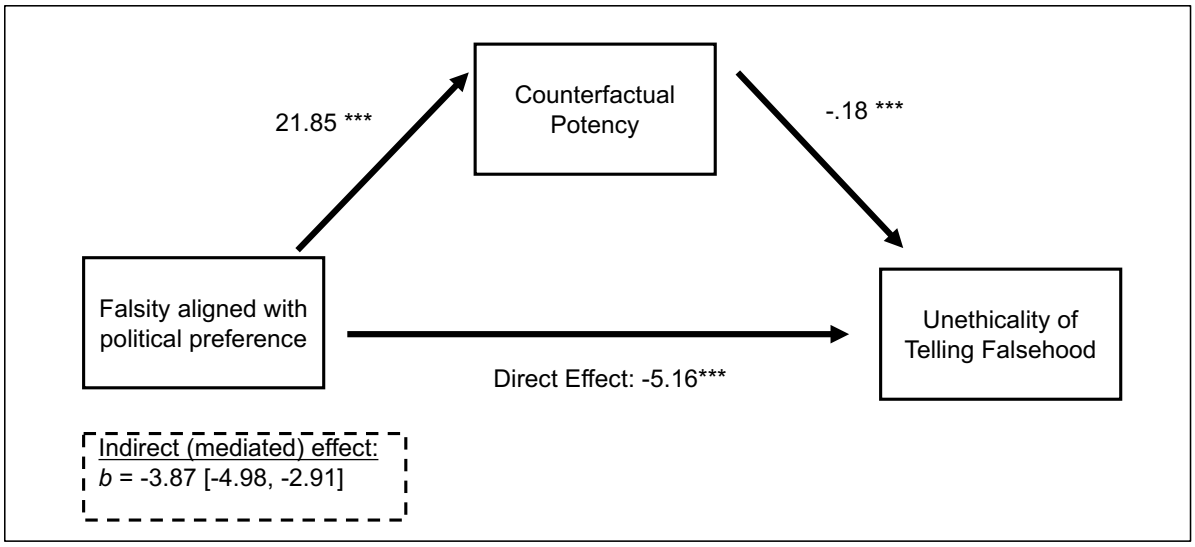
Mean Unethicality Rating, by Condition and Alignment, ±95% CI, in Study 2



Note. Means and 95% CIs computed from mixed regression analysis. Full scale of unethicality ratings is 0-100.

Figure 4

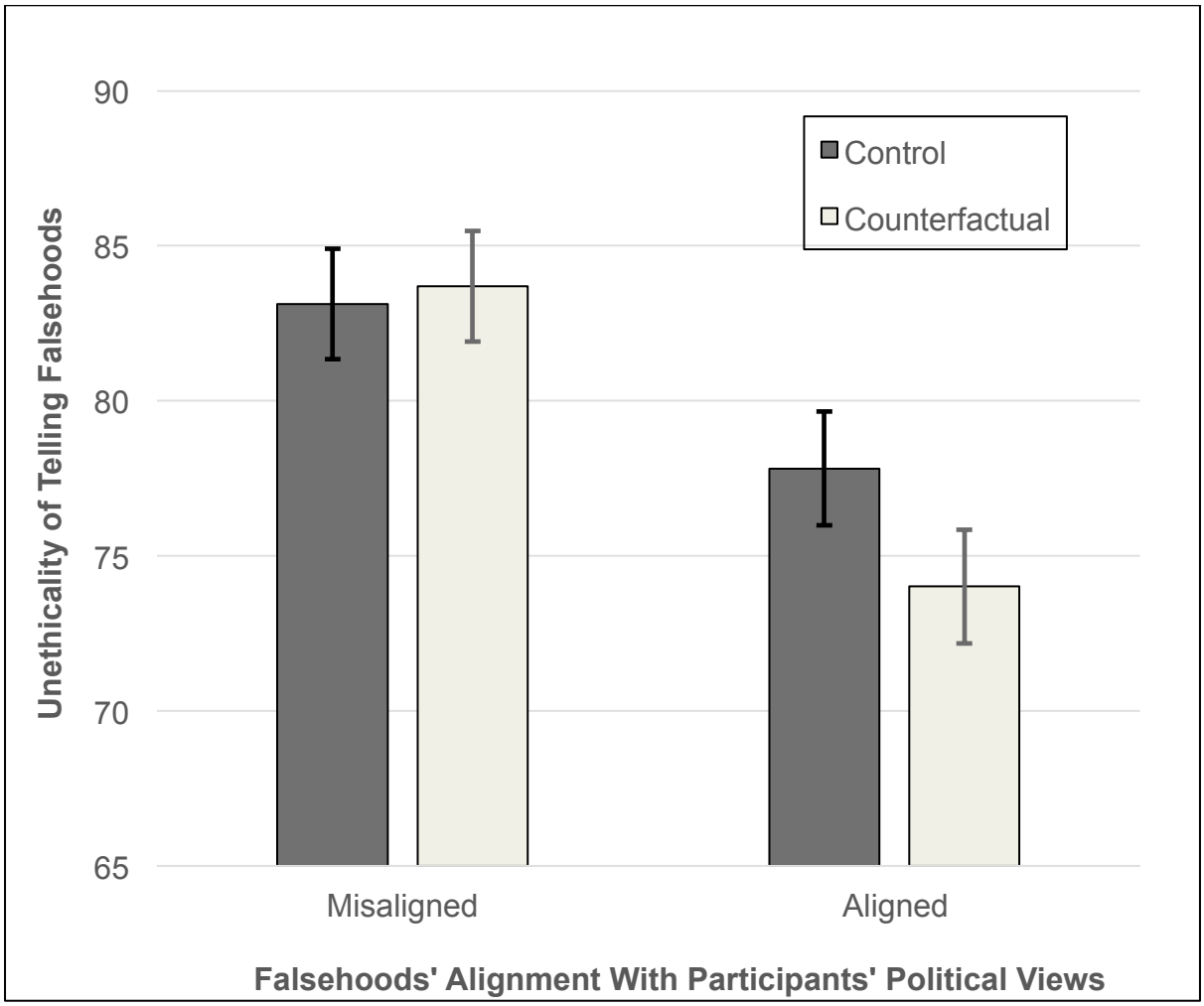
Indirect Effect of Alignment on Unethicality Judgments, Through Counterfactual Potency, in Study 2's Counterfactual Condition



