Online Appendix for Causal Inference with Latent Treatments*

Contents

*Intended for online publication only.

A The Extensive Use of Latent Treatments, Extended Table

Table 1 is a list of the 51 articles published in the *American Journal of Political Science* and the *American Political Science Review* that use a text-based treatment, along with the limitations associated with their research designs. We code a research design as focused on a latent variable if the key variable of interest is a function of the text, rather than the actual text itself. We say that a treatment is aliased if more than the latent treatment of interest varies across treatments. And we say that there is only a single vignette if only text is written to deliver the latent treatment. This is perhaps clearest with conjoint studies. It is true, in principle, that many different values in a conjoint study could be used to express the same "high-income" treatment, the papers that we evaluated tended to use only a single level of the conjoint to deliver a latent treatent.

B Table of Terminology

Table 2 defines our key terms and variables.

C Relationship to Causal Mediation

Our context resembles causal mediation, but there are crucial differences which imply that the key concerns that motivate the identification and estimation issues in mediation are not directly applicable to the study of latent treatments. Like mediation, we seek to study the effect of a text on an outcome through some intermediate variable. In causal mediation, the analyst seeks to study the effect of a treatment through some intermediate outcome. But in the latent treatment setting, rather than an intermediate outcome, the intermediate variable is a deterministic function of the text. As a result, given the text, there is no variation in the latent treatment. A given campaign advertisement is negative or not; there is no way that, given the text, the advertisement is sometimes negative and sometimes not.

To see this key issue more concretely, we show that the individual level indirect effect cannot possibly be defined with texts. Blending notation from our framework with the usual mediation notation, let the text (the treatment) be $\boldsymbol{X}_i$ and the latent treatment (the mediator) be Z_i. Defining unit-level indirect effects would require defining terms of the following form:

$$Y_i(\boldsymbol{X}_i = t, Z_i = 1) - Y_i(\boldsymbol{X}_i = t, Z_i = 0)$$

The challenge addressed by mediation analysis is the possibility that at most one of $Y_i(\boldsymbol{X}_i = t, Z_i = 1)$ and $Y_i(\boldsymbol{X}_i = t, Z_i = 0)$ can be observed for any given unit. In our setting, the problem is deeper: either $Y_i(\boldsymbol{X}_i = t, Z_i = 1)$ or $Y_i(\boldsymbol{X}_i = t, Z_i = 0)$ is undefined, because any given text either possesses the latent treatment or does not possess it, because the mediator is a deterministic function of the text. In addition to the theoretical challenge posed by the inability to define direct effects, this property of latent treatments renders popular designs for causal mediation invalid. For

Table 1: Literature Review

Study	Type	Latent	Aliased	Single
Ward (2019)	Conjoint Experiment	✓		✓
Hankinson (2018)	Conjoint Experiment	✓		✓
Teele, Kalla and Rosenbluth (2018)	Conjoint Experiment	✓		✓
Carnes and Lupu (2016)	Conjoint Experiment	✓		✓
Huff and Kertzer (2018)	Conjoint Experiment	✓		✓
Butler and Hassell (2018)	Field Experiment	✓	✓	✓
Grossman and Michelitch (2018)	Field Experiment	✓	✓	✓
Valenzuela and Michelson (2016)	Field Experiment	✓	✓	✓
Kalla and Broockman (2018)	Field Experiment			
Han (2016)	Field Experiment	✓	✓	
McConnell et al. (2018)	Field Experiment	✓		✓
Hemker and Rink (2017)	Field Experiment	✓		✓
Karpowitz, Monson and Preece (2017)	Field Experiment	✓	✓	✓
Foos and de Rooij (2017)	Field Experiment	✓	✓	✓
Einstein and Glick (2017)	Field Experiment	✓		✓
Corbacho et al. (2016)	Field Experiment	✓	✓	✓
Driscoll et al. (2018)	Field Experiment	✓		✓
Brierley, Kramon and Ofosu (2019)	Field Experiment	✓		
Sheffer et al. (2018)	Lab Experiment	✓	✓	✓
Enos and Gidron (2018)	Lab Experiment	✓	✓	✓
Druckman, Levendusky and McLain (2018)	Lab Experiment	✓	✓	✓
Ryan (2017)	Lab Experiment	✓		✓
De Benedictis-Kessner et al. (2019)	Survey Experiment	✓	✓	✓
Carlson (2019)	Survey Experiment	✓		✓
Mummolo and Peterson (2019)	Survey Experiment	✓	✓	✓
Barber and Pope (2019)	Survey Experiment	✓		✓
Li (2018)	Survey Experiment	✓		✓
Anoll (2018)	Survey Experiment	✓	✓	✓
Simonovits, Kezdi and Kardos (2018)	Survey Experiment	✓	✓	✓
Healy, Kosec and Mo (2017)	Survey Experiment	✓	✓	✓
McEntire, Leiby and Krain (2015)	Survey Experiment	✓	✓	✓
Bloom, Arikan and Courtemanche (2015)	Survey Experiment	✓	✓	✓
Huddy, Mason and Aarøe (2015)	Survey Experiment	✓	✓	✓
Baker (2015)	Survey Experiment	✓	✓	✓
Reeves and Rogowski (2018)	Survey Experiment	✓	✓	✓
Thachil (2017)	Survey Experiment	✓		✓
Kertzer and Zeitzoff (2017)	Survey Experiment	✓	✓	✓
Pérez and Tavits (2017)	Survey Experiment	✓		✓
Gaikwad and Nellis (2017)	Survey Experiment	✓		✓
Christenson and Kriner (2017)	Survey Experiment	✓	✓	✓
Clayton, O'Brien and Piscopo (2019)	Survey Experiment	✓		✓
Mattes and Weeks (2019)	Survey Experiment	✓		✓
Kosmidis (2018)	Survey Experiment	✓	✓	✓
Klar (2018)	Survey Experiment	✓		✓
Balmas (2018)	Survey Experiment	✓	✓	✓
Bisgaard and Slothuus (2018)	Survey Experiment	✓		✓
Corstange and York (2018)	Survey Experiment	✓	✓	✓
Clifford and Jerit (2018)	Survey Experiment	✓	✓	✓
Bisgaard (2019)	Survey Experiment	✓	✓	✓
Hill and Huber (2019)	Survey Experiment	✓	✓	
Broockman, Ferenstein and Malhotra (2019)	Survey Experiment	✓	✓	✓

