Supporting Materials for

# It Might Become True: How Prefactual Thinking Licenses Dishonesty

**Table of Contents**

## Study 1 and 2

All measures and analyses are reported in the main manuscript.

## Study 3

### Exploratory Measures of Voting Behavior, and Political Knowledge and Attitudes

**Voting behavior.** In Study 3, we predicted that considering how a falsehood might become true in the future would reduce condemnation of the falsehood more effectively when that falsehood-prefactual pair fit, rather than conflicted with a participants' political beliefs. In addition to the measures of political orientation reported in the main manuscript, participants also reported their recent voting behavior: "Did you vote in the most recent Presidential election? (Yes / No)" Participants who indicated "Yes" then reported: "Whom did you vote for? (A) Donald Trump, (B) Hillary Clinton, (C) None of the above". Participants who indicated "No" were then asked: "If you had voted, whom would you have voted for? (A) Donald Trump, (B) Hillary Clinton, (C) None of the above." We included these questions for exploratory purposes.

Of participants in Study 3, 669 voted in the previous election and 66 did not. Of those who voted, 293 voted for Donald Trump and 376 voted for Hillary Clinton. Of those who did not vote, 39 reported that, if they had voted, they would have voted for Donald Trump, whereas 27 reported they would have voted for Hillary Clinton.

**Political knowledge and attitudes.** At the end of the survey, we assessed participants' general political knowledge. Participants read the following: "Now we have a set of questions concerning various public figures. We want to see how much information about them gets out to the public from television, newspapers and the like. What job or political office does each of the following people [Mike Pence, Paul Ryan, Angela Merkel, Vladimir Putin, John Roberts] now hold? (Type your answer. If you don't know, please provide your best guess)." Finally, participants rated: How much do you approve or disapprove of the way Donald Trump is handling his job as President? From -3(Strongly disapprove) to +3(Strongly approve). On average, participants reported they "slightly

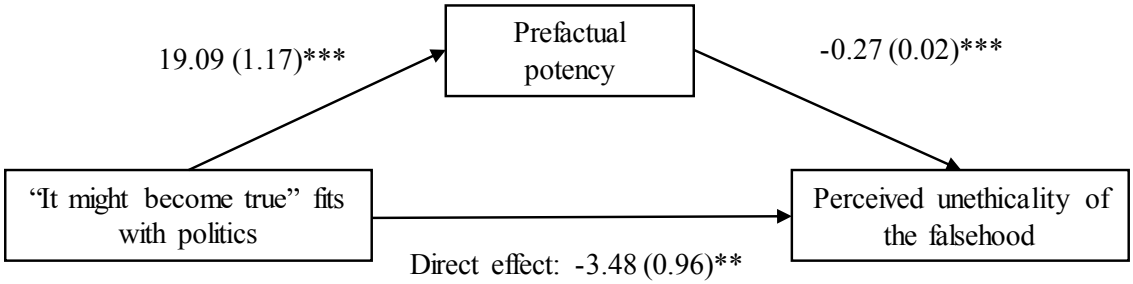
disapproved" of the way Donald Trump was handling his job as President *(M* = -0.81, *SD* = 2.26).

**Ease of Imagining the Prefactual: Prefactual Potency.**

We theorized that it is easier to imagine a falsehood might become true when that possibility fits with one's motivations and beliefs, and that the easier it is to imagine becoming true, the less unethical the falsehood will seem. In line with our theorizing, there was a significant indirect effect from fit with participants' politics, to prefactual potency, to moral judgments of the falsehood, *b* = -5.10, 95% CI [-6.00, -4.20] (see main manuscript). This analysis is limited to participants in the prefactual condition, because they were the only ones who saw and rated prefactuals about how the falsehood might become true. We illustrate this indirect effect in Figure SOM-1.

**Figure SOM-1**

*Mediation Analysis of Political Fit on Judgments of Unethicality through Prefactual Potency in Study 3*



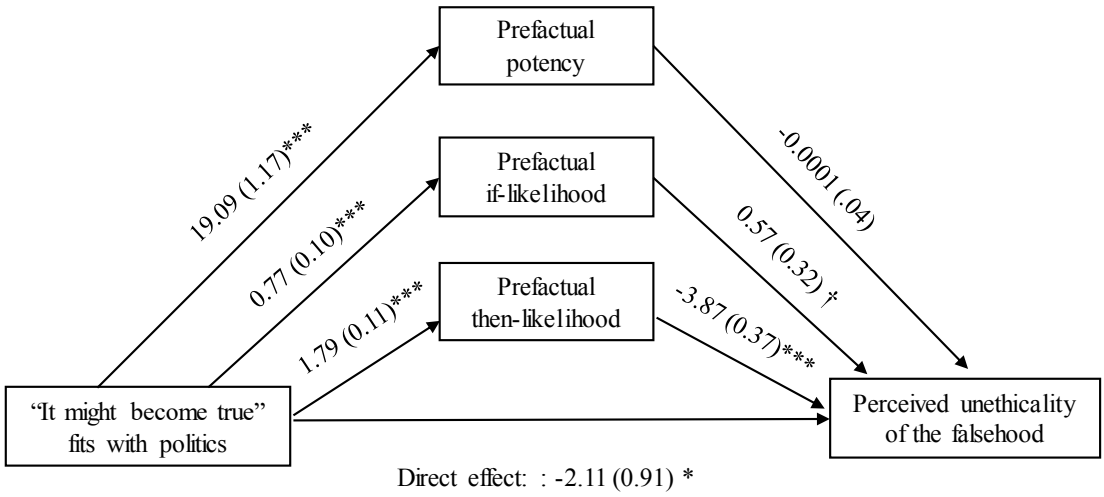

*Note.* Falsehood's fit with participants' political beliefs coded 1 = fit, 0 = conflict. Unstandardized coefficients shown.

