

## Talking About What Would Happen Versus What Happened: Tracking Congressional Speeches During COVID-19

Rinseo Park<sup>1</sup>, Young Min Baek<sup>\*2</sup>

<sup>a</sup>Department of Statistics and Data Science, Yonsei University, Seoul, South Korea

<sup>b</sup>Department of Communication, Yonsei University, Seoul, South Korea

\*Department of Communication, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul, South Korea, 03722. ymbaek@yonsei.ac.kr

In counterfactual thinking, an imagined alternative to the reality that comprises an antecedent and a consequent is widely adopted in political discourse to justify past behaviors (i.e., counterfactual explanation) or to depict a better future (i.e., prefactual). However, they have not been properly addressed in political communication literature. Our study examines how politicians used counterfactual expressions for explanation of the past or preparation of the future during COVID-19, one of the most severe public health crises. All Congressional speeches of the Senate and House in the 116th Congress (2019-2020) were retrieved, and counterfactual expressions were identified along with time-focusing in each speech, using recent advances in natural language processing (NLP) techniques. The results show that counterfactuals were more practiced among Democrats in the Senate and Republicans in the House. With the spread of the pandemic, the use of counterfactuals decreased, maintaining a partisan gap in the House. However, it was nearly stable, with no party differences in the Senate. Implications of our findings are discussed, regarding party polarization, institutional constraints, and the quality of Congressional deliberation. Limitations and suggestions for future research are also provided.

counterfactuals, time-focusing, Congressional deliberation, COVID-19, natural language processing (NLP)

### Non-Technical Summary

#### 1. Background

Since the outbreak of COVID-19 in March, compared to other countries, the United States (US) had time, even if insufficient, to prepare for the approaching threat of COVID-19. However, there were massive administrative failures in the early stages of COVID-19 spread, and preventative public health policies were not appropriately executed, resulting in many sufferings and deaths that could be saved. Regarding political communication, it is important to examine how US politicians (1) suggest taking preemptive actions at the early stage and (2) ask accountability for the inaction of preventative policies. Against this backdrop, this study focuses on US politicians' use of counterfactual expressions.

#### 2. Why was this study done?

Counterfactual expressions, such as the “as if ...” or “what if ...” statements, are popular and used practically in political as well as ordinary discourse. Despite the widespread practice of counterfactuals, however, they have not been properly addressed in political communication literature. In terms of persuasion, people use counterfactuals (1) to justify past behaviors (i.e., counterfactual explanation) or (2) to depict a better future (i.e., prefactual). Under which circumstances do US politicians use different type of counterfactuals in Congressional speech? Our research goal is (a) to detect counterfactual expressions utilizing recent advances in natural language processing (NLP) techniques and (b) to understand how US politicians practice counterfactual expressions during COVID-19. This study is the first empirical research tracking the use of counterfactuals in Congressional speeches.

### 3. What did the researchers do and find?

We examined US politicians' use of counterfactual expressions in Congressional speeches during COVID-19. The results show that counterfactuals were more practiced among Democrats in the Senate and Republicans in the House. As the pandemic worsened, the use of counterfactuals decreased, maintaining a partisan gap in the House. In the Senate, however, it was nearly stable, without party differences. Additionally, with time, politicians put more emphasis on the past (i.e., counterfactual expression) and less highlighted the future (i.e., prefactual). Those time trends in counterfactuals and time-focusing show that politicians are more likely to talk about "what happened" than "what would happen" in the later stages of the pandemic.

### 4. What do these findings mean?

We discuss the findings from theoretical perspectives of Congressional deliberation, party polarization, and bicameral representation. This study found that compared to the Senate, House members showed decreasing time trends with party differences when using counterfactuals. Those patterns suggest that House members' use of counterfactuals is under the influence of (1) institutional constraints (e.g., shorter terms and lower possibilities of re-election) and (2) political polarization. While time trends of counterfactuals and future-focusing were decreasing in both partisans, Republicans' use of counterfactuals was always higher but their future-focusing was lower than that of Democrats. The fact that Republicans put less emphasis on the future hints at their justification motives for the administrative failures. Our findings reflect the bipartisan, bicameral political system in the US. Faced with the unparalleled public health risk, partisans in the Senate and the House demonstrated diverging patterns of counterfactuals use. However, their differences are not unique to the pandemic situation and are attributable to the systematic differences between the two parties and chambers in the US.

On March 11, 2020, the World Health Organization (WHO) declared the novel coronavirus (COVID-19) as a pandemic (WHO, 2020). Among the developed countries of representative democracy, the United States (US) had time, even if insufficient, to prepare for the approaching threat of COVID-19. However, there were "massive failures of judgment and inaction" (Wallach & Myers, 2020) in the early stages of COVID-19 spread, resulting in many sufferings and deaths that could be saved (Worldometer, 2021). Regarding political communication, it is important to examine how US politicians (1) suggest taking preemptive actions at the early stage and (2) ask accountability for the inaction of preventative policies.

This research focuses on politicians' use of counterfactual expressions—defined as linguistic expressions asking the audience "an imagined alternative to the reality about the past" or "the future" (Byrne, 2016, pp. 137-138; also see Epstude & Roese et al., 2008; Tetlock & Belkin, 1996)—in the US Congressional speech. Although alternative fact has been criticized as a pejorative term equated with "fake news" or "misinformation" in recent political discourse (e.g., Barrera et al., 2020; Kakutani, 2018), the use of counterfactuals has been continuously practiced in political as well as ordinary discourse. For example, ancient Greek politicians, according to Thucydides' (ca. 400 B.C.E./1974) portrayal in *History of the Peloponnesian War*, frequently exercised counterfactual expressions to persuade citizens to declare wars against the rivals. Similarly, today's politicians also adopt counterfactuals while justifying, defending, or excusing the negative consequences caused by their actions in the past (Effron, 2018; McGraw, 1991).

The paper proceeds as follows. In the next section, we define counterfactuals and discuss their political meanings in politicians' rhetoric. After examining the counterfactuals research in social sciences and computational linguistics, their use by politicians is analyzed according to party identity, chamber type, and the stage of the COVID-19 spread. After reporting our

findings in the Results section, implications for Congress members' use of counterfactuals are discussed.

## The Role and Functions of Counterfactuals

Under democracy, politicians are asked to make collective decisions that might influence the country's future (Aristotle, ca. 400 B.C.E./1991; Lee, 2011b; Tetlock & Belkin, 1996), and to take accountability for the consequences of their decisions (Weber, 1919). In the collective decision-making process, politicians exercise a variety of strategies to persuade or dissuade others—the colleagues of the rivals as well as the same party, and the general public (Lee, 2011b; Rippere, 2016). Even after the decision is made, politicians are obliged to defend their decisions, especially when negative outcomes are associated (McGraw, 1991). Throughout the process, counterfactual expressions are commonly adopted as communication strategies persuading others and/or justifying the legitimacy of their former positions (Byrne, 2016; Effron, 2018; McGraw, 1991). Unfortunately, there is a lack of political communication studies examining the use of counterfactuals by politicians.

Counterfactuals are comprised of the antecedent as a cause and consequent as a result (Pearl & Mackenzie, 2018; Tetlock & Belkin, 1996). As counterfactuals build upon the imagined antecedent, at least since Hume (1748/1988), philosophers have been disputing the empirical content of counterfactuals (Dawid, 2000; Pearl, 2000). While the issue remains philosophically unresolved, counterfactuals, such as the “as if ...” or “what if ...” statements, are popular and used practically (Barrera et al., 2020; Pearl & Mackenzie, 2018). Moreover, medical findings (e.g., McNamara et al., 2003; Roese et al., 2008) imply that counterfactual thinking is inevitable for healthy human brains because individuals with injuries to prefrontal cortex or Parkinson's disease reveal the loss of counterfactual thinking.

In terms of persuasion, there are two main reasons for the use of counterfactuals (Byrne, 2016). On the one hand, people use counterfactuals for explanation of the past, called “counterfactual explanation” (Byrne, 2016, p. 136). A variety of motives may appear, such as justifying the past decision or the present consequent (e.g., “If X had not been put in place, the situation would have been worse.”), regretting (e.g., “If X were taken earlier, more lives could be saved.”) or acknowledging the past decision with the relief in the present (e.g., “Without X's assistance, my life would be miserable.”).

Although the motive of regret or relief has been more frequently observed among ordinary people (Byrne, 2016; Epstude & Roese, 2008; Pearl & Mackenzie, 2018), justification has been dominant in politicians' discourse, especially when their past decision is suspected to be responsible for the present misfortune (Barrera et al., 2020; Effron, 2018; McGraw, 1991). As COVID-19-caused damages become apparent (Wallach & Myers, 2020), it is a common behavior to blame elected officials (e.g., the President or politicians of the incumbent party) or appointed officials (e.g., Director for Centers for Disease Control and Prevention) for their inaction.

On the other hand, people use counterfactuals for preparation for the future, termed “prefactual” or “preparatory counterfactual” (Byrne, 2016, p. 138). While counterfactual explanations empathize with the mutable antecedent (i.e., the alternative to reality about the past), prefactuals highlight the mutable consequent (i.e., the alternative to reality about the future) and this should be distinguished from the past-focused “counterfactual” (Epstude et al., 2016). Prefactuals are used as suggestions for a better future, frequently in public health campaigns (e.g., “If you quit smoking, your life expectancy will be longer”), commercial advertisement (e.g., “If X is implemented, your company's efficiency will be promoted”), or public deliberation (e.g., “Without X, our workers will lose their job”).