Table 2: Definition of Key Terms

Term	Definition
Latent	Property obtained from applying a function to a high-dimensional intervention, where the function maps from a higher dimension to a lower dimension.
Latent Treatment	Treatment vector obtained from applying codebook function g to a text.
Unmeasured Latent Treatment	Vector of features (potentially treatments in other settings) obtained from applying a function h to texts
$\boldsymbol{X}_i$	Text assigned to unit i
g	Codebook function, applied to obtain latent treatments
h	Unmeasured treatment codebook function, applied to obtain unmeasured latent treatments for theoretical purposes but unknown in practice
$\boldsymbol{Z}_i$	The vector of latent treatments obtained from applying g to $\boldsymbol{X}_i$
Z_i	The binary latent treatment obtained from applying g to $\boldsymbol{X}_i$ if the analyst measures only a single latent treatment
$\boldsymbol{B}_i$	The vector of latent confounders obtained from applying h to $\boldsymbol{X}_i$

example, in the parallel design from Imai et al. (2011), the researcher must assume that both the treatment and the mediator are as good as randomized—a sequential ignorability assumption. In our setting, it is impossible to independently randomize the treatment and the mediator. The inability to separate the treatment assignment from the value of the mediator motivates a new, strong assumption that differs from the assumptions of causal mediation analysis.

D Relationship to Dafoe, Zhang, and Caughey (2018)

Our paper builds on Dafoe, Zhang, and Caughey (2018) (DZC), but it has important distinctions and offers some conceptual clarity to the original paper. In this section, we discuss the distinctions with DZC. We emphasize, however, that because our work is focused on a distinct problem, most of the issues that we describe here could be placed within the DZC framework as a prior step in their framework.

D.1 Comparing Problem Setups

DZC are interested in isolating the effect of a particular belief on an outcome of interest. For example, we might be interested in how believing an adversarial country is a democracy affects support for war with that country. Of course, if we tell someone that a country is democratic they might infer other features about that country—it is more likely to be in Europe, wealthy, etc. These other beliefs confound what DZC call the epistemic effect of merely manipulating democracy.

To define their notation, DZC suppose that there is a treatment $Z_i \in \{0, 1\}$ and that this treatment potentially affects background beliefs $B_i \in \{0, 1\}$ and a belief of interest $D_i \in \{0, 1\}$. That is, they are interested in the effects of a respondent's beliefs on the outcome and use texts to influence those beliefs. This is fundamentally different than our enterprise; we are interested in the effect of exposure to the text, and are agnostic about the specific beliefs the text generates. Even so, it is possible and instructive to combine the two sets of concerns into a single graph.

Figure 1 combines latent treatments from our framework and beliefs from DZC to emphasize that they are distinct concerns and that confounding from latent treatments can arise even in the absence of confounding of beliefs . Figure 1a shows a setting where we manipulate a text-based treatment to affect beliefs, but the background beliefs don't confound the belief of interest. Yet, the issues that we raise in this paper still remain, because we are only able to manipulate the text $\boldsymbol{X}_i$. This occurs because in our setting we explicitly allow for the outcome to depend on the unmeasured treatments. If those happen to be correlated with the measured treatments, there is confoudning.

Figure 1b shows our framework fully mixed with the DZC framework. This raises some additional concerns from the DZC paper that analysts should consider if attempting to isolate the effect of texts on beliefs. Specifically, the measured and unmeasured treatments can affect both the background beliefs and the treatment of interest. This can complicate the conditions in DZC.

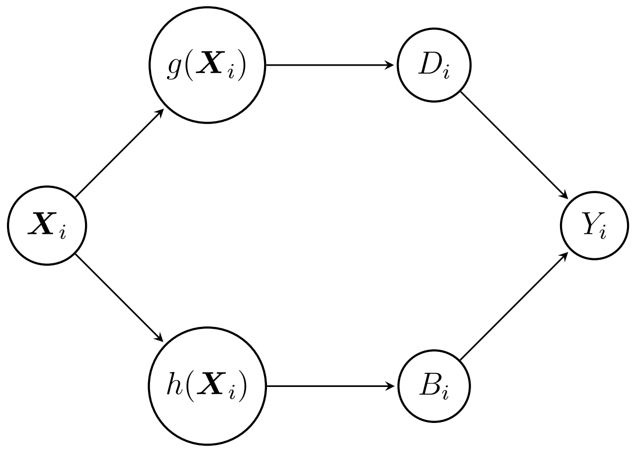
D.2 Research Design Differences

DZC primarily recommend the embedded natural experiment as a strategy to mitigate the effect of background variables. This is an important research strategy for addressing beliefs, but will be less effective with texts. Specifically, it is difficult to independently randomize the measured treatment and leave the unmeasured treatments in place. If the embedded natural experiment were successful, we would require a "no-interaction" assumption for this effect to correspond to the effects of interest in our paper. We might be worried, for example, that by providing information about a lottery or other natural experiment we have revealed other information that affects the results of our study. This could be through beliefs, or it could be because respondents are distracted by text length or because the randomization itself is a feature of the text that affects the outcome.

Our core research design principle is that experiments about text-based treatments should enable lots of different ways to say the same treatment, which enables us to adjust for potential unmeasured treatments ex post. This is similar in spirit to DZC's covariate control method, but differs in crucial ways. The covariate control

Figure 1: Alternative Directed Acyclic Graph for Causal Text Diagram

(a) Focusing on texts shows that the issues with text-based manipulations remain, even if issues in DZC are not present.



(b) Incorporating our framework into DZC's framework.