† p < .10, *** p < .001.

Exploratory analyses suggest that the mediating effect of prefactual potency was driven by participants' judgments of the then-likelihood of the prefactual. We modelled the indirect effect of political fit on participants' moral judgments of the falsehood through the if-likelihood, then-likelihood, and prefactual potency in a multilevel mediation model with Stata's *gsem* command. There was a significant indirect effect of fit with participants' politics on moral judgments through the then-likelihood, $b$ = -6.90, 95% CI [-8.47, -5.34], but not if-likelihood, $b$ = 0.44, 95% CI [-0.05, 0.93] or potency, $b$ = 0.001, 95% CI [-1.69, 1.69] (Figure SOM-2). This analysis was limited to participants in the prefactual condition because they were the only ones who saw and rated prefactuals about how the falsehood might become true.

**Figure SOM-2**

*Mediation Analysis of Political Fit on Judgments of Unethicality through If-Likelihood, Then-Likelihood, and Potency in Study 3*




*Note.* Falsehood's fit with participants' political beliefs coded 1 = fit, 0 = conflict. Unstandardized coefficients shown.

† p < .10, *** p < .001.

## Study 4

### Exploratory Measures of Voting Behavior and Political Knowledge and Beliefs

**Voting behavior.** We included the same exploratory measures of voting behavior as in Study 3. Of participants in Study 4, 669 voted in the previous election and 134 did not. Of those who voted, 199 voted for Donald Trump, 369 voted for Hillary Clinton, and 101 reported they voted for neither of the above. Of those who did not vote, 35 reported that, if they had voted, they would have voted for Donald Trump; 46 reported that they would have voted for Hillary Clinton; and 53 reported that they would have voted for neither of the above.

**Political knowledge and approval.** At the end of the survey, we examined participants' general political knowledge. Participants read the following: "Now we have a set of questions concerning various public figures. We want to see how much information about them gets out to the public from television, newspapers and the like. What job or political office does each of the following people [Mike Pence, Paul Ryan, Angela Merkel, Vladimir Putin, John Roberts] now hold? (Type your answer. If you don't know, please provide your best guess)." Finally, participants rated: How much do you approve or disapprove of the way Donald Trump is handling his job as President? From 1(Strongly disapprove) to 7(Strongly approve). On average, participants reported they "slightly disapproved" of the way Donald Trump was handling his job as President *(M* = -1.20, *SD* = 2.09).

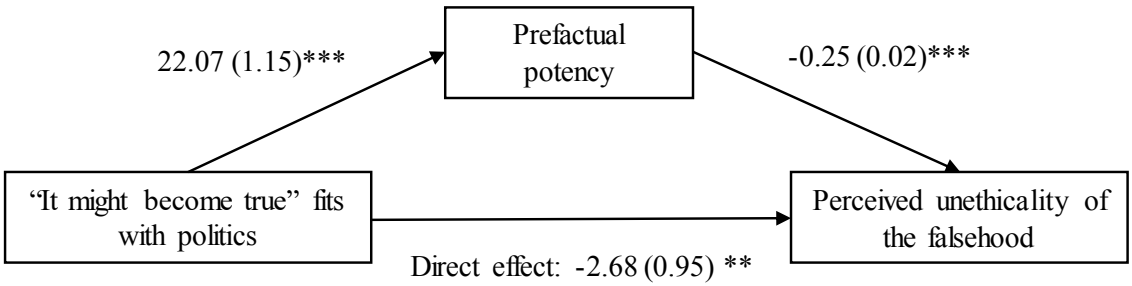
### Ease of Imagining the Prefactual: Prefactual Potency

We theorized that it is easier to imagine a falsehood might become true when that possibility fits with one's motivations and beliefs, and that the easier it is to imagine becoming true, the less unethical the falsehood will seem. In line with our theorizing, there was a significant indirect effect from fit with participants' politics, to prefactual potency, to

moral judgments of the falsehood, $b$ = -8.05, 95% CI [-∞, -4.73] (see main manuscript). This analysis is limited to participants in the prefactual condition, because they were the only ones who saw and rated prefactuals about how the falsehood might become true. We illustrate this indirect effect in Figure SOM-3.

**Figure SOM-3**

*Mediation Analysis of Political Fit on Judgments of Unethicality through Prefactual Potency in Study 4*




*Note.* Falsehood's fit with participants' political beliefs coded 1 = fit, 0 = conflict. Unstandardized coefficients shown.

