The Persistence of Survey Experimental Treatment Effects

Alexander Coppock*

February 21, 2017

Abstract

Do treatments deployed in survey experiments have persistent effects? The prototypical survey experiment collects background information, delivers experimental treatments, and records outcomes all in the space of a few minutes. Survey experimental treatments may cause real, underlying attitude change, or they may simply induce a temporary increase in subjects' probability of choosing one response over another. Further, some treatments may have more persistent effects than others. I hypothesize that priming and framing treatments that make certain considerations more accessible have fleeting effects whereas treatments that make considerations newly applicable or provide new information persist longer. Results from 18 panel survey experiments show that on average, survey experimental treatment effects persist after 10 days, albeit at approximately half their original magnitudes. This suite of evidence also provides strong support for the theory that information treatments have more persistent effects than framing treatments.

Word Count: 5,981

*Alexander Coppock is Assistant Professor, Department of Political Science, Yale University. Portions of this research were funded by the Time-sharing Experiments in the Social Sciences (TESS) organization. This research was reviewed and approved by the Institutional Review Board of Columbia University (IRB-AAAP1312). I am grateful to James N. Druckman and Donald P. Green for their guidance and support.

A common criticism of survey experiments is that they uncover real but fleeting effects. If so, the causal relationships studied in survey experiments may be little more than laboratory curiosities. In the words of Gaines, Kuklinski and Quirk (2007, p. 6) "The implications of survey-experimental results for politics depend crucially on how long the effects last, with relevant periods measured in weeks, or months, not minutes."

Why might survey experimental treatment effects fade? One hypothesis holds that large initial effects are due to experimenter demand: subjects answer how they think researchers want immediately after treatment, but do not face such pressure after some time has passed. Another view is that subjects may simply forget the information provided to them in a persuasive treatment. If it is the information itself that is responsible for short-term shifts in attitudes, the effects should dissipate as the memory of the information becomes more distant. Framing treatments in particular are hypothesized to have fleeting effects because subjects may naturally encounter counterframes in the days and weeks after treatment.

Understanding the persistence of persuasive effects is especially important in an era of long electoral campaigns. If the effects of even very persuasive messages evaporate within a few hours, then campaigns would be best served by not spending large amounts of money on early advertisements, only to see those effects disappear by election day. If, however, persuasive information can induce durable attitude change, the cumulative effects of two years' campaigning might have a substantial impact on election results.

Previous work has found mixed evidence on persistence. Those who find little or no persistence include de Vreese (2004), who showed that subjects exposed to "strategic" news about the enlargement of the EU reported higher (0.44 scale points, 5-point scale) political cynicism than subjects who saw news focused on the issues surrounding enlargement. This difference was almost non-existent in a follow up conducted one week later (0.02 scale points). Similarly, Druckman and Nelson (2003) find that a special-interests frame increases support for a bill by 0.71 scale points (7-point scale); 10 days later this effect drops[1] to 0.41. Mutz and Reeves (2005) report that videos featuring uncivil discourse decreased political trust by 0.44 standard deviations relative to civil discourse videos but that the difference (not reported) was no longer statistically significant approximately one month after treatment.

[1] Approximately 50% of the sample completed the follow-up, leading to a loss of power to detect treatment effects. The follow-up estimate is not significantly different from zero – or from the immediate estimate of 0.71.

Some survey experiments have found strong evidence of persistence: Tewksbury et al. (2000) report that exposure to a pro-regulation news story increased support for the regulation of hog farms by 24.1 percentage points relative to an anti-regulation story; this difference remained at 25.2 percentage points three weeks later. Lecheler and de Vreese (2011) show that a positive economic-benefit frame increased support for the inclusion of Bulgaria and Romania in the EU by 1.1 scale points (7-point scale); this effect was 0.93 a day later, 1.35 a week later, and 0.81 after two weeks.