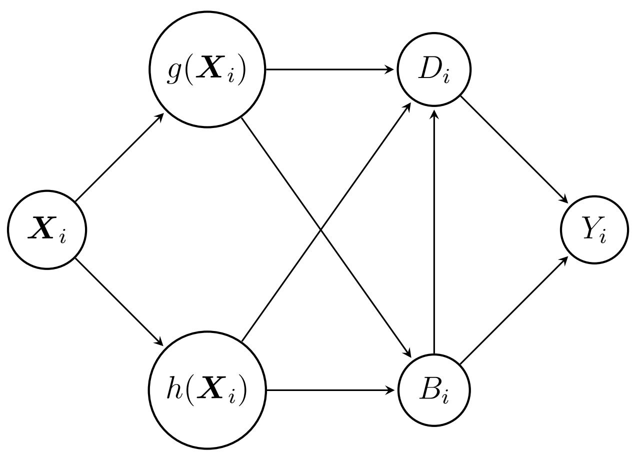
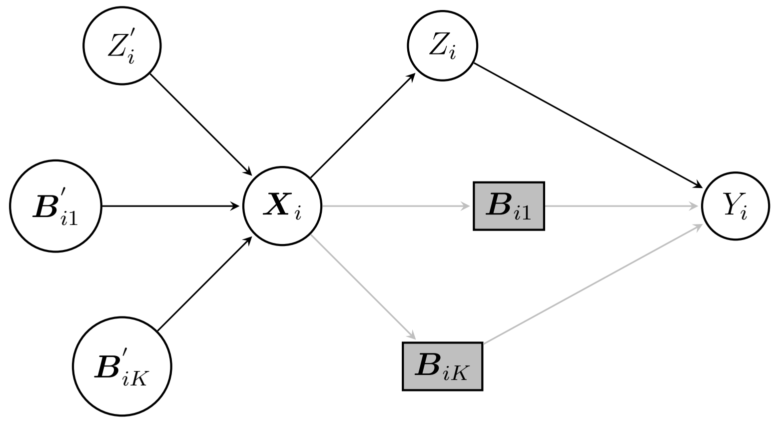


Figure 2: Alternative Directed Acyclic Graph for Causal Text Diagram



method in DZC encourages analysts to identify potential background beliefs and fix them as part of the experimental design ex ante. While our proposed research design accommodates researchers building the texts theemselves and deliberately varying possible unmeasured treatments, it also accommodates using texts that occur in nature to capture natural variation in the unmeasured treatments. In either case, our design retains the ability to explicitly measure and adjust for additional unmeasured treatments ex-post without running a new experiment.

E Alternative Representation of DAG

An alternative DAG can clarify the key issues with measured and unmeasured latent treatments. We construct that DAG in Figure 2 below. This makes clear the following important points from our model:

- The latent characteristics that drive the production of the model Z_i' and $\boldsymbol{B}_{i,1}', \ldots, \boldsymbol{B}_{i,K}'$ all produce the text. But once we observe the text, these characteristics are no longer necessary to be observed.

- The outcome Y_i depends on the measured treatment of interest Z_i and the unmeasured treatments $\boldsymbol{B}_{i,1}, \ldots, \boldsymbol{B}_{i,K}$. This means that if we observe the unmeasured treatments, then we are able to estimate Z_i directly.

- If we don't observe these unmeasured treatments, then we need some additional assumption to close this path. Our Assumption 4 addresses this either by assuming that the measured and unmeasured treatments are independent or that the unmeasured treatments don't have an effect on the outcome. This would correspond to the arrows from $\boldsymbol{B}_{i,1}, \ldots, \boldsymbol{B}_{i,K}$ to Y_i being absent from the diagram.

F Proof of Proposition 1

In this section we prove Proposition 1, which we restate below for clarity.

Proposition 1. *If Assumptions 1-4 hold, then Equation 2.1 is identified and can be estimated by taking the difference in means between individuals who received a document with the treatment and individuals who receive a document without the treatment.*

Proof. Invoking Assumption 1 to write potential outcomes as a function of only the text seen by i and invoking the common support component of Assumption 3, the desired equality for identification can be formally stated as

$$\mathbb{E}[Y_i(\boldsymbol{X}_i)|g(\boldsymbol{X}_i)=1] - \mathbb{E}[Y_i(\boldsymbol{X}_i)|g(\boldsymbol{X}_i)=0] = \tag{F.1}$$
$$\sum_{b\in\mathcal{B}}\{\mathbb{E}[Y_i(\boldsymbol{X}_i)|g(\boldsymbol{X}_i)=1,h(\boldsymbol{X}_i)=\boldsymbol{b}] - \mathbb{E}[Y_i(\boldsymbol{X}_i)|g(\boldsymbol{X}_i)=0,h(\boldsymbol{X}_i)=\boldsymbol{b}]\}Pr(\boldsymbol{B}_i=\boldsymbol{b})$$

$Pr(\boldsymbol{B}_i=\boldsymbol{b})$ is a shorthand for $\sum_{\boldsymbol{x}\in\mathcal{X}}\mathbf{1}\{h(\boldsymbol{x})=\boldsymbol{b}\}Pr(\boldsymbol{X}_i=\boldsymbol{x})$. By Assumption 3, the right hand side is interpretable as the causal effect of the latent treatment.

By Assumption 2, the left hand side of Equation F.1 (the difference in means) is identified by the observed difference in means between respondents who got a text with the treatment and respondents who got a text without the treatment. Expanding over $\mathcal{B}$, the left hand side is equal to

$$\sum_{\boldsymbol{b}\in\mathcal{B}}\mathbb{E}[Y_i(\boldsymbol{X}_i)|g(\boldsymbol{X}_i)=1,h(\boldsymbol{X}_i)=\boldsymbol{b}]Pr(h(\boldsymbol{X}_i)=\boldsymbol{b}|g(\boldsymbol{X}_i)=1)- \tag{F.2}$$
$$\mathbb{E}[Y_i(\boldsymbol{X}_i)|g(\boldsymbol{X}_i)=0,h(\boldsymbol{X}_i)=\boldsymbol{b}]Pr(h(\boldsymbol{X}_i)=\boldsymbol{b}|g(\boldsymbol{X}_i)=0)$$