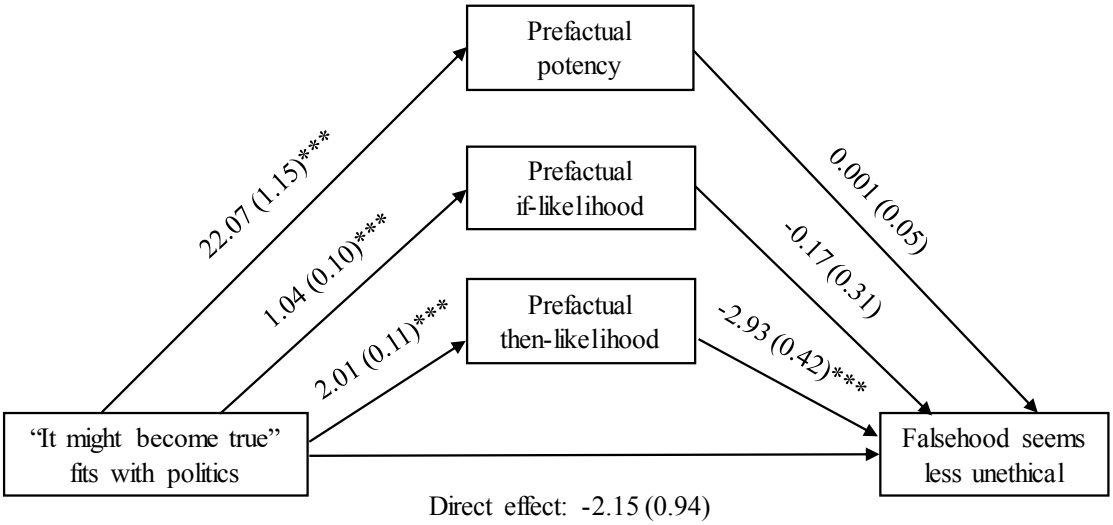
† p < .10, *** p < .001.

Exploratory analyses in Study 4 replicate the findings from Study 3 in suggesting that the mediating effect of prefactual potency was driven by participants' judgments of the then-likelihood. There was a significant indirect effect of fit with participants' politics on moral judgments through the then-likelihood, $b$ = -5.90, 95% CI [-7.66, -4.13], but not if-likelihood, $b$ = -0.17, 95% CI [-0.81, 0.46], or potency, $b$ = 0.01, 95% CI [-2.08, 2.11] (Figure SOM-4). We examined these indirect effects as multilevel mediation model with Stata's *gsem*

command, limited to participants in the prefactual condition because they were the only ones who saw and rated prefactuals about how the falsehood might become true.

**Figure SOM-4**

*Mediation Analysis of Political Alignment on Judgments of Unethicality through If-Likelihood, Then-Likelihood, and Potency in Study 4*




*Note.* Falsehood's fit with participants' political beliefs coded 1 = fit, 0 = conflict. Unstandardized coefficients shown.

† p < .10, *** p < .001.

**Intentions to Promote the Falsehood**

We examined whether relevant prefactuals increased participants' intentions to promote the falsehood online (i.e., increased their intentions to like and share it and decreased their intentions to post a negative comment about or block/unfollow the author of the falsehood). There was a significant indirect effect prefactual manipulation on all four promoting behaviors, through making the falsehood seem less unethical to tell. Moreover,

these indirect effects were significantly stronger when the falsehood-prefactual pair fit, rather than conflicted, with participants' politics (see main manuscript).

We also examined the total effect of the prefactual manipulation on intentions to promote the falsehood by submitting each of the four intentions to share measures to a mixed regression model with condition (1 = relevant prefactual, 0 = irrelevant prefactual) as a fixed effect, item as a fixed effect, and participant as a random effect. We found a total effect of the prefactual manipulation on reducing participants' intentions to block/unfollow the person who posted the falsehood, $b = -0.17$, $z = -1.80$, $p = .036$ (pre-registered one-tailed test), but not on any of the promoting behaviors (see Table SOM-2). There was also a significant interaction of prefactual thinking and political fit on participants' intentions to share the falsehood on social media, $b = 0.11$, $z = 1.74$, $p = .042$, but not the other three promoting intentions (see Table SOM-1).

**Table SOM-1**

*Regression Analyses Predicting Intentions to Promote the Falsehood in Study 4.*

| Predictor | *b* | *SE(b)* | *z* | *p* | 95% CI of *b* (two-tailed) |
|---|---|---|---|---|---|
| **Dependent variable: Intentions to Like the falsehood** | | | | | |
| Step 1 | | | | | |
| Condition | 0.04 | 0.09 | 0.44 | .308 | [-0.13, 0.21] |
| Constant | 2.21 | 0.07 | 31.90 | <.001 | [2.08, 2.35] |
| Step 2 | | | | | |
| Condition | -0.02 | 0.10 | -0.22 | .829 | [-0.21, 0.17] |
| Political fit | 0.32 | 0.05 | 6.78 | <.001 | [0.23, 0.41] |
| Condition x political fit | 0.08 | 0.06 | 1.26 | .110 | [-0.05, 0.21] |
| (Constant) | 2.01 | 0.08 | 25.26 | <.001 | [1.85, 2.17] |
| **Dependent variable: Intentions to Share the Falsehood** | | | | | |
| Step 1 | | | | | |
| Condition | 0.07 | 0.09 | 0.78 | .218 | [-0.11, 0.25] |