As Aristotle (ca. 400 B.C.E./1991) clarified over two thousand years ago, the key function of “democratic assembly” is “judging about future happenings” (p. 48). From the emergence of

COVID-19 in the US until its nationwide spread as a clear and imminent risk, we expect politicians to deploy prefactuals to ask about preemptive disinfection policies and prepare medical supplies before the situation worsens. By contrast, their use of counterfactuals will differ at later stages, when people may have become inured to the death toll and prolonged restrictions. Moreover, even before the initial stages, when scant attention was paid to the virus, divergent time trends were also conceivable. It was only after March that several political figures, including President Trump, and the public began to perceive the looming pandemic as a real risk. Thus, we suspect nonlinear time effects on politicians' use of counterfactuals during the COVID-19 pandemic.

When it comes to reference points in counterfactuals (i.e., “future” for prefactual and “past” for counterfactual explanation), existing studies also suggest that individual differences need to be considered since temporal orientations of language use convey personal attitudes toward the present situation. Studies conducted in linguistic psychology have found that individuals' time-focusing displayed in their language use reveals important clues for individual and psychological features, such as gender, age, personality, or value priorities (Pan et al., 2018; Pennebaker & King, 1999).

More explicitly, previous studies have identified contextual factors of counterfactuals, such as self-serving motives and social power (Epstude et al., 2016; Scholl & Sassenberg, 2015). Individuals are less likely to engage in prefactual thoughts if they have more power or responsibilities over given tasks; but more likely to use counterfactuals after failure (i.e., counterfactual explanation). Especially in politics, counterfactuals are often deployed as communication strategies by increasing group identification (Milesi & Catellani, 2011) and effectively attacking the target (i.e., counterfactual attack; see Bertolotti & Catellani, 2018). Therefore, individual and contextual differences need to be considered when examining politicians' use of counterfactuals.

In summary, people use counterfactuals for (1) explanation of the “past” (i.e., counterfactual explanation) or (2) preparation for the “future” (i.e., prefactual). To understand how politicians practice counterfactual expressions with different time-focusing, we examined Congressional speech data during COVID-19.

## **Factors Influencing Counterfactual Expressions: Party, Chamber Type, and Time**

Politicians' rhetorical strategies hinge on a variety of factors, which could be individual (e.g., gender, age, or endorsed ideology), institutional (e.g., ascribed role or discretionary power), or time-contingent (e.g., timing or moment). Regarding their COVID-19 deliberation, our study focuses on three crucial factors influencing the use of counterfactuals and time-focusing: (1) party identity (i.e., Republicans or Democrats), (2) chamber type (i.e., Senate or House), and (3) the stage of the COVID-19 spread (i.e., time).

First, politicians' party identity would influence their use of counterfactuals. In the US political culture, research has shown that Republicans and Conservatives are more likely to use words denoting loyalty or authority and favor past-focusing expressions, while Democrats and Liberals are more likely to use words showing benevolence or equality and favor future-focusing expressions (Brady et al., 2019; Robinson et al., 2015). Although not directly examining counterfactuals, previous findings imply that Republicans' language use, compared to that of Democrats, would focus on the reality rather than its alternatives, because of their preference for conservative and past-related values, such as tradition or conformity (Caprara et al., 2006; Jones et al., 2018). Thus, it can be inferred that Republicans are less likely to adopt counterfactuals since they are an “imagined alternative to reality.” In contrast, Democrats, who cherish liberal values and desire to accomplish the mutable future, are more likely to use counterfactuals.

Considering that misleading statements based on irrational projections or conspiracy beliefs are defended as “alternative facts” among Republicans supporting Trump Administration (e.g., Effron, 2018; Kakutani, 2018), however, it might be too hasty to predict that Republicans intend to reserve counterfactuals because they seek certainty by focusing on the immutable past rather than the mutable future. Focusing on partisans’ psychological and personal differences, Democrats are more likely to describe the better future through counterfactuals (Robinson et al., 2015). However, it is also not implausible to expect the opposite relationship, as Democrats are more likely to emphasize the current state of American politics (Effron, 2018). Given the lack of previous research, we propose the following research question, rather than a theory-driven research hypothesis:

**RQ1:** When deliberating the issue of COVID-19 in the US Congress, how is politicians’ use of counterfactuals related to their party identity?

Second, politicians’ chamber type affects their use of counterfactuals and time orientation in speeches. In bicameral representation, the US Constitution aimed to make two chambers “as little connected” (Lee, 2011a, pp. 272-273), and Congress members’ institutional status and discretionary power vary. For example, US Senators are more experienced than House representatives with longer terms, and they represent states rather than small districts (Lee, 2011a). Thus, Senators are relatively free from popular short-term sentiments and work like semi-independent lawmakers (Lee, 2011a; Sinclair, 2000). Such differences between the Senate and House have been pointed out as the institutional factor shaping a different culture for each chamber. Partisan conflicts are easily observed in the House whose members are more likely to reflect affective polarization between Republican and Democratic citizens (Jamieson & Falk, 2000; Rippere, 2016). However, such conflicts have been less intense in the Senate because Senators consider each other as esteemed colleagues and engage more in content-oriented deliberation (Rippere, 2016) than blame games.

Moreover, numerous studies report that counterfactuals are frequently practiced by more educated and experienced people because they demand additional cognitive efforts and causal knowledge about the phenomena of creating plausible alternatives to the reality (Byrne, 2016, p. 137; Tetlock & Belkin, 1996). Historically, demographic profiles of members of Congress have shown that Senators, compared to House representatives, are older, with higher educational degrees, and serve for a longer time (Congressional Research Service, 2012, 2020). If such conjectures are correct, it is expected that Senators are more likely to use counterfactuals as preparative than for explanations of the past. By contrast, we expect that House representatives are less likely to use counterfactuals, considering their demographic profiles (e.g., younger age and lower levels of educational attainment) which are less conducive to cognitive demands of counterfactual thinking, as well as relatively insecure status (e.g., shorter terms and lower possibilities of re-election; see Congressional Research Service, 2012, 2020). Therefore, we shape the second research question, regarding chamber type:

**RQ2:** When deliberating the issue of COVID-19 in the US Congress, how is politicians’ use of counterfactuals related to their chamber type?

Third, time is the critical factor in predicting the role of counterfactuals in politicians’ speech. In the early stages of COVID-19, the use of counterfactuals would be more motivated to prepare the future when the infection risk is heightened (i.e., prefactuals); however, in later stages, they would be used in a retrospective sense (i.e., counterfactual explanations), wherein the meanings of counterfactuals are inferred using temporal orientations of speech. Furthermore, partisans are expected to deploy different strategies across the stages of the COVID-19 spread. The time effects on counterfactual expression patterns are likely to depend



on politicians' individual differences, namely, party identity or chamber type. Therefore, we shape a series of time-related research questions:

**RQ3:** When deliberating the issue of COVID-19 in the US Congress, how does politicians' use of counterfactuals vary across the stage of the COVID-19 spread?

**RQ3a:** How does politicians' use of counterfactuals vary across the stage of the COVID-19 spread, depending on their party identity?

**RQ3b:** How does politicians' use of counterfactuals vary across the stage of the COVID-19 spread, depending on their chamber type?

## Method

### Data Gathering Procedure

To construct the Congressional speech data during COVID-19, we systematically collected the speeches of the Congressional Record from Congress.gov, an official website. The collection was limited to the records of Senate and House section in the 116<sup>th</sup> Congress (2019-2020), containing either "COVID" or "coronavirus" as a search term. The raw text was then parsed into individual speeches. Each speech was matched with the identified speaker, serving as a member of the 116th Congress (Senate or House) and whose political party is either Republican or Democratic (speeches from three Independent members were dropped from the corpus; two Senators and one Congressperson).

Based on the procedure described, we identified 4,446 speeches between January to November 2020. For our analysis, we filtered out (1) the speeches uttered by Independent members ( $n = 16$ ), (2) the speeches made by two House members on the Senate floor ( $n = 2$ ), and (3) the speeches with less than or equal to one valid sentence ( $n = 241$ ). To exclude invalid sentences (e.g., abbreviations with a period, such as "H. Res. 925"), we applied the results of part-of-speech (POS) tagging and dependency parsing using the R "spacyr" package (version 1.2.1; Benoit & Matsuo, 2020). In our analysis, a sentence is treated as "valid" only if its root word is a verb, and only valid sentences were included. Further, because the speeches with one valid sentence are merely courtesy expressions or debate rules (e.g., "I reserve the balance of my time."), we required speeches to have more than just one valid sentence. Out of 241 removed speeches, seven had none of the valid sentences and 234 had only one valid sentence. As a result, an individual speech corpus, featuring 4,187 speeches with 191,296 valid sentences uttered by 500 unique members of Congress, was constructed. In the corpus, each of the 500 members gave a Congressional speech at least once.