Note. Coefficients are unstandardized.

Figure 5

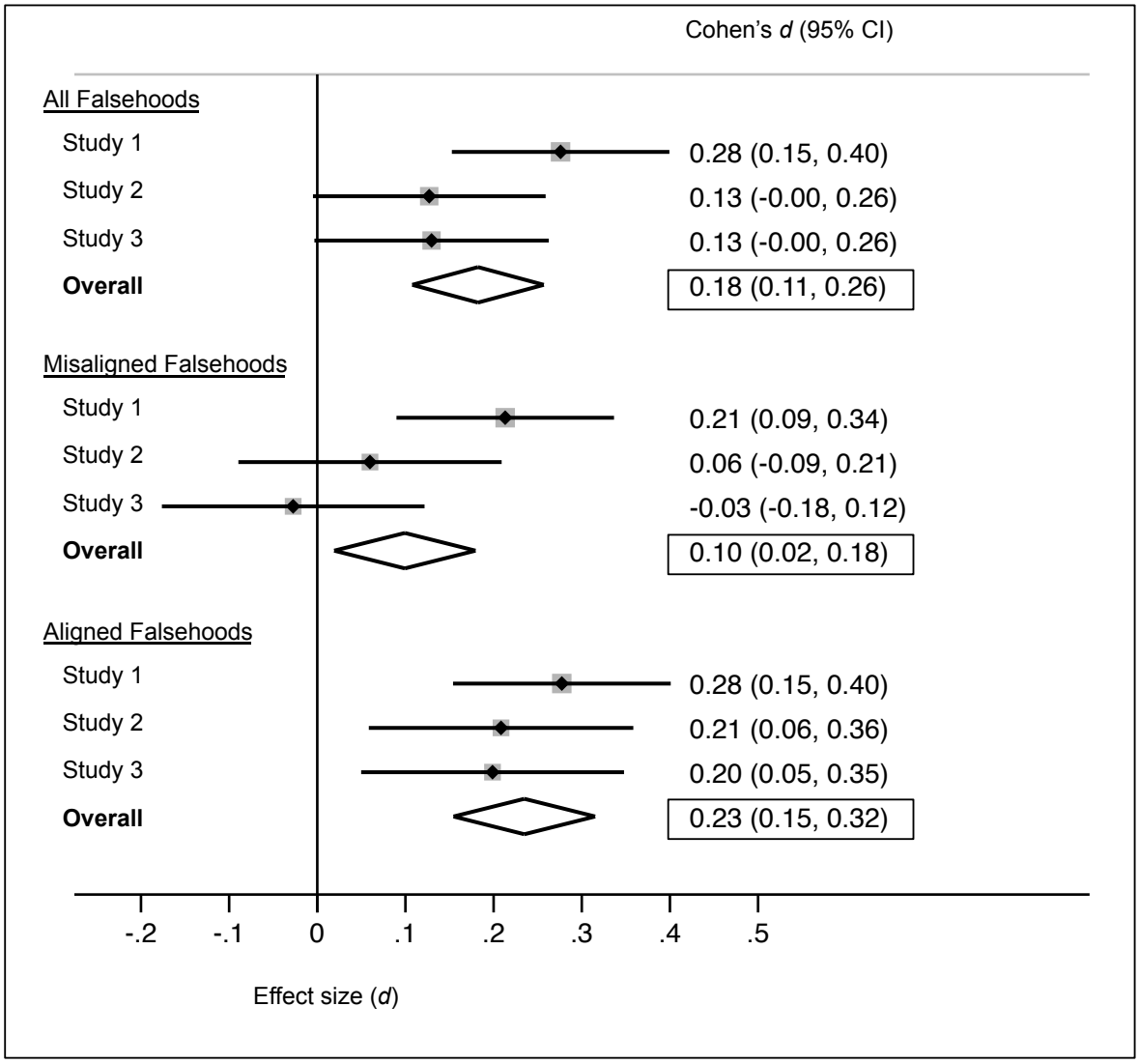
Mean Unethicality Ratings, by Condition and Alignment, ±95% CI, in Study 3



Note. Means and 95% CIs computed from mixed regression analysis. Full scale of unethicality ratings is 0-100.