| | | | | | |
|---|---|---|---|---|---|
| Constant | 2.16 | 0.07 | 30.40 | <.001 | [2.02, 2.30] |
| Step 2 | | | | | |
| Condition | -0.001 | 0.10 | -0.01 | .992 | [-0.20, 0.20] |
| Political fit | 0.24 | 0.05 | 5.22 | <.001 | [0.15, 0.33] |
| Condition x political fit | 0.12 | 0.06 | 1.86 | .032 | [-0.01, 0.24] |
| (Constant) | 2.02 | 0.08 | 24.82 | <.001 | [1.86, 2.18] |
| **Dependent variable:** | | | | | |
| **Intentions to Leave a Negative Comment on the Falsehood** | | | | | |
| Step 1 | | | | | |
| Condition | 0.01 | 0.10 | 0.10 | .540 | [-0.19, 0.21] |
| Constant | 2.13 | 0.08 | 26.70 | <.001 | [1.97, 2.28] |
| Step 2 | | | | | |
| Condition | -0.07 | 0.11 | -0.61 | .543 | [-0.29, 0.15] |
| Political fit | -0.27 | 0.05 | -5.13 | <.001 | [-0.37, -0.16] |
| Condition x political fit | 0.04 | 0.07 | 0.54 | .707 | [-0.10, 0.18] |
| (Constant) | 2.40 | 0.09 | 26.38 | <.001 | [2.22, 2.58] |
| **Dependent variable:** | | | | | |
| **Intentions to Block / Unfollow the Person Who Posted the Falsehood** | | | | | |
| Step 1 | | | | | |
| Condition | -0.17 | 0.10 | -1.80 | .036 | [-0.36, 0.02] |
| Constant | 1.89 | 0.08 | 24.76 | <.001 | [1.74, 2.04] |
| Step 2 | | | | | |
| Condition | -0.23 | 0.11 | -2.13 | .033 | [-0.44, -0.02] |
| Political fit | -0.33 | 0.05 | -6.68 | <.001 | [-0.42, -0.23] |
| Condition x political fit | -0.03 | 0.07 | -0.49 | .313 | [-0.16, 0.10] |
| (Constant) | 2.19 | 0.09 | 25.38 | <.001 | [2.02, 2.36] |

*Note.* Condition is coded 1 = relevant prefactual, 0 = irrelevant prefactual. Political fit is coded 1 = fits, 0 = conflicts, with a participant's political beliefs. The mixed regression model also includes participant random effects and item fixed effects. We performed one-tailed significance tests to examine the effect of prefactual thinking and is interaction with political fit (pre-registered).

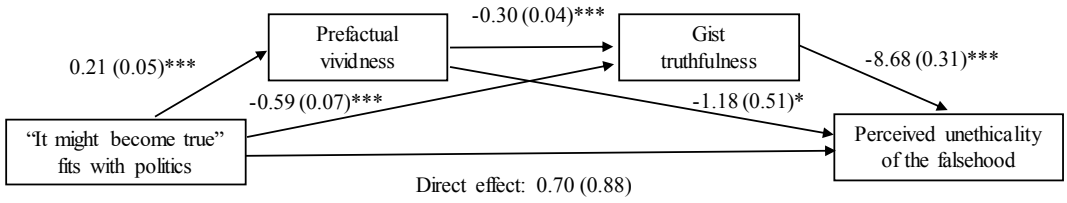
## Study 5

### Ease of Imagining the Prefactual: Prefactual Vividness

We theorized that it is easier to imagine a falsehood might become true when that possibility fits with one's motivations and beliefs, and that the easier it is to imagine becoming true, the more the gist of the falsehood would seem true, and the less unethical the falsehood will seem. In line with our theorizing, there was a significant serial mediation from fit with participants' politics, to prefactual potency, to gist, to moral judgments of the falsehood, $b$ = -0.54, 95% CI [-0.83, -0.26], (see main manuscript). This analysis is limited to participants in the prefactual condition, because they were the only ones who saw and rated prefactuals about how the falsehood might become true. We illustrate this indirect effect in Figure SOM-5.

**Figure SOM-5**

*Mediation Analysis of Political Fit on Judgments of Unethicality through Prefactual Potency in Study 5*




*Note.* Falsehood's fit with participants' political beliefs coded 1 = fit, 0 = conflict. Unstandardized coefficients shown.

* p < .05, *** p < .001.

### Intentions to Promote the Falsehood

Imagining how a falsehood might become true made it seem less unethical to tell, which in turn predicted stronger intentions to share the falsehood online, $b = 0.16$, 95% CI [0.09, ∞], and this indirect effect was significantly stronger when the falsehood-prefactual pair fit with participants' political beliefs, $b = 0.04$, 95% CI [-0.005, ∞] for the index of moderated mediation (see main manuscript for details of analyses).

We also examined the manipulation's total effect on sharing intentions. Participants reported greater intentions to share the falsehood when they wrote about how it might become true (*prefactual condition*: $M = 2.25$, $SD = 1.59$) versus how an irrelevant claim might become true (*irrelevant prediction condition*: $M = 2.05$, $SD = 1.55$). This difference was marginally significant in a mixed regression model with condition (1 = prefactual, 0 = irrelevant prediction) as a fixed effect, item as a fixed effect, and participant as a random effect, $d = 0.13$, $z = 1.58$, $p = .057$.