## Measures

### *Dependent Variable*

#### *Counterfactual expression*

To identify counterfactual expressions in Congressional speech corpus, we combined token-based and POS-based rules with syntactic dependencies. There has been significant work on identifying discourse relations, mainly (1) building annotated corpora, such as the Penn Discourse Treebank (PDTB; Prasad et al., 2008) or Rhetorical Structure Theory Discourse Treebank (RST-DT; Carlson et al., 2001), and (2) automating the annotation procedure as a parser (Feng & Hirst, 2012; Lin et al., 2014). Unlike other all-purpose parsers such as the Lin parser (Lin et al., 2014), our approach focuses on identifying the predefined counterfactual forms. To reflect the unidirectional relationship between the antecedent and consequent in

counterfactual expressions, we represented each sentence as a CoNLL-style dependency tree.<sup>1</sup> Then, as a top-down approach, we identified the single root verb and its leaf verb(s)—which depends on the root—per sentence. The root and leaf verbs are candidates of the consequent and antecedent, respectively. Moreover, prior to that, each sentence is tokenized, lemmatized, and POS tagged. As a result, each word and punctuation (“token”) is identified with its basic form (“lemma”) and grammatical information (i.e., “POS tag”). As previously mentioned, we used the R “spacyr” package (version 1.2.1) for parsing and tagging. Finally, we extracted the counterfactual forms and types presented in Table 1.

Table 1. Three Forms of Counterfactual Expression

Form	Example
Conditional Conjunction	<i>If we <u>do not move</u> this bill, the death <u>will be</u> in our districts.</i> (Mr. GARAMENDI)
Verb Inversion	<i><u>Had</u> he (= Mr. HOYER) <u>been</u> serious about that, there <u>would have</u> <u>been</u> a discussion before this bill ever came to the floor.</i> (Mr. PALMER)
Wish/Should	<i>I <u>wish</u> this bill <u>was not</u> necessary, but unfortunately, it is now more imperative than ever.</i> (Mr. LOFGREN)

*Note.* The **root** verb, consequent, antecedent and *keyword* are highlighted.

In Table 1, we introduced three popular forms of counterfactual expression (Janocko et al., 2016; Son et al., 2017; Yang et al., 2020). Examples are drawn from our Congressional speech corpus (see Section A of the Supplementary Materials for details; our detection rules are illustrated through the examples in Table 1).

Regarding the “Conditional Conjunction” form, the consequent consists of a modal verb, and the antecedent consists of a base verb, a past-tense verb, or a modal verb, preceded by a conditional conjunction, such as “If” or “Unless.” For the “Verb Inversion” form, the sentence begins with a verb inversion (e.g., “Had he ...” or “Were I ...”), followed by a past-tense antecedent and a modal consequent. In the “Wish/Should” form, the clause in the past tense following “wish” or “should” is the antecedent and the consequent is omitted, yet implied as a better situation.

In total, we detected 1,843 (0.96%) counterfactual expressions out of 191,296 sentences. This is consistent with the previous finding that only 1% of tweets were counterfactual (Son et al., 2017). Among the three forms in Table 1, conditional conjunction occurred most frequently at 1,326 expressions. As a result, we calculated the total number of counterfactual expressions in each Congressional speech ( $M = .44$ ,  $SD = .96$ ).

### Time-Focusing

To identify time-focusing in Congressional speech, we utilized tense information of root verbs in a valid sentence. This is comparable to Linguistic Inquiry and Word Count (LIWC; Pennebaker et al., 2015) approach, but our measures are sentence-level using syntactic parsers (Feng & Hirst, 2012; Lin et al., 2014), rather than word-level using LIWC-styled dictionary. If the root verb in a sentence is a modal verb (e.g., will, would, etc.), the sentence is labeled as “future.” Similarly, we labeled the sentence “present” or “past” according to its root verb tense. Of the 191,296 sentences, 35,038 (18.3%) were past-oriented, 130,551 (68.2%) were present-oriented and 22,108 (11.6%) were future-oriented. Less than 2% of total sentences were unlabeled, most of which were perfect modal verbs and classified as counterfactuals.

In summary, we have three measures of time-focusing: (1) the number of past-focused sentences ( $M = 8.37$ ,  $SD = 15.31$ ), (2) the number of present-focused sentences ( $M = 31.18$ ,  $SD = 36.91$ ), and (3) the number of future-focused sentences ( $M = 5.28$ ,  $SD = 6.84$ ).

## Independent Variables

### Party identity

Congress members' party identity was considered to measure the partisan gap in language use. Our Congressional speech corpus examined 500 unique members of the 116th Congress, each of them affiliated with either the Republican or the Democratic party. Out of 500 members, 235 (47%) were Republicans and 265 (53%) were Democrats. This variable was effect-coded when entered in the regression models (Republican = 1, Democrat = -1).

### Type of chamber

To reflect the bicameral nature of the US Congress, chamber type was considered as another independent variable. Our data included 98 (19.6%) members of the Senate and 402 (80.4%) members of the House. For the Senate ( $n = 98$ ), 53 (54.1%) were Republicans and 45 (45.9%) were Democrats. For the House ( $n = 402$ ), 182 (45.3%) were Republicans and 220 (54.7%) were Democrats. This variable was also effect-coded for our regression models (Senate = 1, House = -1).

### Calendar time as the stage of COVID-19 spread

Our data range between January 23<sup>rd</sup> and November 10<sup>th</sup>, 2020. Since the pandemic spread in the US as well as other countries with a tremendous number of confirmed cases and deaths, it is reasonable to expect members from the two parties serving in two separate chambers to differently interpret and judge the situation at each time point. We treated calendar time as a measure of blame risk severity for two reasons. First, the total counts of confirmed cases and deaths due to COVID-19 in the US continue to rise during the period (The New York Times, 2020; see the “New Confirmed Cases” column in Table 2). Second, calendar time has strong positive correlation with the confirmed cases ( $r = .97, p < .001$ ) and deaths ( $r = .99, p < .001$ ). When estimating the time effect on Congress members' use of counterfactual expressions and time-focusing, we first transformed calendar dates into numbers, and then standardized them.

Further, its quadratic form was added as there may exist nonlinear time trends. The number of Congressional speeches addressing COVID-19 fluctuated greatly as the number of confirmed cases continued to resurge. Table 2 confirms that until February, COVID-19 was yet to become a major or urgent risk in the US, with both number of speakers and number of speeches below 100 ( $N_{Speaker} = 72, N_{Speech} = 82$ ). In March, however, the number of speakers soared to 300 and the number of speeches to over 600 ( $N_{Speaker} = 300, N_{Speech} = 608$ ), showing that the majority of US Congress members began to regard the pandemic with greater concern. The increase appeared to be non-negligible, even compared to the increase in the number of unique session dates ( $N_{Session} = 10$  in February;  $N_{Session} = 20$  in March). The ensuing time trends also resembled the pattern of the COVID-19 spread in the US, hitting its highest peak in July ( $N_{Speaker} = 370, N_{Speech} = 846$ ; also see the “New Confirmed Cases” column).

Table 2. Monthly Totals of New Confirmed Cases in the United States and COVID-19 Related Congressional Speech in the 116<sup>th</sup> Congress

Month	New Confirmed Cases	Congressional Speech		
		$N_{Session}$	$N_{Speaker}$	$N_{Speech}$
2020-01	7	2	2	2
2020-02	63	10	72	82
2020-03	188,391	20	300	608
2020-04	887,295	4	157	167
2020-05	723,546	15	243	496
2020-06	854,019	20	324	799
2020-07	1,918,348	12	370	846
2020-08	1,473,786	9	143	201
2020-09	1,217,279	13	294	695



2020-10	1,946,192	9	162	279
2020-11	1,428,492	2	9	12

*Note.* “ $N_{Session}$ ” is the number of unique session dates, “ $N_{Speaker}$ ” is the number of unique speakers, and “ $N_{Speech}$ ” is the number of Congressional speeches in each month.

## Statistical Methods

Our dependent measures are count variables to which conventional ordinary least square regression approach cannot be applied. Congress members from different states can make multiple speeches across different time points, which violates the independence assumption taken in general linear models (GLMs). Thus, special care should be taken when investigating our research questions. Specifically, multi-level zero-inflated Poisson (MZIP) model, along with multi-level Poisson (MP) model, with the total number of sentences as an offset variable were adopted for three reasons (for modeling details, see Section B of the Supplementary Materials). To estimate the MZIP and MP models, we used the R package “glmmTMB” (version 1.0.2.1; Brooks et al., 2017).

In the next section, we reported results of MZIP model for the number of counterfactual expressions; results of MP model for the number of (1) past, (2) present, and (3) future-focused sentences in a Congressional speech. In MZIP and MP models, we estimated the main effects of three independent variables (i.e., party, chamber, and calendar time) and their two-way interaction effects. Three-way interaction effect term was not included in the models as it fails to achieve conventional statistical significance.

## Results

### Descriptive Statistics for Counterfactual Expressions in Congressional Speech

Table 3 presents the average proportions of counterfactual sentences in speech. The numbers are tallied in each chamber and party of members, presenting the means of speech-wise ratios of “the number of counterfactual sentences” to “the total number of sentences” (see the “Counterfactual” column). Similarly, average proportions of time-focusing sentences are the means of speech-wise ratios of “the number of past-, present-, or future-focused sentences” to “the total number of sentences” (see the “Time-focusing” columns).

Two findings are notable in Table 3. First, Congress members have different strategic preferences regarding counterfactual expressions, depending on their party identity and type of chamber. In the Senate, Democrats (0.94%) used counterfactual expressions more frequently than Republicans (0.77%). In the House, however, Republicans (1.05%) used counterfactual expressions more frequently than Democrats (0.79%). Compared to the Senate (0.17%), the absolute partisan difference was larger in the House (0.26%). Second, members used past-focused sentences more frequently in the Senate (19%) than in the House (13%). Accordingly, Senators were less likely to use present- or future-focused sentences, compared to House representatives.