Figure 6

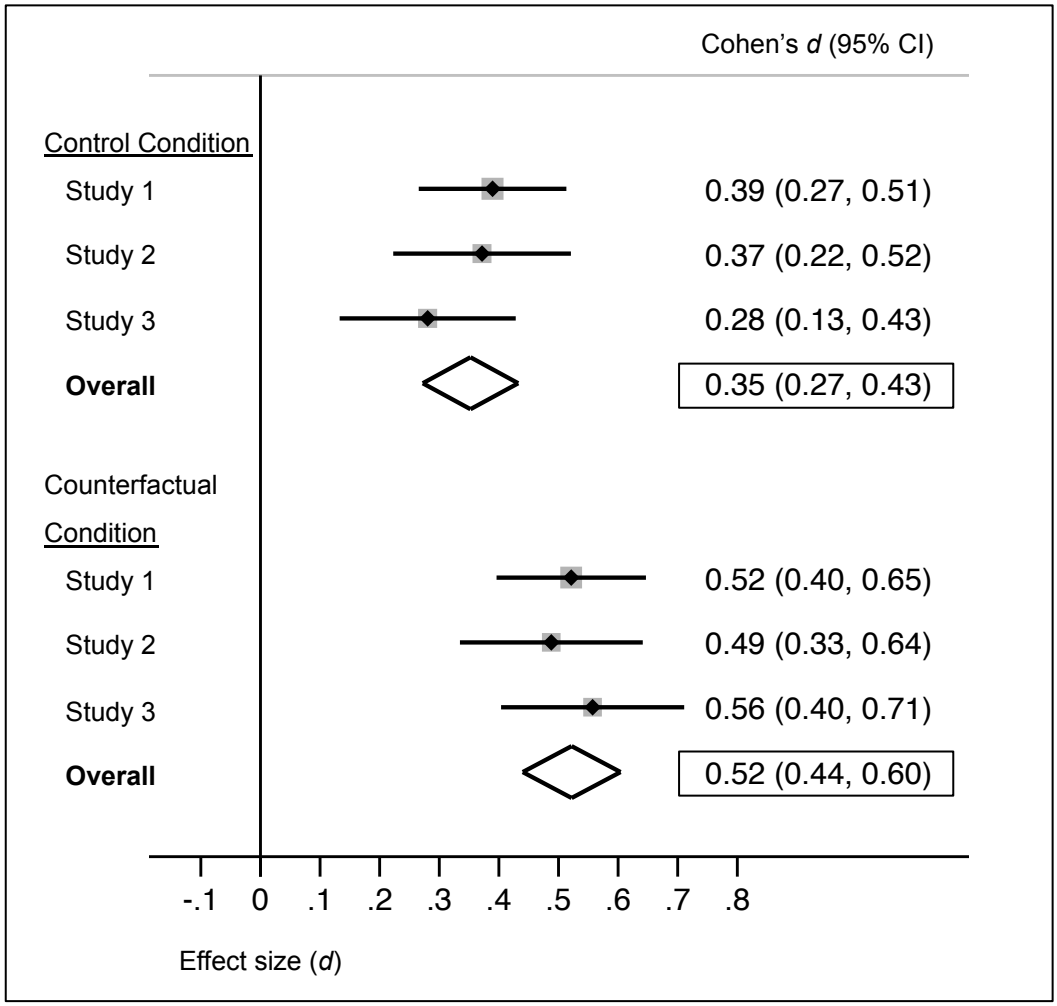
Meta-Analysis: The Counterfactual Manipulation Decreased Unethicality Judgments, Especially for Falsehoods Aligned With Political Preferences



Note. The larger the effect size, the more the counterfactual manipulation reduced the falsehoods' perceived unethicality. Lines show 95% CIs. Studies 2 and 3 pre-registered one-tailed tests, so 95% CIs underestimate statistical significance. Diamonds show 95% CI for meta-analytic effect across studies. The size of the gray square is proportional to the sample size.

Figure 7

Meta-Analysis: The Counterfactual Manipulation Increased Political Polarization



Note. The larger the effect size, the greater the tendency to perceive falsehoods as less unethical when aligned (vs. misaligned) with one's political views. Thus, larger effect sizes indicate greater political polarization. The graph shows that the counterfactual condition increased political polarization relative to the control condition. The size of the gray square is proportional to the sample size.