Moreover, the total effect of the prefactual manipulation on sharing was significantly greater when the falsehood fit, rather than conflicted, with participants' politics, $b = -0.14$, $z = 2.15$, $p = .016$. For this analysis, we added a dummy code for political fit (1 = fit, 0 = conflicts, with participants' political beliefs), and its interaction with the prefactual manipulation, to the regression model. When the falsehood-prefactual pair fit with participants' political views, they reported greater intentions to share the falsehood on social media if they wrote about how it might become true in the future ($M = 2.49$, $SD = 1.75$) than if they wrote about an irrelevant prediction ($M = 2.20$, $SD = 1.65$), $d = 0.17$, $z = 2.06$, $p = .040$. When the falsehood-prefactual pair conflicted with their politics there was not a significant difference in intentions to share the falsehood between participants who wrote about how it might become true in the future ($M = 2.00$, $SD = 1.55$) and participants who wrote about an irrelevant prediction ($M = 1.89$, $SD = 1.55$), $d = 0.28$, $z = 0.97$, $p = .331$. These simple slope analyses are two-tailed, because we did not pre-register this analysis.

## Study 6

### Political Fit and Fact-check Responses

We used a fact-check measure to examine participants' memory and beliefs about the falsehoods. In line with our intention that the fact-check measure capture beliefs (as well as memory) about the falsehoods, participants failed the fact-check more often when the falsehood fit with their politics (74.11%) than when the falsehood conflicted with their politics (85.30%), $b = -0.77$, $z = -10.27$, $p < .001$ (we did not pre-register this analysis, so this is a two-tailed test). This analysis was conducted in a mixed logistic regression analysis with fixed effects for condition (1 = relevant prefactual, 0 = irrelevant prefactual), political fit (1 = fits, 0 = conflicts with participants' politics), and items, and random intercepts for participants.

### Pre-Registered Robustness Checks

As pre-registered robustness checks, we repeated our prior analyses (a) with only responses to the dependent measures that corresponded to correct fact-checks, and (b) when controlling for the fact-check measure. We report the results of the robustness checks below. The results remained in the same direction and with the same statistical significance with one exception that we report in the main text; the marginally significant effect of the prefactual manipulation on perceptions of unethicality when the falsehood-prefactual pair fit with participants' politics became statistically significant when including only responses that corresponded to correct fact-checks.

**Gist mechanism.** Results of the pre-registered robustness checks supported the unique mediating role of perceived gist truthfulness. The main effect of the prefactual condition on *gist truthfulness* remained significant and in the same direction when we repeated our prior analyses retaining only responses to the dependent measure that corresponded to correct fact-checks, $b = 0.14$, $z = 2.55$, $p = .006$, and when we retained all

data but statistically controlled for the fact-check measure, $b = 0.12$, $z = 2.58$, $p = .005$. The main effect of the prefactual condition on *verbatim truthfulness* remained non-significant when we repeated our main analyses retaining only responses to the dependent measure that corresponded to correct fact-checks, $b = -0.03$, $z = -0.35$, $p = .724$, and when we retained all data but statistically controlled for the fact-check measure, $b = -0.01$, $z = -0.12$, $p = .901$.

We performed the same robustness checks on the indirect effects of the prefactual manipulation on perceptions of unethicality through gist truthfulness. Attesting to the robustness of the findings, when we repeated our prior analyses retaining only responses to the dependent measure that corresponded to correct fact-checks, the indirect effect through gist truthfulness was statistically significant and in the same direction, $b = -1.05$, 95% CI [$-\infty$, -0.37] and the indirect effect through verbatim truthfulness was not $b = 0.06$, 95% CI [-0.27, 0.39]. Providing further evidence of robustness, results for the indirect effects were the same when we retained all data but statistically controlled for the fact-check measure (*gist truthfulness*: $b = -0.74$, 95% CI [$-\infty$, -0.11]; *verbatim truthfulness*: $b = 0.11$, 95% CI [-0.28, 0.50]).

As in the analyses retaining all data, the total effect of the prefactual manipulation on perceptions of unethicality was not significant when we retained only responses to the dependent measure that corresponded to correct fact-checks, $b = -1.76$, $z = -1.31$, $p = .096$, and when we retained all data but statistically controlled for the fact-check measure, $b = -1.55$, $z = -1.24$, $p = .108$.

**Moderation by political fit.** Results of the robustness check provide further evidence that the effect of the prefactual manipulation on perceptions of unethicality was significantly larger when the falsehood-prefactual pair fit, rather than conflicted, with participants' politics (H3). This interaction was significant when we retained only responses to the dependent measure that corresponded to correct fact-checks, $b = -3.72$, $z = -2.39$, $p = .009$, and when we

retained all data but statistically controlled for the fact-check measure, $b$ = -2.65, $z$ = -1.92, $p$ = .028.