Table 3. Descriptive Statistics for Counterfactual Expression and Time-Focusing in Congressional Speech

Source	$N_{Speech}$	$N_{Sentence}$	Average Proportion of Sentences in Speech (%)			
			Counterfactual Use	Past-focusing	Present-focusing	Future-focusing
Senate Republicans	989	60,366	0.77%	18.4%	68.8%	10.4%

Democrats	762	58,380	0.94%	18.7%	67.7%	11.8%
Total	1,751	118,746	0.84%	18.6%	68.3%	11.0%
<b>House</b>						
Republicans	981	32,284	1.05%	14.4%	70.6%	12.0%
Democrats	1,455	40,266	0.79%	11.8%	72.3%	13.9%
Total	2,436	72,550	0.89%	12.8%	71.6%	13.1%

*Note.*  $N = 4,187$ . The number of Congressional speeches (“ $N_{Speech}$ ”) and their sentences totals (“ $N_{Sentence}$ ”) are reported. Average proportions of counterfactuals (time-focusing) in speech are the means of speech-wise ratios of “the number of counterfactual (past-, present-, or future-focused) sentences” to “the total number of sentences.” The sum of the proportions across the three types of time-focusing could be less than 100% as there is a fraction of sentences whose tense cannot be identified.

**Table 3** allows us a sketch of understanding patterns of counterfactual expression in Congressional speech. Still, major characteristics of our dependent measures (i.e., count variables with excessive zeros and multiple hierarchies) and time effects (i.e., calendar time) were not considered in this preliminary analysis. Therefore, we present the MP and MZIP regression results in the next section.

## Results Predicting Counterfactual Expression and Three Types of Time-Focusing

In **Table 4**, the MZIP regression result for the counterfactual expression (see the “Counterfactual” columns) and MP models for three types of time-focusing (see the “Time-focusing” columns) are displayed.

Table 4. Regression Models for Counterfactual Expression and Three Types of Time-Focusing

Parameter	Counterfactual		Time-focusing		
	Logistic	Poisson	Past-focusing	Present-focusing	Future-focusing
<b>Fixed effect</b>					
Intercept	-2.73*** (0.55)	-4.63*** (0.06)	-1.91*** (0.03)	-0.36*** (0.01)	-2.12*** (0.02)
Party (R = 1, D = -1) <sup>a</sup>	-0.90 (0.59)	0.08 <sup>†</sup> (0.04)	0.06* (0.03)	-0.002 (0.01)	-0.05*** (0.02)
Senate (S = 1, H = -1) <sup>b</sup>	-0.05 (0.42)	0.02 (0.04)	0.20*** (0.03)	-0.03*** (0.01)	-0.05*** (0.02)
Party × Senate	0.01 (0.37)	-0.15*** (0.04)	-0.02 (0.03)	< 0.001 (0.01)	-0.01 (0.02)
Time <sup>c</sup>	-1.04 <sup>†</sup> (0.54)	-0.08* (0.03)	0.07*** (0.01)	-0.01*** (0.003)	-0.04*** (0.01)
Time (squared)	-0.28 (0.42)	0.03 (0.03)	0.01* (0.01)	-0.01* (0.003)	-0.003 (0.01)
Party × Time	-1.10 <sup>†</sup> (0.60)	-0.02 (0.03)	0.01 (0.01)	-0.01 (0.003)	0.01 (0.01)
Senate × Time	0.03 (0.39)	0.08** (0.03)	0.03*** (0.01)	-0.02*** (0.003)	0.02** (0.01)
<b>Random effect</b>					

$\sigma^2$ (member)	N/A	0.32	0.43	0.07	0.23
$\sigma^2$ (state)	N/A	0.10	< 0.001	0.01	0.05
<b>Goodness of fit</b>					
AIC	5966.30		25078.77	24329.71	18796.72
-2 Log Likelihood	5930.30		25058.77	24309.71	18776.72
LR statistics					
$\chi^2(df)^d$	29.73*** (8)		N/A	N/A	N/A

*Note.*  $N = 4,187$ . Cell entries are regression coefficients of the multi-level zero-inflated Poisson regression model for “Counterfactual” and multi-level Poisson regression models for three types of “Time-focusing” with total number of sentences as an offset variable. Standard errors are reported in parentheses. Models were fit using the R “glmmTMB” package (version 1.0.2.1).

<sup>a</sup>R = Republican and D = Democrat. <sup>b</sup>S = Senate and H = House. Both variables were effect coded to interpret the coefficients. <sup>c</sup>Time denotes calendar dates in numbers, standardized for interpretation.

<sup>d</sup>Likelihood ratio test statistics against the multi-level Poisson regression model not estimating zero part are reported.

<sup>†</sup> $p < .10$ .  $*p < .05$ .  $**p < .01$ .  $***p < .001$ .

For counterfactual expression, the MZIP regression result can be interpreted as follows (see the “Counterfactual” columns). Overall, independent variables affect the occurrence of counterfactual expressions (see the “Poisson” column); however, they barely played the role in predicting whether to use counterfactual expression in Congressional speech (see the “Logistic” column). Those findings are reasonable, given that most of the sentences in Congress members’ speech were not counterfactual expressions. In other words, the MZIP regression succeeded to filter out members who never use counterfactual expressions (see the bottom row in [Table 4](#) confirming additional estimation predicting the zero occurrence of counterfactual expressions contributes to a significant increment of explanatory power:  $\chi^2(8) = 29.73$ ,  $p < .001$ ).

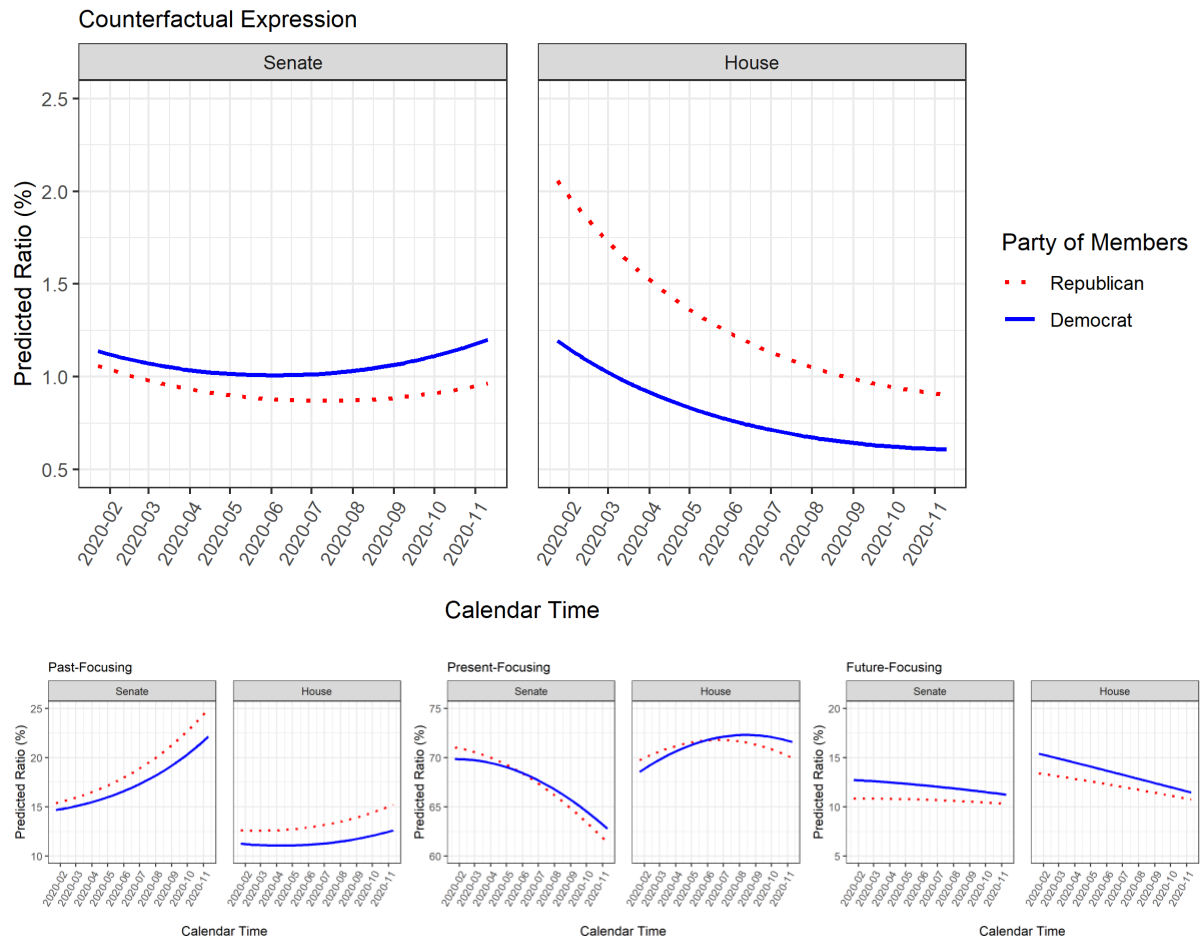
To facilitate understanding of the patterns in [Table 4](#), we adduced [Figure 1](#) showing how the occurrence of counterfactual expressions is related to Congress members’ party and chamber type. Because we have modeled the rate of expressing counterfactual expression in a speech by treating the total number of sentences as an offset variable in the MZIP model, this can be interpreted as the expected proportion of counterfactual sentences in the total number of sentences.