$Pr(h(\boldsymbol{X}_i)=\boldsymbol{b}|g(\boldsymbol{X}_i)=z)$ is shorthand for $\frac{\sum_{\boldsymbol{x}\in\mathcal{X}}\mathbf{1}\{g(\boldsymbol{x})=z,h(\boldsymbol{x})=\boldsymbol{b}\}Pr(\boldsymbol{X}_i=\boldsymbol{x})}{\sum_{\boldsymbol{x}\in\mathcal{X}}\mathbf{1}\{g(\boldsymbol{x})=z\}Pr(\boldsymbol{X}_i=\boldsymbol{x})}$.
Substituting this into the left-hand side of Equation F.1 and rearranging yields

$$\sum_{\boldsymbol{b}\in\mathcal{B}}\mathbb{E}[Y_i(\boldsymbol{X}_i)|g(\boldsymbol{X}_i)=1,h(\boldsymbol{X}_i)=\boldsymbol{b}][Pr(\boldsymbol{B}_i=\boldsymbol{b})-Pr(\boldsymbol{B}_i=\boldsymbol{b}|Z_i=1)] = \tag{F.3}$$
$$\sum_{\boldsymbol{b}\in\mathcal{B}}\mathbb{E}[Y_i(\boldsymbol{X}_i)|g(\boldsymbol{X}_i)=0,h(\boldsymbol{X}_i)=\boldsymbol{b})][Pr(\boldsymbol{B}_i=\boldsymbol{b})-Pr(\boldsymbol{B}_i=\boldsymbol{b}|Z_i=0)]$$

Technically, identification requires only that Equation F.3 holds. Assumption 4 merely provides two sufficient conditions for Equation F.3. However, this minor tightening of the assumptions yields substantial gains in interpretability. Suppose, as in Assumption 4, $\boldsymbol{B}_i$ and Z_i are independent. Then $Pr(\boldsymbol{B}_i=\boldsymbol{b})-Pr(\boldsymbol{B}_i=\boldsymbol{b}|Z_i=z)=0$ for $z\in\{0,1\}$, so both sides of Equation F.3 are 0 and the equality holds. Alternatively, suppose $\mathbb{E}[Y_i(\boldsymbol{X}_i)|g(\boldsymbol{X}_i)=z,h(\boldsymbol{X}_i)=\boldsymbol{b}]=\mathbb{E}[Y_i(\boldsymbol{X}_i)|g(\boldsymbol{X}_i)=z,h(\boldsymbol{X}_i)=\boldsymbol{b}']$ for all $\boldsymbol{b},\boldsymbol{b}'\in\mathcal{B}$. Then in both summations, $\mathbb{E}[Y_i(\boldsymbol{X}_i)|g(\boldsymbol{X}_i)=z,h(\boldsymbol{X}_i)=\boldsymbol{b}]$ (for any arbitrary value of $\boldsymbol{b}$) 8 can be factored out, leaving a summation of $\sum_{\boldsymbol{b}\in B}Pr(\boldsymbol{B}_i=\boldsymbol{b})-Pr(\boldsymbol{B}_i=\boldsymbol{b}|Z_i=z)$ on each side. This summation is equal to 0, so here too the equality holds.

One advantage of Equation F.3 is that the difference between the left and right hand sides of the equation characterizes the size of the inconsistency. Thus, $\boldsymbol{B}_i$ has a small effect on the outcome or Z_i and $\boldsymbol{B}_i$ are nearly independent, the inconsistency is small.

It is easy to see that if there were some subset of the elements of $\boldsymbol{B}_i$ that had no effect on the outcome and the remainder of $\boldsymbol{B}_i$ were independent of Z_i, the proof would still follow. The proof would simply marginalize over the elements of $\boldsymbol{B}_i$ that had no effect on the outcome and then the invoke independence on the remainder of $\boldsymbol{B}_i$. We leave the statement of Assumption 4 as more restrictive than necessary because it more concisely conveys the intuition practitioners need to assess whether unmeasured treatments threaten inference. □

G Generalization to the AMCE

When we are interested in the effect of more than one measured latent treatment, our estimand is the average marginal component effect (AMCE). The AMCE for the k^{th} component is defined as

$$\begin{aligned}\text{AMCE}_k \quad = \quad & \sum_{\boldsymbol{z}_{-k} \in \mathcal{Z}_{-k}} \sum_{\boldsymbol{b} \in \mathcal{B}} \{\mathbb{E}[Y_i(Z_k = 1, \boldsymbol{Z}_{i,-k} = \boldsymbol{z}_{-k}, \boldsymbol{B}_i = \boldsymbol{b})] - \mathbb{E}[Y_i(Z_k = 0, \boldsymbol{Z}_{i,-k} = \boldsymbol{z}_{-k}, \boldsymbol{B}_i = \boldsymbol{b})] \\ & \times Pr(\boldsymbol{B}_i = \boldsymbol{b}) \times m(\boldsymbol{z}_{-k})\end{aligned}$$

where $-k$ is an index that indicates all but the kth components of $\boldsymbol{Z}_i$ and $\mathcal{Z}$. $m(\boldsymbol{z}_{-k})$ is an analyst-specified density on measured treatments besides the latent treatment of interest.

The proof from Appendix F generalizes to identifying the AMCE$_k$. For compactness, let $Z_{i,k}$ be the kth element of $g(\boldsymbol{X}_i)$ and let $\boldsymbol{Z}_{i,-k}$ be the other $k-1$ elements. The desired equality for identification is instead

$$\begin{aligned}& \sum_{\boldsymbol{z}_{-k} \in \mathcal{Z}_{-k}} \{\mathbb{E}[Y_i(\boldsymbol{X}_i) | Z_{i,k} = 1, \boldsymbol{Z}_{i,-k} = \boldsymbol{z}_{-k}] - \\ & \qquad \mathbb{E}[Y_i(\boldsymbol{X}_i) | Z_{i,k} = 0, \boldsymbol{Z}_{i,-k} = \boldsymbol{z}_{-k}]\} \times m(\boldsymbol{z}_{-k}) = \\ & \sum_{\boldsymbol{z}_{-k} \in \mathcal{Z}_{-k}} \sum_{\boldsymbol{b} \in \mathcal{B}} \{\mathbb{E}[Y_i(\boldsymbol{X}_i) | Z_{i,k} = 1, \boldsymbol{Z}_{i,-k} = \boldsymbol{z}_{-k}, h(\boldsymbol{X}_i) = \boldsymbol{b}] - \\ & \qquad \mathbb{E}[Y_i(\boldsymbol{X}_i) | Z_{i,k} = 0, \boldsymbol{Z}_{i,-k} = \boldsymbol{z}_{-k}, h(\boldsymbol{X}_i) = \boldsymbol{b}]\} \\ & \qquad \times Pr(\boldsymbol{B}_i = \boldsymbol{b}) m(\boldsymbol{z}_{-k})\end{aligned} \tag{G.1}$$

$m(\boldsymbol{z}_{-k})$ is an analyst-specified density, which for practical purposes is often $Pr(\boldsymbol{Z}_{-k} = \boldsymbol{z}_{-k})$.