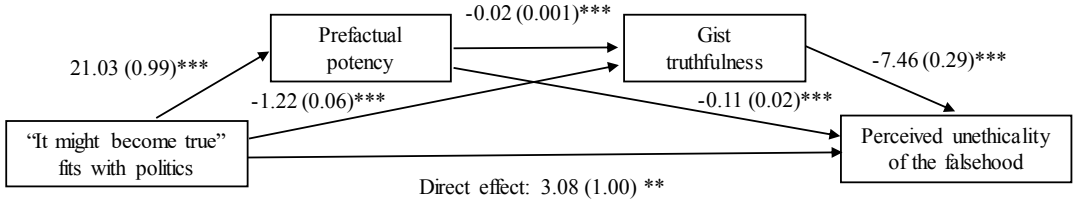
Robustness checks also support the simple slopes findings in the main text. When the falsehood-prefactual pair *fit* with participants' politics, considering a relevant prefactual led participants to rate the falsehood as less unethical to tell than considering an irrelevant prefactual: $b$ = -3.71, $z$ = -2.36, $p$ = .018, when retaining only responses to the dependent measure that corresponded to correct fact-checks, and $b$ = -2.80, $z$ = -1.97, $p$ = 0.049, when retaining all data but statistically controlled for the fact-check measure. When the falsehood-prefactual pair *conflicted* with participants' politics, considering a relevant prefactual did not lead participants to rate the falsehood as less unethical to tell: $b$ = 0.01, $z$ = 0.01, $p$ = 0.995, when retaining only responses to the dependent measure that corresponded to correct fact-checks, and $b$ = -0.15, $z$ = -0.11, $p$ = .915, when retaining all data but statistically controlled for the fact-check measure. These tests are two-tailed, because we did not pre-register hypotheses for these simple slopes.

**Ease of Imagining the Prefactual: Prefactual Potency**

We theorized that it is easier to imagine a falsehood might become true when that possibility fits with one's motivations and beliefs, and that the easier it is to imagine becoming true, the more truthful the gist will seem, and the less unethical people will judge the falsehood. In line with our theorizing, there was a significant serial mediation from fit with participants' politics, to prefactual potency, to gist, to moral judgments of the falsehood, $b$ = -3.74, 95% CI [-4.31, -3.17] (see main manuscript). This analysis is limited to participants in the prefactual condition, because they were the only ones who saw and rated prefactuals about how the falsehood might become true. We illustrate this indirect effect in Figure SOM-6.

**Figure SOM-6.**

*Mediation Analysis of Political Fit on Judgments of Unethicality through Prefactual Potency and Gist in Study 6*




*Note.* Falsehood's fit with participants' political beliefs coded 1 = fit, 0 = conflict. Unstandardized coefficients shown.

** p < .01, *** p < .001

**'Liking' Measure**

In contrast to our pre-registered predictions, there was not a significant indirect effect of prefactual thinking on liking the falsehood, $b = 0.04$, one-tailed 95% CI [-0.03, ∞]. We computed this indirect effect as a multi-level indirect effects model using the gsem and nlcom functions in Stata, with Prefactual condition as the predictor (1 = relevant prefactual, 0 = irrelevant prefactual), unethicality as the mediator, and liking as the dependent variable. We computed the indirect effect by multiplying the *a*- and *b*-paths together (i.e., using the *gsem* and *nlcom* functions in Stata). We specified a logit function on the b-path because our behavioral measure of liking was a dichotomous measure

As an exploratory analysis, we examined a serial mediation model through ratings of gist and unethicality. In line with our theorizing, there was an indirect effect from prefactual condition to ratings of gist, to ratings of unethicality, to liking in a serial mediation model, but this serial mediation was only marginally significant, $b = -0.02$, 95% CI [-0.0005, 0.04],

computed with item fixed effects and participant random effects, and an independent covariance structure (Stata's default). We did not pre-register this analysis so we consider these results preliminary.

**Gist Mediation without Verbatim Mediation**

In the main manuscript, we report the significant indirect effect of prefactual condition on unethicality judgments through perceived gist truthfulness and verbatim truthfulness. To examine the robustness of the mediation through gist truthfulness, we also examined the indirect effect of prefactual condition on unethicality through gist truthfulness without computing a path through verbatim truthfulness. Supporting the robustness of the indirect effect through gist, this indirect effect remains significant when removing parallel mediation through verbatim truthfulness from the model, $b$ = -0.84, 95% CI [-∞, -0.13]. We conducted this analysis as a generalized structural equation model with prefactual condition as the independent variable (1 = relevant prefactual, 0 = irrelevant prefactual), gist truthfulness as the mediator, and perceptions of unethicality as the dependent variable, with random effects for participants and fixed effects for item to account for the data's multilevel structure. We computed the indirect effects by multiplying the *a*- and *b*-paths together using the *gsem* and *nlcom* functions in Stata.

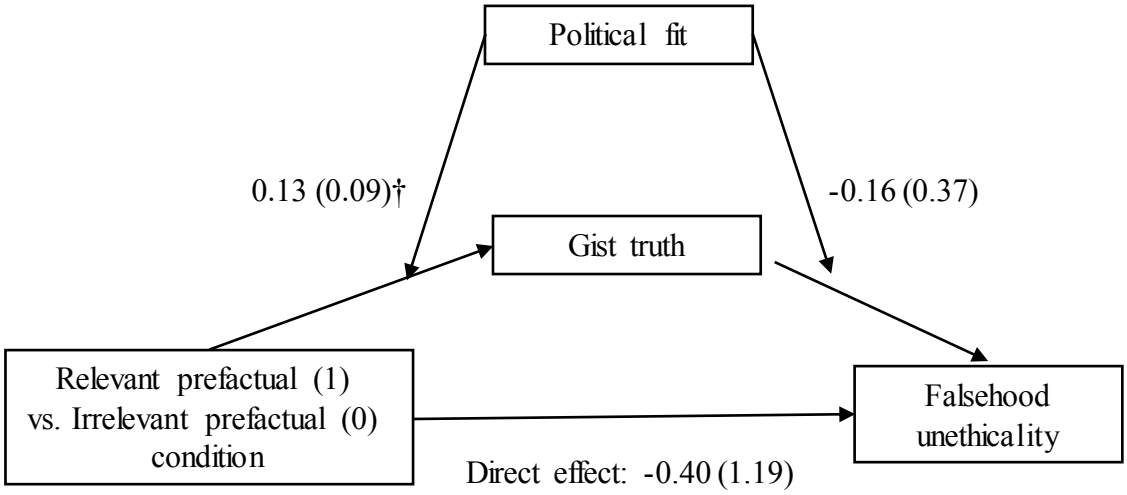
**Indirect of Prefactual Manipulation on Unethicality Through Gist, Conditional on Political Fit**