First, we focus on the interaction between party and chamber type. According to the upper panel of [Figure 1](#), Republicans were more likely to use counterfactual expressions in the House, whereas there was no substantial difference in the use of counterfactuals between the different party representatives in the Senate. The interaction effect term between the party and chamber type confirms that such pattern is statistically significant ( $b_{\text{Party} \times \text{Senate}} = -.15$ ,  $p < .001$ ). Second, the time effect is contrasted between two chambers. As shown in the upper panel of [Figure 1](#), House members are less likely to make counterfactual expressions when the pandemic situation becomes worse, but Senators’ practice of counterfactual expressions less varies throughout the COVID-19 spread. The difference in the time effect is statistically significant ( $b_{\text{Senate} \times \text{Time}} = .08$ ,  $p = .01$ ).

Three implications for counterfactual expression can be derived from the findings reported in [Table 4](#) and [Figure 1](#). First, counterfactual expressions are favored among Republicans in the House and Democrats in the Senate. Second, party differences in the use of counterfactuals

are larger in the House than the Senate. Third, counterfactual expressions are more likely to be reserved, as the pandemic situation becomes worse.

Figure 1. Model-Based Predicted Ratios for Counterfactual Expression and Three Types of Time-Focusing



*Note.* Using the fitted Poisson models in Table 4, predicted ratios (0-100%) for counterfactual expression (upper panel) and three types of time-focusing (lower panel) in Congress members' speeches are illustrated.

For three types of time-focusing, the MP regression results for each type are provided (see the "Time-focusing" columns in Table 4). The lower panel of Figure 1 is drawn to assist the understanding of the patterns.

First, past-focusing exponentially increases with time ( $b_{\text{Time}} = .07, p < .001$ ;  $b_{\text{Time(squared)}} = .01, p = .01$ ) and such trend is found more dominantly among Senators, compared to House members ( $b_{\text{Senate} \times \text{Time}} = .03, p < .001$ ). While Republicans favor past-focused expressions more than Democrats ( $b_{\text{Party}} = .06, p = .02$ ), such partisan gap is quite stable throughout the pandemic ( $b_{\text{Party} \times \text{Time}} = .01, p = \text{n.s.}$ ).

Second, future-focusing continuously decreases ( $b_{\text{Time}} = -.04, p < .001$ ;  $b_{\text{Time(squared)}} = -.003, p = \text{n.s.}$ ), but the decreasing pattern is more prominent among House members, compared to Senators ( $b_{\text{Senate} \times \text{Time}} = .02, p = .01$ ). Unlike the case of past-focusing, Republicans exercise future-focused expressions more frequently than Democrats ( $b_{\text{Party}} = -.05, p = .001$ ), but such partisan gap does not vary across time ( $b_{\text{Party} \times \text{Time}} = -.01, p = \text{n.s.}$ ).

Third, present-focusing decreases among Senators but vary less among House members ( $b_{\text{Time}} = -.01, p < .001$ ;  $b_{\text{Time(squared)}} = -.01, p = .05$ ;  $b_{\text{Senate} \times \text{Time}} = -.02, p < .001$ ). Additionally, there was no distinct partisan differences for the use of present-focused expressions in both chambers ( $b_{\text{Party}} = -.002, p = \text{n.s.}$ ;  $b_{\text{Party} \times \text{Time}} = -.005, p = \text{n.s.}$ ).

As shown in Table 4 and Figure 1, our findings suggest that Republicans are more inclined to focus on the past, compared to Democrats. Also, the speeches have become dominantly past-focused in later stages of the pandemic.

## Discussion and Conclusion

This study examined US politicians' use of counterfactual expressions in Congressional speeches during COVID-19. The results show that counterfactuals were more practiced among Democrats in the Senate and Republicans in the House. As the pandemic worsened, the use of counterfactuals decreased, maintaining a partisan gap in the House. In the Senate, however, it was nearly stable, without party differences. Additionally, with time, politicians put more emphasis on the past and highlighted the future less. Those time trends in counterfactuals and time-focusing show that politicians are more likely to talk about "what happened" than "what would happen" in the later stages of the pandemic. We now discuss the findings from theoretical perspectives of Congressional deliberation, party polarization, and bicameral representation.

First, this study shows diverging time trends in the use of counterfactuals between two chambers. During the pandemic, Senators used counterfactual expressions more frequently than House members. There was no substantial difference between Republicans and Democrats in the Senate, but party differences were maintained in the House. As aimed in the US Constitution, Senators are older and more experienced than House members with longer terms and more discretionary power. Moreover, party polarization in the Senate is less severe than that in the House (Jamieson & Falk, 2000; Rippere, 2016). Senators' increasing past-focused and decreasing present- or future-focused utterances imply that their use of counterfactuals shifted to "counterfactual explanations" focusing on a mutable antecedent, expressing regrets over the present negative consequences.

However, the steady party differences in the House suggest that, compared to the Senate, political polarization might be more intense there. As language use reveals the speaker's value priorities, including those related to party identity (Jones et al., 2018; Robinson et al., 2015), party differences in counterfactuals would imply the disparity between Republican and Democratic identities. Despite certain limitations in representing polarization precisely, this approach is in line with previous studies, where differences in word usage were adduced as evidence of polarization concerning counterfactuals (Effron, 2018) or elite communication during the COVID-19 pandemic (Green et al., 2020; Hart et al., 2020).

Second, a more fascinating finding is the decreasing time trends among House members, with party differences when using counterfactuals. Compared to the Senate, two questions regarding House members' speech could be raised: (1) Why are the time trends decreasing? (2) Why do Republicans prefer to use counterfactuals? For the first question, one plausible answer could be found in the US Constitution. At the closure of the 116th Congress, the terms of all House members ended, but only one-third of the Senators ended. As "being re-elected" is the top priority for most Congress members (e.g., Lee, 2011b; Hood, 2011), House members are less inclined to make counterfactuals about the unparalleled pandemic which is exposing unresolved issues to the upcoming 117th Congress. Hence, House members who are uncertain about the re-election would be less motivated to make prefactuals or counterfactual explanations for the next Congress.

Regarding the second question, we cautiously make different interpretations of counterfactuals from Republicans and Democrats. While time trends of counterfactuals and future-focusing were decreasing in both partisans, Republicans' use of counterfactuals was always higher but their future-focusing was lower than that of Democrats. In other words, the counterfactuals in Republicans' speech put less emphasis on the future, implying they would be closer to counterfactual explanations than prefactuals (for similar past findings, see Robinson et al., 2015; Sylwester & Purver, 2015).



The use of counterfactual explanations indicates Republicans' political power and ascendancy under the Trump Administration, hinting at their justification motives for the massive failures (Scholl & Sassenberg, 2015). Such interpretation could be linked with recent studies about Republican supporters' reliance on "alternative facts" or ideological affinities to the former president Trump, popularizing negatively framed counterfactual thinking (Barrera et al., 2020; Effron, 2018; Kakutani, 2018). In this sense, Republican politicians' use of counterfactual explanations can be analyzed as counterfactual attacks (Bertolotti & Catellani, 2018). Continued failures of the Trump Administration left Republican politicians out of favor, driving them to adopt counterfactual strategies to confront their opponents (for a similar case, see Milesi & Catellani, 2011).

Moreover, experienced politicians in the House usually attempt to go forward to the Senate, so it seems plausible those Republicans' counterfactuals, mainly driven by so-called "post-truth" motives, would pervade in the Senate after several years. Such expectation should be taken with extreme caution because this study only identified formal linguistic, not semantic features in Congressional speeches.

Our findings reflect the bipartisan, bicameral political system in the US. Faced with the unparalleled public health risk, partisans in the Senate and the House demonstrated diverging patterns of counterfactuals use. However, their differences are not unique to the pandemic situation and are attributable to the systematic differences between the two parties and chambers in the US. Thus, we cautiously predict that the patterns are robust to the changing time trends of the pandemic. Even if the numbers of confirmed cases or deaths change, the systematic differences may remain consistent.

While this study is the first to detect and model the uncovered politicians' counterfactuals, several limitations should be overcome in future studies. First, due to time and resource constraints, we indirectly inferred the characteristics of counterfactuals by treating time-focusing displayed in speeches as contextual information. To the best of our knowledge, no NLP tools were launched to judge or predict the semantic information of counterfactuals. In this sense, manual coding would be attemptable to understand the semantics of politicians' use of counterfactuals.

Second, it is necessary to extract more textual cues, such as emotions displayed (e.g., hope or fear for prefactuals; relief or regret for counterfactual explanations), direction (e.g., upward or downward counterfactuals; subtractive or additive counterfactuals) or focus of mutability (e.g., mutable antecedent or mutable consequent), or target (e.g., self or others; Byrne, 2016; Epstude & Roese, 2008).

Third, politicians' characteristics could be added to the analyses. In Congress, politicians need colleagues' agreement to proceed with the legislative process, meaning party identity is one of several different factors influencing politicians' Congressional speech. For example, different partisans reach an agreement based on district-based interests, and demographic features of districts or states constrain Congress members' behaviors (e.g., Lee, 2011b).