Assumptions 1 and 2 can be used without modification, but the proof requires a modification to Assumptions 3 and 4.

Assumption 5. *There exists some function $h : \mathcal{X} \to \mathcal{B}$ such that if $g(\boldsymbol{x}) = g(\boldsymbol{x}')$ and $h(\boldsymbol{x}) = h(\boldsymbol{x}')$ for $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}$, then $\mathbb{E}[Y_i(\boldsymbol{X}_i = \boldsymbol{x})] = \mathbb{E}[Y_i(\boldsymbol{X}_i = \boldsymbol{x}')]$. Additionally, $0 < Pr(Z_{i,k} = 1 | \boldsymbol{Z}_{i,-k} = \boldsymbol{z}_{-k}, \boldsymbol{B}_i = \boldsymbol{b}) < 1$ for all $\boldsymbol{z}_{-k} \in \mathcal{Z}_{-k}$ such that $m(\boldsymbol{z}_{-k}) > 0$ and all $\boldsymbol{b} \in \mathcal{B}$.*

Assumption 6. *At least one of the following is true:*

- *The kth measured latent treatment and the unmeasured latent treatments are independent, given the other $k-1$ measured latent treatments:* $Pr(Z_{i,k} = z_k, \boldsymbol{B}_i = \boldsymbol{b} | \boldsymbol{Z}_{i,-k} = \boldsymbol{z}_{-k}) = Pr(Z_{i,k} = z_k | \boldsymbol{Z}_{i,-k} = \boldsymbol{z}_{-k}) Pr(\boldsymbol{B}_i = \boldsymbol{b} | \boldsymbol{Z}_{i,-k} = \boldsymbol{z}_{-k})$.
- *The unmeasured treatments are unrelated to the outcome:* $\mathbb{E}[Y_i(\boldsymbol{Z}_i, \boldsymbol{B}_i)] = \mathbb{E}[Y_i(\boldsymbol{Z}_i, \boldsymbol{B}'_i)]$ *for* $\boldsymbol{B}_i$ *and* $\boldsymbol{B}'_i$.

Additionally, the presence of multiple latent treatments raises the possibility

Proposition 2. *If Assumptions 1, 2, 5, and 6 hold, then the right hand side of Equation G.1 is identified by the left hand side of Equation G.1.*

Proof. The proof follows from the proof strategy in Appendix F. The left hand side of Equation G.1 can be rewritten as

$$\begin{aligned}\sum_{\boldsymbol{z}_{-k} \in \mathcal{Z}_{-k}} \sum_{b \in \mathcal{B}} \{ & \mathbb{E}[Y_i(\boldsymbol{X}_i) | Z_{i,k} = 1, \boldsymbol{Z}_{i,-k} = \boldsymbol{z}_{-k}, \boldsymbol{B}_i = \boldsymbol{b}] \\ & \times Pr(\boldsymbol{B}_i = b | Z_{i,k} = 1, \boldsymbol{Z}_{i,-k} = \boldsymbol{z}_{-k}) - \\ & \mathbb{E}[Y_i(\boldsymbol{X}_i) | Z_{i,k} = 0, \boldsymbol{Z}_{i,-k} = \boldsymbol{z}_{-k}, \boldsymbol{B}_i = \boldsymbol{b}] \\ & \times Pr(\boldsymbol{B}_i = b | Z_{i,k} = 0, \boldsymbol{Z}_{i,-k} = \boldsymbol{z}_{-k}) \} \times m(\boldsymbol{z}_{-k})\end{aligned} \tag{G.2}$$

Substituting this into the left-hand side of Equation G.1 yields

$$\begin{aligned}& \sum_{\boldsymbol{z}_{-k} \in \mathcal{Z}_{-k}} m(\boldsymbol{z}_{-k}) \sum_{\boldsymbol{b} \in \mathcal{B}} \mathbb{E}[Y_i(\boldsymbol{X}_i) | Z_{i,k} = 1, \boldsymbol{Z}_{i,-k} = \boldsymbol{z}_{-k}, h(\boldsymbol{X}_i) = \boldsymbol{b}] \\ & \qquad \times [Pr(\boldsymbol{B}_i = \boldsymbol{b} | \boldsymbol{Z}_{i,-k} = \boldsymbol{z}_{-k}) - Pr(\boldsymbol{B}_i = \boldsymbol{b} | Z_{i,k} = 1, \boldsymbol{Z}_{i,-k} = \boldsymbol{z}_{-k})] = \\ & \sum_{\boldsymbol{z}_{-k} \in \mathcal{Z}_{-k}} m(\boldsymbol{z}_{-k}) \sum_{\boldsymbol{b} \in \mathcal{B}} \mathbb{E}[Y_i(\boldsymbol{X}_i) | Z_{i,k} = 0, \boldsymbol{Z}_{i,-k} = \boldsymbol{z}_{-k}, h(\boldsymbol{X}_i) = \boldsymbol{b})] \\ & \qquad \times [Pr(\boldsymbol{B}_i = \boldsymbol{b} | \boldsymbol{Z}_{i,-k} = \boldsymbol{z}_{-k}) - Pr(\boldsymbol{B}_i = \boldsymbol{b} | Z_{i,k} = 0, \boldsymbol{Z}_{i,-k} = \boldsymbol{z}_{-k})]\end{aligned} \tag{G.3}$$

The proposition requires only this equality, but Assumption 6 provides intuitive sufficient conditions. □

This generalization to the AMCE highlights a path forward in instances in which the analyst seeks the ATE but is concerned about an unmeasured confounder that is correlated with both the outcome and the treatment of interest. This proof of the AMCE allows for $\boldsymbol{Z}_{i,-k}$ to be correlated with $Z_{i,k}$.