We examined the indirect effect of the prefactual manipulation on unethicality of the falsehood through perceptions of gist truthfulness, conditional on whether the falsehood-prefactual pair fit with participants' politics. When the possibility that the falsehood might become true *fit* with participants' politics, there was a significant indirect effect on perceptions of unethicality through gist truthfulness, $b$ = -1.38, 95% CI [-2.49, -0.27]. When

the possibility that the falsehood might become true *conflicted* with participants' politics, there was not a significant indirect effect of perceptions of unethicality through gist truthfulness, $b$ = -0.27, 95% CI [-1.34, 0.80]. However, this difference in the indirect effects was not statistically significant, $b$ = -1.11, 95% CI [-∞, 0.11] for the index of moderated mediation (Figure SOM-7). We conducted this analysis as a generalized structural equation model with prefactual condition as the independent variable (1 = relevant prefactual, 0 = irrelevant prefactual), gist truthfulness as the mediator, perceptions of unethicality as the dependent variable, with moderation by political fit on the *a*- and *b*-paths. The model also included random effects for participants and fixed effects for item to account for the data's multilevel structure. We computed the indirect effects by multiplying the *a*- and *b*-paths together using the *gsem* and *nlcom* functions in Stata.

**Figure SOM-7**

*Conditional Indirect Effect of Prefactual Condition on Judgments of Unethicality through Gist in Study 6*




*Note.* Falsehood's fit with participants' political beliefs coded 1 = fit, 0 = conflict.

Unstandardized coefficients shown.

† $p < .10$.

## Supplemental Study 1: Intuitive versus Deliberative Thinking

In a supplementary study, we examine whether inducing participants to think carefully and deliberatively, rather than rapidly and intuitively, could mitigate the effect of prefactual thinking on excusing falsehoods. Results from this study replicate the effect of prefactual thinking on reducing condemnation on falsehoods, and find that this effect does not depends on whether participants are thinking deliberatively or intuitively.

### Method

Participants evaluated the unethicality of six falsehoods in a 2 (condition: prefactual versus control; between subjects) x 2 (mode of thinking: intuitive versus deliberative; between subjects) design. In previous studies, participants responded to falsehood-prefactual pairs that either fit or conflicted with their political beliefs. Results showed that prefactuals reduced condemnation of falsehoods both when the falsehood-prefactual pair fit and when it conflicted with participants' beliefs, but this effect was greater for the former than the latter. In this study, we limited analyses to falsehoods that fit with a participants' political beliefs, such that Democrat and Republican participants saw separate sets of falsehoods that fit with their political views.

**Participants.** We recruited a sample of 800 American participants on Prolific Academic following the same exclusion criteria as Studies 3-5. Our previous studies included a greater proportion of Democrats than Republicans, so in this study we targeted an equal number of Republicans and Democrats. After our pre-registered exclusion criteria (non-US or duplicate IP addresses, duplicate participant IDs), we had a sample of 856 participants who completed all dependent measures for at least one falsehood. This number is slightly greater than our targeted number because of how the Prolific Academic software handles sign-ups.

Participants were 388 men and 385 women ($M_{age}$ = 36, $SD_{age}$ = 13). Of these participants, 436 participants identified as Democrat and 420 identified as Republican.

**Procedure.** Participants were randomly assigned to one of four between-subjects conditions in a 2(prefactual versus control) x 2(intuitive versus deliberative) design. At the beginning of the study, participants in the intuitive thinking condition were instructed to "make each rating based on your first instinct. Pay attention to your feelings, and don't think too hard," whereas participants in the deliberative condition were instructed to "take time to deliberate. Think very hard about the statements, try to ignore any gut feelings, and generate clear reasons for your judgment" (adapted from Nordgren & Dijksterhuis, 2009).

Then, all participants viewed a fact—labeled as "a proven fact"—about a controversial political issue (e.g., "It's a proven fact that the average top CEO makes much less than 500 times more money than the average American worker."). After viewing the fact, participants in the prefactual condition considered the prediction, "If politicians don't do anything about it, then…" followed by a statement about how a falsehood that contradicted the fact might become true in the future (e.g., "If politicians don't do anything about it, then top CEOs will make 500 times more money than the average American worker."), and rated the likelihood of that prefactual from 1 (*Not at all likely*) to 11 (*Extremely likely*). Participants in the control condition did not read or rate a prefactual. Next, participants in the intuitive condition completed a rapid word association task to prime intuitive thinking, whereas participants in the deliberative condition provided one reason why they believed making the given statement was ethical or unethical. Participants then rated the falsehood's unethicality and their intentions to share the falsehood online.