Congress should remain as a discursive system integrating different voices in democracy. Since the effectiveness of democracy frequently depends on the quality of Congressional deliberation, tracking textual features of Congressional speech, including counterfactuals, would be a linguistic diagnosis of the discursive health of the democracy. Our study is meaningful and valuable since it is the first empirical research tracking the use of counterfactuals and prefactuals in Congressional speeches.

## **Supplementary Materials**

The Supplementary Materials comprise two sections (for access see Index of Supplementary Materials below). In Section A, the counterfactual detection algorithm is explained in detail.

In Section B, we described the statistical methods adopted in the study (i.e., the multi-level zero-inflated Poisson and multi-level Poisson models).

### **Index of Supplementary Materials**

Park, R. & Baek, Y. M. (2021). *Supplementary Materials to “Talking About What Would Happen Versus What Happened: Tracking Congressional Speeches during COVID-19”* [Additional information and analyses]. PsychOpen GOLD. [DOI NEED TO BE ASSIGNED].

### **Funding**

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2020S1A5C2A03093177).

### **Competing Interests**

The authors have declared that no competing interests exist.

### **Acknowledgements**

The authors have no support to report.

## References

- <bok>Aristotle. (1991). *On rhetoric* (G. A. Kennedy, Trans.). New York, NY, USA: Oxford University Press (Original work published ca. 400 B.C.E.)</bok>
- <jrn>Barrera, O., Guriev, S., Henry, E., & Zhuravskaya, E. (2020). Facts, alternative facts, and fact checking in times of post-truth politics. *Journal of Public Economics*, 182, Article 104123. <https://doi.org/10.1016/j.jpubeco.2019.104123></jrn>
- <eref>Benoit, K., & Matsuo, A. (2020). spacyr (R package version 1.2.1) [Computer software]. Retrieved from <https://spacyr.quanteda.io></eref>
- <jrn>Bertolotti, M., & Catellani, P. (2018). The effects of counterfactual attacks on the morality and leadership of different professionals. *Social Psychology*, 49(3), 154–167. <https://doi.org/10.1027/1864-9335/a000338></jrn>
- <jrn>Brady, W. J., Wills, J. A., Burkart, D., Jost, J. T., & Van Bavel, J. J. (2019). An ideological asymmetry in the diffusion of moralized content on social media among political leaders. *Journal of Experimental Psychology: General*, 148, 1802–1813. PubMed <https://doi.org/10.1037/xge0000532></jrn>
- <jrn>Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Mächler, M., & Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal*, 9, 378–400. <https://doi.org/10.32614/RJ-2017-066></jrn>
- <jrn>Byrne, R. M. (2016). Counterfactual thought. *Annual Review of Psychology*, 67, 135–157. PubMed <https://doi.org/10.1146/annurev-psych-122414-033249></jrn>
- <jrn>Caprara, G. V., Schwartz, S., Capanna, C., Vecchione, M., & Barbaranelli, C. (2006). Personality and politics: Values, traits, and political choice. *Political Psychology*, 27, 1–28. <https://doi.org/10.1111/j.1467-9221.2006.00447.x></jrn>
- <conf>Carlson, L., Marcu, D., & Okurowski, M. E. (2001). Building a discourse-tagged corpus in the framework of Rhetorical Structure Theory. In *SIGDIAL '01: Proceedings of the Second SIGdial Workshop on Discourse and Dialogue* (Vol. 16). <https://doi.org/10.3115/1118078.1118083></conf>
- <eref>Congressional Research Service. (2012, February 17). *Representatives and senators: Trends in member characteristics since 1945*. Retrieved from <https://fas.org/sgp/crs/misc/R42365.pdf></eref>
- <eref>Congressional Research Service. (2020, December 17). *Membership of the 116th congress: A profile*. Retrieved from <https://fas.org/sgp/crs/misc/R45583.pdf></eref>
- <jrn>Dawid, A. P. (2000). Causal inference without counterfactuals. *Journal of the American Statistical Association*, 95, 407–424. <https://doi.org/10.1080/01621459.2000.10474210></jrn>
- <jrn>Effron, D. A. (2018). It could have been true: How counterfactual thoughts reduce condemnation of falsehoods and increase political polarization. *Personality and Social Psychology Bulletin*, 44, 729–745. PubMed <https://doi.org/10.1177/0146167217746152></jrn>
- <jrn>Epstude, K., & Roese, N. J. (2008). The functional theory of counterfactual thinking. *Personality and Social Psychology Review*, 12, 168–192. PubMed <https://doi.org/10.1177/1088868308316091></jrn>
- <jrn>Epstude, K., Scholl, A., & Roese, N. J. (2016). Prefactual thoughts: Mental simulations about what might happen. *Review of General Psychology*, 20, 48–56. <https://doi.org/10.1037/gpr0000064></jrn>

- <edb>Feng, V. W., & Hirst, G. (2012). Text-level discourse parsing with rich linguistic features. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics* (Vol. 1: Long Papers, pp. 60-68). Retrieved from <https://aclanthology.org/P12-1007></edb>
- <jrn>Green, J., Edgerton, J., Naftel, D., Shoub, K., & Cranmer, S. J. (2020). Elusive consensus: Polarization in elite communication on the COVID-19 pandemic. *Science Advances*, 6, Article eabc2717. PubMed <https://doi.org/10.1126/sciadv.abc2717></jrn>
- <jrn>Hart, P. S., Chinn, S., & Soroka, S. (2020). Politicization and polarization in COVID-19 news coverage. *Science Communication*, 42, 679–697. <https://doi.org/10.1177/1075547020950735></jrn>
- <bok>Hood, C. (2011). *The blame game: Spin, bureaucracy, and self-preservation in government*. Princeton, NJ, USA: Princeton University Press.</bok>
- <bok>Hume, D. (1988). *An enquiry concerning human understanding*. Chicago, IL, USA: Open Court Press. (Original work published in 1748)</bok>
- <edb>Jamieson, K. H., & Falk, E. (2000). Continuity and change in civility in the House. In J. R. Bond & R. Fleisher (Eds.), *Polarized politics: Congress and the President in a partisan era* (pp. 96-108). Washington, DC, USA: CQ Press.</edb>
- <eref>Janocko, A., Larche, A., Raso, J., & Zembroski, K. (2016). *Counterfactuals in the language of social media: A natural language processing project in conjunction with the World Well-Being Project*. Retrieved from [https://www.seas.upenn.edu/~cse400/CSE400\\_2015\\_2016/reports/report\\_15.pdf](https://www.seas.upenn.edu/~cse400/CSE400_2015_2016/reports/report_15.pdf)</eref>
- <jrn>Jones, K. L., Noorbaloochi, S., Jost, J. T., Bonneau, R., Nagler, J., & Tucker, J. A. (2018). Liberal and conservative values: What we can learn from congressional tweets. *Political Psychology*, 39, 423–443. <https://doi.org/10.1111/pops.12415></jrn>
- <bok>Kakutani, M. (2018). *The death of truth: Notes on falsehood in the age of Trump*. New York, NY, USA: Tim Duggan Books.</bok>
- <edb>Lee, F. E. (2011a). Bicameral representation. In E. Schickler & F. E. Lee (Eds.), *The Oxford handbook of the American Congress* (pp. 272-292). New York, NY, USA: Oxford University Press.</edb>
- <jrn>Lee, F. E. (2011b). Congress and the executive branch: Delegation and presidential dominance. In E. Schickler & F. E. Lee (Eds.), *The Oxford handbook of the American Congress* (pp. 789-811). New York, NY, USA: Oxford University Press.</jrn>
- <jrn>Lin, Z., Ng, H., & Kan, M. (2014). A PDTB-styled end-to-end discourse parser. *Natural Language Engineering*, 20, 151–184. <https://doi.org/10.1017/S1351324912000307></jrn>
- <jrn>McGraw, K. M. (1991). Managing blame: An experimental test of the effects of political accounts. *The American Political Science Review*, 85, 1133–1157. <https://doi.org/10.2307/1963939></jrn>
- <jrn>McNamara, P., Durso, R., Brown, A., & Lynch, A. (2003). Counterfactual cognitive deficit in persons with Parkinson's disease. *Journal of Neurology, Neurosurgery, and Psychiatry*, 74, 1065–1070. PubMed <https://doi.org/10.1136/jnnp.74.8.1065></jrn>
- <jrn>Milesi, P., & Catellani, P. (2011). The day after an electoral defeat: Counterfactuals and collective action. *British Journal of Social Psychology*, 50, 690–706. PubMed <https://doi.org/10.1111/j.2044-8309.2011.02068.x></jrn>