It is important to note that even if $Z_{i,k}$ is independent of $\boldsymbol{B}_i$, there is no guarantee that $Z_{i,k}$ is independent of $\boldsymbol{B}_i$ given $\boldsymbol{Z}_{i,-k}$. Verbally, if there is only one unmeasured treatment that is correlated with both the treatment of interest and the outcome, there is no guarantee that measuring and controlling for that unmeasured latent treatment will lead to identification, because there might be another unmeasured treatment that is correlated with both the previously unmeasured treatment and the outcome. This could lead to a violation of the conditional independence assumption required for identification. Even so, we recommend explicitly measuring and adjusting for any treatment that seems to be correlated with both the outcome and the treatment of interest, because the estimator is guaranteed to be inconsistent if you do not and merely risks being inconsistent if you do.

H Hong Kong Message Experiment

To construct the arguments for the Hong Kong survey experiment, we collected all congressional floor speeches made between the start of the protests (March, 2019) and when we launched the first iteration of the experiment (November, 2019). This yielded 16 speeches from the both chambers and both parties. The authors then read these floor speeches to identify common arguments made in support of the protesters and different ways the arguments were presented. This process generated the seven latent treatments described in the main text: commitment, bravery, mistreatment, flags, threat, economy, and violation. The experiment includes many different versions of each text.

The commitment texts vary (1) whether the law is described merely as a bill or the Hong Kong Policy Act, (2) the language and verb tense used to describe the commitment, (3) the presentation of the timing of the bill's passage, i.e. whether it was in 1992, 27 years ago, during George H.W. Bush's Administration, or some time ago, and (4) whether the commitment protects Hong Kong's freedom, autonomy, right to govern itself, or some combination thereof.

The bravery texts vary (1) whether the protester's willingness to be beaten, imprisoned, or tortured is attributable to a desire to be free, (2) whether the protesters are described as the Hong Kong protesters or merely as some protesters, (3) whether their dangers are described as being beaten, imprisoned, or tortured or as being targeted by brutal police tactics and intimidation, and (4) whether their actions are explicitly described as brave.

The mistreatment texts vary (1) whether China is described as authoritarian, totalitarian, communist, or just as China, (2) whether China is described as having a long history of mistreating its citizens or more specifically described as not allowing free speech, free press, or the due process of law, (3) whether China is explicitly described as attempting to destroy freedom in Hong Kong, (4) whether the Tiananmen Square Massacre in 1989 is recounted, (5) whether the text alleges that 18 million Chinese citizens have been killed by Chinese government policy, (6) whether Chinese mistreatment of the Uyghurs is detailed.

The flags texts vary (1) whether the text praises the flag as a symbol for freedom everywhere, (2) whether waving flags is attributed to a desire to emulate the American Revolution, and (3) whether waving flags is described as affirming a commitment to American values.

The threat texts vary (1) whether the text mentions possible future aggression to Japan or Taiwan, (2) whether the text describes efforts to catch up to American military capabilities, (3) whether the text mentions purchasing arms from Russia, and (4) whether the text discusses the growth of the Chinese aircraft carrier fleet.

The economy texts vary whether (1) respondents are told that China has been governed separately from China for the last 150 years, (2) the use of the term free-market, (3) whether Hong Kong is described as having a special status within China, (4) whether Hong Kong is described as one of the richest cities in the world, and (5) whether Hong Kong's autonomy is described as self-government.

The violation texts vary whether (1) the date in which Britain returned Hong Kong to China is mentioned, (2) China's promise not to interfere in Hong Kong's affairs for 50 years is mentioned, (3) the Sino-British Joint Declaration is mentioned by name, and (4) the texts says that protesters allege China's actions violate the Sino-British Joint Declaration.

I Trump Message Experiment

I.1 Example Tweets

Treatment 1

- Any negative polls are fake news, just like the CNN, ABC, NBC polls in the election. Sorry, people want border security and extreme vetting.
- The FAKE NEWS media (failing @nytimes, @NBCNews, @ABC, @CBS, @CNN) is not my enemy, it is the enemy of the American People!
- Crooked Hillary Clinton now blames everybody but herself, refuses to say she was a terrible candidate. Hits Facebook & even Dems & DNC.

Treatment 2

- Congratulations to Roy Moore on his Republican Primary win in Alabama. Luther Strange started way back & ran a good race. Roy, WIN in Dec!
- The approval process for the biggest Tax Cut & Tax Reform package in the history of our country will soon begin. Move fast Congress!
- Since November 8th, Election Day, the Stock Market has posted $3.2 trillion in GAINS and consumer confidence is at a 15 year high. Jobs!

Treatment 3

- Insurance companies are fleeing ObamaCare - it is dead. Our healthcare plan will lower premiums & deductibles - and be great healthcare!
- The Republican Senators must step up to the plate and, after 7 years, vote to Repeal and Replace. Next, Tax Reform and Infrastructure. WIN!
- The Democrats will only vote for Tax Increases. Hopefully, all Senate Republicans will vote for the largest Tax Cuts in U.S. history.

Treatment 4

- Why is the NFL getting massive tax breaks while at the same time disrespecting our Anthem, Flag and Country? Change tax law!
- Today we remember the men and women who made the ultimate sacrifice in serving. Thank you, God bless your families & God bless the USA!
- Sports fans should never condone players that do not stand proud for their National Anthem or their Country. NFL should change policy!

Treatment 5

- Just completed call with President Moon of South Korea. Very happy and impressed with 15-0 United Nations vote on North Korea sanctions.
- Since November 8th, Election Day, the Stock Market has posted $3.2 trillion in GAINS and consumer confidence is at a 15 year high. Jobs!
- Stock Market at all time high, unemployment at lowest level in years (wages will start going up) and our base has never been stronger!

No Treatments

- Today I will meet with Canadian PM Trudeau and a group of leading business women to discuss women in the workforce. https://t.co/bFAHPRXHdP
- Today I signed the Veterans (OUR HEROES) Choice Program Extension & Improvement Act @ the @WhiteHouse. S544 Watch 45.wh.gov/7x5n53
- Honored to host a luncheon for African leaders this afternoon. Great discussions on the challenges & opportunities facing our nations today.
- Melania and I were thrilled to join the dedicated men and women of the @USEmbassyFrance, members of the U.S. Military and their families.
- Such an honor to have my good friend, Israel PM @Netanyahu, join us w/ his delegation in NYC this afternoon. UNGA

I.2 Correlation of Treatments

Table 3 shows that there are relatively low levels of correlations between the latent Trump tweets. This suggests that little is lost by not including interactions between the treatments in the linear regression used to estimate the AMCEs. Moreover, an F-test that nests the baseline specification inside a model that includes all first order interactions between the treatments fails to reject the null hypothesis that the baseline model fits the data as well as the model with first-order interactions, yielding p-values of 0.78 for Republicans, 0.12 for Democrats, and 0.49 for independents.