These divergent results have led scholars to pose more nuanced theoretical questions. Instead of asking if survey experimental treatment effects persist in general, they ask, "Which treatments exhibit stronger or weaker persistence?" Drawing on theories of framing, priming, and persuasion, Baden and Lecheler (2012) hypothesize that three features of an experimental treatment in particular will influence its durability. Some primes and cues operate by making subjects' preexisting knowledge more *accessible*. Treatments that operate primarily though this channel are hypothesized to have fleeting effects. Other treatments may operate by making preexisting knowledge newly *applicable*. For example, drawing a connection between a policy proposal and a social movement might induce subjects to consider how the movement's principles are applicable in the policy context. Such treatments may persist longer. Finally, some treatments supply subjects with *new information*, possibly leading to enduring belief change. Most treatments are likely to be a bundle of all three processes – the idiosyncratic mix at work in each treatment may influence durability.

A major complication in the study of persistence is that stronger arguments may be associated both with larger initial effects *and* more enduring effects. A strong argument might induce a large the initial attitude change as well as increasing certainty with which subjects hold those new attitudes. If this pattern holds, we would expect to observe a positive correlation between the size of effects and their durability. However, a far less elaborate theory of persistence could account for why some effects endure and others fade: large effects last and small effects do not. In order distinguish between these possibilities, a body of evidence with large effects generated by both strong and weak arguments is required.

The remainder of this article will proceed as follows. First, I will describe the theoretical framework alluded to above in greater detail. I will then describe how the set of treatments used to explore durability were selected, the panel experimental design, and the statistical model of persistence used here. I will then present results and interpret them in light of the theoretical framework.

1 Theory

The expectancy value model of attitudes (Ajzen and Fishbein, 1980; Nelson, Oxley and Clawson, 1997; Zaller, 1992; Chong and Druckman, 2012) posits that subjects have a set of considerations to which subjective weights are attached. When prompted by a survey question, subjects take a weighted average of their considerations and return a response. This framework can accommodate many of the types of treatments that are commonly studied in survey experiments. Emphasis frames (Druckman, 2001) are hypothesized to affect responses by changing the weights, not the considerations themselves. Information treatments, by contrast, change responses by furnishing subjects with new considerations to take account of when offering a survey response. The manner in which survey responses vary with framing or information treatments can therefore be modeled as:

$$Y(Z) \sim f(\mathbf{W}, \mathbf{C}, Z, \sigma_Y, \ldots),$$

where Y is the survey response, $\mathbf{W}$ is a vector of weights attached to a vector of considerations $\mathbf{C}$, Z is an indicator for the treatment condition, and σ_Y is a measure of prior uncertainty that subjects have about the issue area. This description of how responses might change is very general and includes "..." to indicate that many other factors influence attitudes. This model allows for complex interactions between Z, $\mathbf{W}$, and $\mathbf{C}$. Unfortunately the flexibility of this model renders it difficult to work with when making predictions about the relative durability of different treatments. I will turn instead to a much simplified version of the general model above:

$$Y(Z) = \sum_{1}^{j} w_j(Z) \cdot c_j,$$

where Y is a simple weighted average of considerations. The weights w_j are themselves responsive to Z, indicating that the weights attached to considerations might differ depending on the treatment condition. In this model, there is no requirement that the weights w_j sum to 1. This choice reflects the theoretical possibility that the treatments themselves change the amount of "attention" that a subject gives to a survey response. If a respondent has been asked to carefully weigh a series of arguments for and against a proposition, the total amount of attention (weight) given to that response may be higher than if the subject were solicited for a response without any additional context.

To illustrate how this model helps draw distinctions among treatments, suppose that all subjects' survey responses are a function of just three considerations, c_A, c_B, and c_C, and that the treatment indication Z can take three values: control, an emphasis frame or an information treatment.

$$\begin{aligned}
\text{Response under control: } Y(Z=0) &= 0.5 \cdot c_A + 0.2 \cdot c_B \\
\text{Response under emphasis frame: } Y(Z=1) &= 0.8 \cdot c_A + 0.2 \cdot c_B \\
\text{Response under information treatment: } Y(Z=2) &= 0.5 \cdot c_A + 0.2 \cdot c_B + 0.3 \cdot c_C
\end{aligned}$$

The baseline (control) responses are a weighted average of considerations c_A and c_B, where the weights are 0.5 and 0.2. The responses under an emphasis are also a weighted average of these two considerations, but c_A is given a new weight, 0.8. Under an information treatment, the weights on c_A and c_B remain unchanged, but a new consideration, c_C, with an associated weight of 0.3 is added.