- <jrn>Pan, L., McNamara, G., Lee, J. J., Halebian, J., & Devers, C. E. (2018). Give it to us straight (most of the time): Top managers' use of concrete language and its effect on investor reactions. *Strategic Management Journal*, 39, 2204–2225. <https://doi.org/10.1002/smj.2733></jrn>
- <jrn>Pearl, J. (2000). Comment on A.P. Dawid's, Causal inference without counterfactuals. *Journal of the American Statistical Association*, 95, 428–431.</jrn>
- <bok>Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect*. New York, NY, USA: Basic Books.</bok>
- <conf>Pennebaker, J. W., Booth, R. J., Boyd, R. L., & Francis, M. E. (2015). *Linguistic Inquiry and Word Count: LIWC2015*. Austin, TX, USA: Pennebaker Conglomerates.</conf>
- <jrn>Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology*, 77, 1296–1312. [PubMed https://doi.org/10.1037/0022-3514.77.6.1296](https://doi.org/10.1037/0022-3514.77.6.1296)</jrn>
- <conf>Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A., & Webber, B. (2008). The Penn discourse treebank 2.0. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*. Retrieved from [http://www.lrec-conf.org/proceedings/lrec2008/pdf/754\\_paper.pdf](http://www.lrec-conf.org/proceedings/lrec2008/pdf/754_paper.pdf)</conf>
- <jrn>Rippere, P. S. (2016). Polarization reconsidered: Bipartisan cooperation through bill cosponsorship. *Polity*, 48, 243–278. <https://doi.org/10.1057/pol.2016.4></jrn>
- <jrn>Robinson, M. D., Cassidy, D. M., Boyd, R. L., & Fetterman, A. K. (2015). The politics of time: Conservatives differentially reference the past and liberals differentially reference the future. *Journal of Applied Social Psychology*, 45, 391–399. <https://doi.org/10.1111/jasp.12306></jrn>
- <jrn>Roese, N. J., Park, S., Smallman, R., & Gibson, C. (2008). Schizophrenia involves impairment in the activation of intentions by counterfactual thinking. *Schizophrenia Research*, 103, 343–344. [PubMed https://doi.org/10.1016/j.schres.2007.05.006](https://doi.org/10.1016/j.schres.2007.05.006)</jrn>
- <jrn>Scholl, A., & Sassenberg, K. (2015). Better know when (not) to think twice: How social power impacts prefactual thought. *Personality and Social Psychology Bulletin*, 41, 159–170. [PubMed https://doi.org/10.1177/0146167214559720](https://doi.org/10.1177/0146167214559720)</jrn>
- <edb>Sinclair, B. (2000). Individualism, partisanship, and cooperation in the Senate. In B. A. Loomis (Ed.), *Esteemed colleagues: Civility and deliberation in the U.S. Senate* (pp. 59–77). Washington, DC, USA: Boorkings Institution Press.</edb>
- <conf>Son, Y., Buffone, A., Janocko, A., Larche, A., Raso, J., Zembroski, K., Schwartz, H. A., & Ungar, L. (2017). Recognizing counterfactual thinking in social media texts. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics* (Vol. 2: Short Papers, pp. 654–658). <https://doi.org/10.18653/v1/P17-2103></conf>
- <jrn>Sylwester, K., & Purver, M. (2015). Twitter language use reflects psychological differences between democrats and republicans. *PLoS One*, 10, Article e0137422. [PubMed https://doi.org/10.1371/journal.pone.0137422](https://doi.org/10.1371/journal.pone.0137422)</jrn>
- <bok>Tetlock, P. E., & Belkin, A. (1996). *Counterfactual thought experiment in world politics: Logical, methodological, and psychological perspective*. Princeton, NJ, USA: Princeton University Press.</bok>
- <eref>The New York Times. (2020). Coronavirus (Covid-19) data in the United States. Retrieved November 14, 2020, from <https://github.com/nytimes/covid-19-data></eref>



<bok>Thucydides. (1974). *History of the Peloponnesian War* (R. Warner, Trans.). London, United Kingdom: Penguin Books. (Original work published ca. 400 B.C.E.)</bok>

<eref>Wallach, P. A., & Myers, J. (2020). *The federal government's coronavirus response-public health timeline*. Retrieved from <https://www.brookings.edu/research/the-federal-governments-coronavirus-actions-and-failures-timeline-and-themes></eref>

<bok>Weber, M. (1919). *Politics as a vocation* (R. Livingston, Trans.). Indianapolis, IN, USA: Hackett Books. (Original work published in 1919).</bok>

<eref>World Health Organization. (2020). WHO timeline—COVID-19. Retrieved from <https://www.who.int/news-room/detail/27-04-2020-who-timeline—covid-19></eref>

<eref>Worldometer. (2021, January 17). Reported cases and deaths by country, territory, or conveyance. Retrieved from <https://www.worldometers.info/coronavirus/#countries></eref>

<prpt>Yang, X., Obadinma, S., Zhao, H., Zhang, Q., Matwin, S., & Zhu, X. (2020). *Semeval-2020 task 5: Counterfactual recognition*. arXiv. Retrieved from <https://arxiv.org/pdf/2008.00563.pdf></prpt>

<sup>1</sup><https://universaldependencies.org/u/dep/index.html>

## Supplementary Materials to “Talking About What Would Happen Versus What Happened: Tracking Congressional Speeches during COVID-19”

### Section A. Counterfactual Detection Rules

Section A elaborates on the procedure of counterfactual detection through two tables and one figure. Parsing and tagging results (see [Table S1](#)), the visualized dependency tree (see [Figure S1](#)), and the extended set of counterfactual forms and syntactic rules (see [Table S2](#)) are provided with supplementary text.

Our counterfactual detection rules utilize three types of lexical and syntactic information, based on version 1.2.1 of the R “spacyr” package (Benoit & Matsuo, 2020).<sup>1</sup> After preprocessing (i.e., removing redundant white spaces, special symbols, and stopwords), each sentence in the Congressional speech corpus was split into individual words or punctuation, a process known as “tokenization.” For each token, we identified its “lemma,” “POS (parts-of-speech) tag,” and “dependency relation.” First, “lemma” is a base token form, and “lemmatization” includes converting plural nouns to singular nouns (e.g., “districts” to “district” in [Table S1](#)), finding the root verb of a past-tense verb (e.g., transforming “was” or “were” into “be”), and others.

Second, “POS tagging” is the process in which grammatical information, such as verbs or nouns, are tagged into tokens. POS tags can be either Universal or Detailed. To illustrate, “was” is simply a verb (i.e., “VERB”) in the Universal tag set,<sup>2</sup> but it can also be tagged as Detailed as “VBD” or past-tense verb. Our study utilized the detailed POS tags of the Penn Treebank, as provided in the R “spacyr” package.

Third, the relations between tokens are obtainable through “dependency parsing.” Each token serves a role in each sentence, for example, “to modify” other tokens or “to be modified” by them. In other words, a sentence has a hierarchical structure in which tokens are interdependent to other tokens. Thus, we represented each sentence as a CoNLL-style dependency tree since its dependency structure can reflect the unidirectional relationship between the antecedent and consequent in a counterfactual expression. Please refer to the results of the dependency parsing (see the “Relation” column in [Table S2](#)) and the visualized dependency tree (see [Figure S1](#)). The example sentence in Table 1 (“If we do not move this bill, the death will be in our districts.”) is tokenized into 16 tokens, and each token is labelled with a lemma, a POS tag, and a dependency relation.

Our counterfactual detection rules are primarily based on the single root verb and its leaf verbs.<sup>3</sup> In [Table S1](#), the root verb is “Be,” and “Move” is the only verb depending on the root verb. Therefore, we suspect the “Be” and “Move” clauses to be the consequent and the antecedent, respectively. To identify whether the consequent or the antecedent is in past or present tense, or whether it is a modal verb (e.g., should, could), we additionally extract auxiliary particles (i.e., “aux” or “neg” in the “Relation” column), depending on either of the two verbs. Please note that we are focusing on the verbs only and not other POS (e.g., root and leaf nouns) as we need to identify a hypothetical situation in counterfactuals.

<sup>1</sup> The “spacyr” package is the R wrapper to the Python “spaCy” package. For details, please visit the “spacyr” package webpage: <https://cran.r-project.org/web/packages/spacyr/index.html>

<sup>2</sup> <https://universaldependencies.org/u/pos/all.html>

<sup>3</sup> As an additional note, we also used the root verb to identify time-focusing of the sentence. For example, the consequent in [Table S1](#) is “will be,” so the sentence will be labelled as “future-focusing.” Likewise, if the root verb is past tense (i.e., “VBD”) or base form (i.e., “VB”) without “will,” the time-focusing would be “past-” or “present-focusing” in each case.

Table S1. Example of Tokenization, Lemmatization, POS Tagging, and Dependency Parsing

ID (#)	Token	Lemma	POS Tag <sup>a</sup>	Dependency	
				Head Token ID (#)	Relation <sup>b</sup>
1	If	If	IN	5	mark
2	We	-PRON-	PRP	5	nsubj
3	<u>Do</u>	Do	VBP	5	aux
4	<u>Not</u>	Not	RB	5	neg
5	<u>Move</u>	Move	VB	12	advcl
6	This	This	DT	7	det
7	Bill	Bill	NN	5	dobj
8	,	,	,	12	punct
9	The	The	DT	10	det
10	Death	Death	NN	12	nsubj
11	<u>Will</u>	Will	MD	12	aux
12	<u>Be</u>	Be	VB	12	ROOT
13	In	In	IN	12	prep
14	Our	-PRON-	PRP\$	15	poss
15	districts	District	NNS	13	pobj
16	.	.	.	12	punct

*Note.* Using the `spacy_parse()` function in the R “spacyr” package (version 1.2.1), the annotation results of the first sentence in [Table 1](#) (“If we do not move this bill, the death will be in our districts.”) are displayed. For dependency parsing, “Relation” denotes the relationship between the token and its head token, matched with the “Head Token ID (#).”