Participants repeated this procedure for six falsehoods that fit with their politics. Finally, participants responded to a fact-check measure (see below), the three-item cognitive reflection task (Frederick, 2005), and reported their demographics.

## Measures

**Unethicality.** Participants responded to the same 3-item measure of unethicality from Study 1 ($a$ = .88 for each falsehood).

**Intentions to promote the falsehood on Reddit.** Participants also reported their intentions to promote the falsehood online by "upvoting" the statement on the website Reddit: "If you saw an article with this headline on reddit, how likely would you be to "upvote" it (i.e. increase the chance that more people would see it)?" (from 1 = *Not at all likely* to 5 = *Extremely likely*).

**Fact-check.** At the end of the study, we included a fact-check measure to test that participants acknowledged that the falsehoods were indeed false. Specifically, for each falsehood, we showed participants either the fact or the falsehood they had seen earlier, and participants had to categorize it as true or false. We expected that our main results would be robust after excluding responses to items on which participants did not correctly identify fact from falsehood.

## Results

**Unethicality of telling the falsehood.** Across conditions, participants believed it was moderately unethical to tell the falsehoods ($M$ = 57.80 on a 100-point scale, $SD$ = 23.86). Supporting our main hypothesis and replicating our findings in the main text, participants who considered a prefactual about how the falsehood might become true judged the falsehoods as less unethical to tell ($M$ = 55.04, $SD$ = 22.96) than participants in the control condition ($M$ = 60.59, $SD$ = 24.45). This effect was significant in a mixed regression model with condition as a fixed effect (1 = prefactual, 0 = control), falsehood fixed effects, and random intercepts for participants, $d$ = -0.23, $z$ = -3.71, $p$ < .001.

We did not find support for our prediction that deliberative thinking would attenuate the effect of prefactual thinking on judging falsehoods as less unethical to tell. We added to the regression model a dummy term for mode of thinking (1 = intuitive, 0 = deliberative) and an interaction between mode of thinking (1 = intuitive, 0 = deliberative) and prefactual thinking (1 = prefactual, 0 = control). A non-significant coefficient on the interaction term indicated that the effect of the prefactual-thinking manipulation did not differ between the intuitive- and deliberative-thinking conditions, $b = -.26$, $z = -0.08$, $p = .936$. However, there was a total effect of mode of thinking, such that participants in the intuitive conditions judged the falsehoods to be less unethical to tell ($M = 56.07$, $SD = 23.01$) than participants in the deliberative condition ($M = 59.64$, $SD = 24.63$), $d = -0.15$, $z = -2.38$, $p = .017$.

***Promoting the falsehood online.*** Not only did the prefactual manipulation reduce condemnation of falsehoods, but it also increased participants' intentions to promote the falsehood online. Participants in the prefactual condition reported significantly stronger intentions to "upvote" the falsehood on the popular discussion forum site Reddit ($M = 2.18$, $SD = 1.09$) than participants in the control condition ($M = 2.04$, $SD = 1.07$), $d = 0.13$, $z = 2.05$, $p = .040$.

Perceptions of unethicality mediated the effect of the prefactual condition on participants' intentions to promote the falsehood on Reddit. There was a significant indirect effect from prefactual condition to judgments of unethicality to intentions to promote the falsehood, $b = 0.16$, one-tailed 95% CI [0.12, ∞]. We conducted this mediation analysis as a generalized structural equation model with prefactual condition as the independent variable (1 = prefactual, 0 = control), unethicality as the mediator, and intentions to promote the falsehood as the dependent variable. The model included random effects for participants and fixed effects for item to account for the data's multilevel structure. We computed the indirect

effects by multiplying the *a*- and *b*-paths together using the *gsem* and *nlcom* functions in Stata.

***Fact-check.*** The prefactual manipulation did not influence participants' beliefs of the truth of the statement. Participants in the prefactual condition were as likely to correctly identify the statement's truth (72.94%) as participants in the control condition (73.69%), $b = -0.13$, $z = -0.68$, $p = .495$. Further, the significant effect of prefactual thinking on judgments of ethicality and likelihood to promote the falsehood remained significant in the same direction when controlling for memory, $b = -6.26$, $z = -4.14$, $p < .001$, and when restricting our analyses to those statements that participants correctly identified as true or false, $b = -4.02$, $z = -2.52$, $p = .006$. These findings suggest that prefactuals did not reduce condemnation of falsehoods by confusing participants about whether the statement was true. Rather, prefactuals made participants more lenient of falsehoods they knew to be untrue.

**Discussion**

Supplemental Study 1 replicated our basic effect: People judged falsehoods as less unethical when they imagined how the falsehoods might become true. Moreover, imagining how a falsehood might become true increased participants' intentions to upvote falsehoods on Reddit, and this effect was mediated by judgments of unethicality.

Results did not support our prediction that encouraging people to think deliberatively would mitigate this effect. Imagining how a falsehood might become true led participants to judge the falsehood as less unethical to tell regardless of whether participants thought deliberatively or intuitively. It is possible that even when thinking deliberatively, considering how a falsehood might become true leads people to judge the gist of that falsehood as truer, and that people view this as reason to judge a falsehood as less unethical to tell. Altogether, the results of this supplementary study provide evidence for the robustness of the effect of

prefactual thinking on reduced condemnation of falsehoods and suggest that future research is needed on how to mitigate the effect of prefactual thinking on judging falsehoods' morality.