Table 3: Correlation of Trump Tweet Treatments

	Z1	Z2	Z3	Z4	Z5
Z1	1.00	0.10	0.13	-0.03	0.04
Z2	0.10	1.00	0.12	-0.03	0.22
Z3	0.13	0.12	1.00	0.12	0.23
Z4	-0.03	-0.03	0.12	1.00	0.06
Z5	0.04	0.22	0.23	0.06	1.00

I.3 First-Order Interaction Regressions

Table 4 presents a linear regression of the Trump texts with first-order interactions between the treatments. The negative effects of Treatments 1 and 3 are robust to the inclusion of first-order interactions. The positive effects of Treatments 4 and 5 remain, although they do not reach statistical significance for all specifications.

Table 4: Trump Tweets Results with First-Order Interactions

	Democrats	Independents	Republicans
(Intercept)	-76.04^*	2.67^*	98.79^*
	(1.69)	(1.34)	(1.08)
Z1	-53.35^*	-37.22^*	-21.42^*
	(6.16)	(4.87)	(3.92)
Z2	−4.87	2.33	9.52
	(15.44)	(12.21)	(9.81)
Z3	-41.53^*	-28.68^*	-10.20^*
	(6.49)	(5.13)	(4.12)
Z4	21.31	22.58^*	12.98
	(10.98)	(8.69)	(6.98)
Z5	37.91^*	23.41^*	0.22
	(13.64)	(10.79)	(8.67)
Z1:Z2	15.42	15.11	7.16
	(43.84)	(34.68)	(27.86)
Z1:Z3	33.88^*	19.90	5.01
	(13.42)	(10.44)	(8.53)
Z1:Z4	−5.46	4.61	−18.98
	(45.01)	(27.35)	(28.61)
Z1:Z5	28.09	11.92	14.60
	(55.11)	(43.60)	(35.03)
Z2:Z3	37.46	19.58	4.32
	(21.99)	(17.40)	(13.98)
Z2:Z5	−17.48	−13.32	10.08
	(35.29)	(27.92)	(22.43)
Z3:Z4	−18.74	−8.55	6.23
	(19.88)	(15.73)	(12.64)
Z3:Z5	−5.52	−7.73	4.35
	(18.12)	(14.33)	(11.52)
Z4:Z5	23.45	15.04	26.65
	(29.21)	(23.10)	(18.57)
N	752	752	752

Standard errors in parentheses

* indicates significance at $p < 0.05$

References

Anoll, Allison P. 2018. "What makes a good neighbor? Race, place, and norms of political participation." *American Political Science Review* 112(3):494–508.

Baker, Andy. 2015. "Race, paternalism, and foreign aid: Evidence from US public opinion." *American Political Science Review* 109(1):93–109.

Balmas, Meital. 2018. "Tell me who is your leader, and I will tell you who you are: Foreign leaders' perceived personality and public attitudes toward their countries and citizenry." *American Journal of Political Science* 62(2):499–514.

Barber, Michael and Jeremy C Pope. 2019. "Does party trump ideology? Disentangling party and ideology in America." *American Political Science Review* 113(1):38–54.

Bisgaard, Martin. 2019. "How Getting the Facts Right Can Fuel Partisan-Motivated Reasoning." *American Journal of Political Science* .

Bisgaard, Martin and Rune Slothuus. 2018. "Partisan elites as culprits? How party cues shape partisan perceptual gaps." *American Journal of Political Science* 62(2):456–469.

Bloom, Pazit Ben-Nun, Gizem Arikan and Marie Courtemanche. 2015. "Religious social identity, religious belief, and anti-immigration sentiment." *American Political Science Review* 109(2):203–221.

Brierley, Sarah, Eric Kramon and George Kwaku Ofosu. 2019. "The moderating effect of debates on political attitudes." *American Journal of Political Science* .

Broockman, David E, Gregory Ferenstein and Neil Malhotra. 2019. "Predispositions and the Political Behavior of American Economic Elites: Evidence from Technology Entrepreneurs." *American Journal of Political Science* 63(1):212–233.

Butler, Daniel M and Hans JG Hassell. 2018. "On the limits of officials' ability to change citizens' priorities: A field experiment in local politics." *American Political Science Review* 112(4):860–873.

Carlson, Taylor N. 2019. "Through the Grapevine: Informational Consequences of Interpersonal Political Communication." *American Political Science Review* 113(2):325–339.

Carnes, Nicholas and Noam Lupu. 2016. "Do voters dislike working-class candidates? Voter biases and the descriptive underrepresentation of the working class." *American Political Science Review* 110(4):832–844.

Christenson, Dino P and Douglas L Kriner. 2017. "Constitutional qualms or politics as usual? The factors shaping public support for unilateral action." *American Journal of Political Science* 61(2):335–349.

Clayton, Amanda, Diana Z O'Brien and Jennifer M Piscopo. 2019. "All male panels? Representation and democratic legitimacy." *American Journal of Political Science* 63(1):113–129.

Clifford, Scott and Jennifer Jerit. 2018. "Disgust, anxiety, and political learning in the face of threat." *American Journal of Political Science* 62(2):266–279.

Corbacho, Ana, Daniel W Gingerich, Virginia Oliveros and Mauricio Ruiz-Vega. 2016. "Corruption as a Self-Fulfilling Prophecy: Evidence from a Survey Experiment in Costa Rica." *American Journal of Political Science* 60(4):1077–1092.

Corstange, Daniel and Erin A York. 2018. "Sectarian framing in the syrian civil war." *American Journal of Political Science* 62(2):441–455.

De Benedictis-Kessner, Justin, Matthew A Baum, Adam J Berinsky and Teppei Yamamoto. 2019. "Persuading the enemy: Estimating the persuasive effects of partisan media with the preference-incorporating choice and assignment design." *American Political Science Review* 113(4):902–916.