Based on the dependencies, two hierarchical relations are shaded: (1) “Be (#12, ROOT)” and its dependencies in grey color, (2) “Move (#5, advcl)” and its dependencies in orange color. Note that “Move” is the only verb depending on the root verb, and all the shaded tokens are either verbs (i.e., “VB” in the “POS Tag” column) or auxiliary particles (i.e., “aux” or “neg” in the “Relation” column). Using these verbal hierarchies, the antecedent and consequent candidates are underlined. See [Figure S1](#) for the visualized results of [Table S1](#).

<sup>a</sup>POS (part-of-speech) tags from the Penn Treebank tag set. For shaded areas, VBP = Verb, 3<sup>rd</sup> person singular present; RB = Adverb; VB = Verb, base form; MD = Modal.

<sup>b</sup>Universal dependency relations in the Stanford Dependencies. For shaded areas, aux = auxiliary; neg = negative; advcl = adverbial clause modifier; ROOT = root. More details are available in the R “spacyr” package manual.

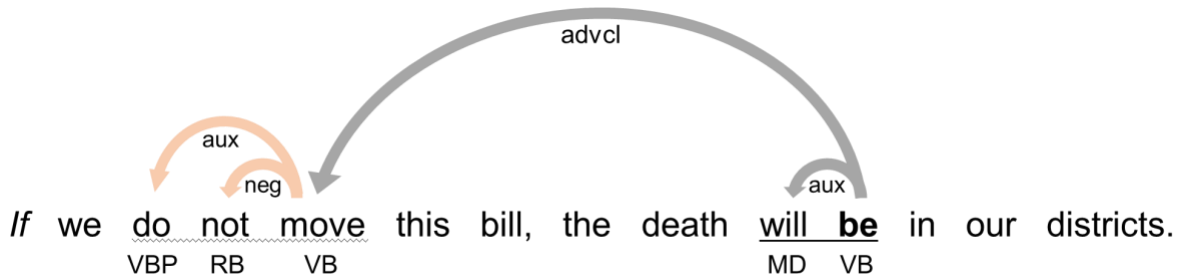
Table S2. Three Forms of Counterfactual Expression and Syntactic Rules

Form	Example	Keyword	Antecedent	Consequent
Conditional Conjunction	<i>If we <u>do not move</u> this bill, the death <u>will be</u> in our districts. (Mr. GARAMENDI)</i>	“If” or “Unless”	Base verb, past-tense verb, or modal verb	Modal verb
	<i><u>Unless</u> we <u>act</u> soon, we <u>will see</u> mass layoffs, devastating tax increases, and a breakdown in public safety and essential services. (Mr. MENENDEZ)</i>			

Verb Inversion	<u>Had</u> he (= Mr. HOYER) been serious about that, there <u>would have been</u> a discussion before this bill ever came to the floor. (Mr. PALMER) <u>Were</u> I (= Judge Ginsburg) to <u>rehearse</u> here what I would say and how I would reason on such questions, I <u>would act</u> injudiciously. (quotes Mr. LANKFORD)	“Had” or “Were”	Past-tense verb	Modal verb
Wish/Should	I <u>wish</u> this bill <u>was not</u> necessary, but unfortunately, it is now more imperative than ever. (Mr. LOFGREN) We <u>shouldn't</u> be asking that question. That <u>should have been</u> a question for the Afghans. (Mr. GOHMERT)	“wish” or “should”	Past-tense (modal) verb	N/A

*Note.* Extending Table 1 in the main text, the root verb, consequent, antecedent and keyword are highlighted in each of the six sentences. Note that the “Verb Inversion” form allows only upper cases of “Had” or “Were” at the beginning of a sentence, whereas the “Conditional Conjunction” form includes both upper and lower cases of “If” or “Unless” (c.f., excluding “as if”). For the “Wish/Should” form, the consequent is omitted yet implied as a better situation. Especially for the “Should” verb, our syntactic rules necessitate the presence of a VBN (Verb, past particle) in the antecedent (e.g., should have “been”).

Figure S1. An Example of CoNLL-style Dependency Tree



*Note:* Based on the annotation results in Table S1, the dependency structure of the antecedent and consequent is presented. Colors of arrow are matched to either of the two hierarchies (i.e., “be” in grey; “move” in orange), along with “Relation” tags underneath the arrows and POS tags below the line.

Using the annotation results, we can extract counterfactual expressions by matching the results to the predefined set of three counterfactual forms. In main text, we have introduced the counterfactual detection rules using the three example sentences in Table 1. In Table S2, the extended set of six counterfactual sentences is presented. As mentioned in main text, the syntactic rules for each form are as follows (please refer to relevant examples in parentheses).

Regarding the “Conditional Conjunction” form, the consequent consists of a modal verb (e.g., “will be” or “will see”) while the antecedent consists of a base verb (e.g., “move” or “act”), a past-tense verb, or a modal verb, preceded by a conditional conjunction, such as “If” or “Unless.” For the “Verb Inversion” form, the sentence begins with a verb inversion, like “Had he ...” or “Were I ...”), followed by a past-tense antecedent and a modal consequent. Accordingly, the “Verb Inversion” form allows only upper cases of “Had” or “Were” at the

beginning of a sentence, whereas the “Conditional Conjunction” form includes both upper and lower cases of “If” or “Unless” (c.f., excluding “as if”). In the “Wish/Should” form, the clause in the past tense, following “wish” (a past-tense verb; e.g., “was”) or “should” (a past-tense modal verb; e.g., “should have been”), is the antecedent, and the consequent is omitted yet implied as a better situation. Especially for the “Should” verb, our syntactic rules necessitate the presence of a VBN (Verb, past particle) in the antecedent (e.g., should have “been”).

## Section B. MZIP and MP Model Description

Section B provides a full description of the multi-level zero-Inflated Poisson (MZIP) and the multi-level Poisson (MP) models adopted in the study. In the following paragraphs, we demonstrated the rationale for using an MP approach with an offset variable. The need and importance of a zero-inflated model were pointed out, especially for counterfactual expressions, but not for time-focusing.

First, it is reasonable to expect that our dependent variables—the number of counterfactual expressions or verbs whose tenses are past, present, or future—are more likely to appear in longer speeches consisting of many sentences. Regarding the first issue, we treated the total number of sentences in a speech as an offset (Long, 1996), modeling the incidence ratio of an individual member using counterfactual expressions or specific time-orientation in a speech.

Second, Congressional speech data are nested with multiple hierarchies. Individual speeches may be correlated if they were made by the same member of Congress, or by those from the common state. To statistically address those cluster effects across members and states, random effects are modeled. In brief, we fitted an MZIP model to estimate the number of counterfactual expressions and time-focusing in Congressional speech.

Finally, as described above, counterfactual expressions are not frequently used in Congressional speech, meaning that there are excessive zeros in the number of counterfactual expressions (i.e., about 1% of total sentences were counterfactual sentences). In other words, conventional regression models for count variables, such as simple Poisson or negative binomial regressions, fail to predict the occurrence of counterfactual expressions adequately. To address this, the zero-inflated Poisson (ZIP) regression model (Cameron & Trivedi, 2013) was adopted. Statistically, the ZIP regression assumes a mixed distribution of zero and positive counts, meaning that it simultaneously estimates whether an event (i.e., counterfactual expression) occurs, as well as the number of times it occurs (i.e., frequency of counterfactual expressions). The ZIP regression, comprising the logistic and Poisson model, predicts zero and non-zero count responses, respectively. Unlike counterfactual expressions, time-focusing variables do not suffer from excessive zeros, so we did not implement zero-inflated models.

The model formula of the MZIP model is as follows: Under the standard ZIP model, the  $i^{\text{th}}$  observation  $Y_i$  follows the mixture distribution composed of two parts, where  $\phi_i$  is the proportion of zero responses from a logistic model, and  $\lambda_i$  is the rate parameter of the Poisson model.

$$P(Y_i = y) = \begin{cases} \phi_i + (1 - \phi_i)e^{-\lambda_i}, & y = 0 \\ (1 - \phi_i) \frac{\lambda_i^y e^{-\lambda_i}}{y!}, & y \geq 1 \end{cases}$$

Let  $Y_{ij}$  denote the  $i^{\text{th}}$  observation of the  $j^{\text{th}}$  subject. Then, we can consider the subject random effect ( $u_j$ ) as a part of the Poisson model. Using the log link function, the rate parameter  $\lambda_i$  is modeled as shown. This can be viewed as an extension of the generalized linear mixed model (GLMM) with a mixture of distributions.



$$\log(\lambda_{ij}) = x_{ij}^T \beta + u_j$$

For time-focusing, by contrast, the dependent measures are count variables clustered across members and states, but presumably not zero-inflated. Over 98% of the total sentences were successfully classified as past (18.3%), present (68.2%), or future (11.6%) focused. Thus, when modeling time-focusing, we adopted MP regression models without implementing a zero-inflated model.

## References

- <eref>Benoit, K., & Matsuo, A. (2020). spacyr (R package version 1.2.1) [Computer software]. Retrieved from <https://spacyr.quanteda.io></eref>
- <bok>Cameron, A. C., & Trivedi, P. K. (2013). *Regression analysis of count data*. Cambridge, United Kingdom: Cambridge University Press.</bok>
- <bok>Long, J. S. (1996). *Regression for categorical and limited dependent variables*. Thousand Oaks, CA, U.S.A.: Sage Publications.</bok>