Driscoll, Amanda, Gabriel Cepaluni, Feliciano de Sá Guimarães and Paolo Spada. 2018. "Prejudice, Strategic Discrimination, and the Electoral Connection: Evidence from a Pair of Field Experiments in Brazil." *American Journal of Political Science* 62(4):781–795.

Druckman, James N, Matthew S Levendusky and Audrey McLain. 2018. "No need to watch: How the effects of partisan media can spread via interpersonal discussions." *American Journal of Political Science* 62(1):99–112.

Einstein, Katherine Levine and David M Glick. 2017. "Does race affect access to government services? An experiment exploring street-level bureaucrats and access to public housing." *American Journal of Political Science* 61(1):100–116.

Enos, Ryan D and Noam Gidron. 2018. "Exclusion and Cooperation in Diverse Societies: Experimental Evidence from Israel." *American Political Science Review* 112(4):742–757.

Foos, Florian and Eline A de Rooij. 2017. "All in the Family: Partisan Disagreement and Electoral Mobilization in Intimate Networks—A Spillover Experiment." *American Journal of Political Science* 61(2):289–304.

Gaikwad, Nikhar and Gareth Nellis. 2017. "The Majority-Minority Divide in Attitudes toward Internal Migration: Evidence from Mumbai." *American Journal of Political Science* 61(2):456–472.

Grossman, Guy and Kristin Michelitch. 2018. "Information dissemination, competitive pressure, and politician performance between elections: A field experiment in Uganda." *American Political Science Review* 112(2):280–301.

Han, Hahrie. 2016. "The organizational roots of political activism: Field experiments on creating a relational context." *American Political Science Review* 110(2):296–307.

Hankinson, Michael. 2018. “When do renters behave like homeowners? High rent, price anxiety, and NIMBYism.” *American Political Science Review* 112(3):473–493.

Healy, Andrew, Katrina Kosec and Cecilia Hyunjung Mo. 2017. “Economic development, mobility, and political discontent: An experimental test of tocqueville’s thesis in Pakistan.” *American Political Science Review* 111(3):605–621.

Hemker, Johannes and Anselm Rink. 2017. “Multiple dimensions of bureaucratic discrimination: Evidence from German welfare offices.” *American Journal of Political Science* 61(4):786–803.

Hill, Seth J and Gregory A Huber. 2019. “On the Meaning of Survey Reports of Roll-Call “Votes”.” *American Journal of Political Science* .

Huddy, Leonie, Lilliana Mason and Lene Aarøe. 2015. “Expressive partisanship: Campaign involvement, political emotion, and partisan identity.” *American Political Science Review* 109(1):1–17.

Huff, Connor and Joshua D Kertzer. 2018. “How the public defines terrorism.” *American Journal of Political Science* 62(1):55–71.

Imai, Kosuke, Luke Keele, Dustin Tingley and Teppei Yamamoto. 2011. “Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies.” *American Political Science Review* 105(4):765–789.

Kalla, Joshua L and David E Broockman. 2018. “The minimal persuasive effects of campaign contact in general elections: Evidence from 49 field experiments.” *American Political Science Review* 112(1):148–166.

Karpowitz, Christopher F, J Quin Monson and Jessica Robinson Preece. 2017. “How to elect more women: Gender and candidate success in a field experiment.” *American Journal of Political Science* 61(4):927–943.

Kertzer, Joshua D and Thomas Zeitzoff. 2017. “A bottom-up theory of public opinion about foreign policy.” *American Journal of Political Science* 61(3):543–558.

Klar, Samara. 2018. “When common identities decrease trust: An experimental study of Partisan women.” *American Journal of Political Science* 62(3):610–622.

Kosmidis, Spyros. 2018. “International constraints and electoral decisions: does the room to maneuver attenuate economic voting?” *American Journal of Political Science* 62(3):519–534.

Li, Zhao. 2018. “How Internal Constraints Shape Interest Group Activities: Evidence from Access-Seeking PACs.” *American Political Science Review* 112(4):792–808.

Mattes, Michaela and Jessica LP Weeks. 2019. “Hawks, Doves, and Peace: An Experimental Approach.” *American Journal of Political Science* 63(1):53–66.

McConnell, Christopher, Yotam Margalit, Neil Malhotra and Matthew Levendusky. 2018. "The economic consequences of partisanship in a polarized era." *American Journal of Political Science* 62(1):5–18.

McEntire, Kyla Jo, Michele Leiby and Matthew Krain. 2015. "Human rights organizations as agents of change: An experimental examination of framing and micromobilization." *American Political Science Review* 109(3):407–426.

Mummolo, Jonathan and Erik Peterson. 2019. "Demand effects in survey experiments: An empirical assessment." *American Political Science Review* 113(2):517–529.

Pérez, Efrén O and Margit Tavits. 2017. "Language Shapes People's Time Perspective and Support for Future-Oriented Policies." *American Journal of Political Science* 61(3):715–727.

Reeves, Andrew and Jon C Rogowski. 2018. "The Public Cost of Unilateral Action." *American Journal of Political Science* 62(2):424–440.

Ryan, Timothy J. 2017. "How do indifferent voters decide? The political importance of implicit attitudes." *American Journal of Political Science* 61(4):892–907.

Sheffer, Lior, Peter John Loewen, Stuart Soroka, Stefaan Walgrave and Tamir Sheafer. 2018. "Nonrepresentative representatives: an experimental study of the decision making of elected politicians." *American Political Science Review* 112(2):302–321.

Simonovits, Gábor, Gabor Kezdi and Peter Kardos. 2018. "Seeing the world through the other's eye: An online intervention reducing ethnic prejudice." *American Political Science Review* 112(1):186–193.

Teele, Dawn Langan, Joshua Kalla and Frances Rosenbluth. 2018. "The Ties That Double Bind: Social Roles and Women's Underrepresentation in Politics." *American Political Science Review* 112(3):525–541.

Thachil, Tariq. 2017. "Do rural migrants divide ethnically in the city? Evidence from an ethnographic experiment in India." *American Journal of Political Science* 61(4):908–926.

Valenzuela, Ali A and Melissa R Michelson. 2016. "Turnout, status, and identity: Mobilizing Latinos to vote with group appeals." *American Political Science Review* 110(4):615–630.

Ward, Dalston G. 2019. "Public attitudes toward young immigrant men." *American Political Science Review* 113(1):264–269.