With this notation in hand, we can define two estimands, the average treatment effect framing and the average treatment effect of information. In this context, the expectation operator $\mathrm{E}[\cdot]$ denotes the true sample mean.

$$\begin{aligned}
\text{Average Framing Effect} &= \mathrm{E}\left[Y(Z=1) - Y(Z=0)\right] \\
&= (0.8 * c_A + 0.2 * c_B) - (0.5 * c_A + 0.2 * c_B) \\
&= 0.3 * c_A
\end{aligned}$$

$$\begin{aligned}
\text{Average Information Effect} &= \mathrm{E}\left[Y(Z=2) - Y(Z=0)\right] \\
&= (0.5 * c_A + 0.2 * c_B + 0.3 * c_C) - (0.5 * c_A + 0.2 * c_B) \\
&= 0.3 * c_C
\end{aligned}$$

The question under study here is, to what extent do these effects endure? In order to answer this question, we must add a time dimension to our model. An attitude measured at time t can be written:

$$Y_t(Z) = \sum_{1}^{j} w_{j,t}(Z) \cdot c_j$$

In this model, the weights are subscripted by consideration j and time t, and may be different depending on treatment status (Z). Using this setup, I define the average persistence of an effect as the ratio of the ATE at time 2 to the ATE at time 1:

$$\frac{\text{ATE Time 2}}{\text{ATE Time 1}} = \frac{\text{E}\left[Y_{t=2}(Z=1) - Y_{t=2}(Z=0)\right]}{\text{E}\left[Y_{t=1}(Z=1) - Y_{t=1}(Z=0)\right]}$$

It should be emphasized that while this estimand is well-defined, it is not equal to the average of all individual-level ratios of time 1 effect to the time 2 effect because the expectation of a ratio is not, in general, equal to the ratio of expectations. This non-equality is due to the possible covariance between individuals' treatment effects at time 1 with their treatment effect at time 2. I recognize that this definition introduces a theoretical inconsistency: the psychological processes hypothesized to be responsible for persistence take place at the individual level but the persistence estimand is defined at the group level. Ultimately, this inconsistency is unavoidable because of the fundamental problem of causal inference: we can no more identify the covariance between effects at the individual level than we can identify the effects themselves.

The average persistence estimand would be misleading in some scenarios. Imagine, for example, that half the sample has a treatment effect of 1 at time 1 but 0 at time 2; the other half has treatment effects of 0 at time 1 but 1 at time 2. For individuals in the first group, the true persistence is $\frac{0}{1} = 0$; for individuals in the second group, the true persistence is undefined: $\frac{1}{0}$. The average of these individual level persistences is likewise undefined. The average persistence, however, is $\frac{0.5}{0.5} = 1$. By contrast, if treatment effects were equal to 1 for the whole sample at time 1 and equal to 0.5 at time 2, then the average of each individual's effect ratio would be equal to the ratio of the average effects: $\frac{1}{N}\sum_1^N \frac{0.5}{1} = \frac{1}{N}\sum_1^N 0.5 / \frac{1}{N}\sum_1^N 1 = 0.5$.

With this caveat in mind, I define the following persistence estimands:

$$
\begin{aligned}
\text{Framing Effect Persistence} &= \frac{\mathrm{E}\left[Y_{t=2}(Z=1) - Y_{t=2}(Z=0)\right]}{\mathrm{E}\left[Y_{t=1}(Z=1) - Y_{t=1}(Z=0)\right]} \\
&= \frac{(0.6 * c_A + 0.2 * c_B) - (0.5 * c_A + 0.2 * c_B)}{(0.8 * c_A + 0.2 * c_B) - (0.5 * c_A + 0.2 * c_B)} \\
&= \frac{0.1 * c_A}{0.3 * c_A} \\
&= \frac{1}{3}
\end{aligned}
$$

$$
\begin{aligned}
\text{Information Effect Persistence} &= \frac{\mathrm{E}\left[Y_{t=2}(Z=2) - Y_{t=2}(Z=0)\right]}{\mathrm{E}\left[Y_{t=1}(Z=2) - Y_{t=1}(Z=0)\right]} \\
&= \frac{(0.5 * c_A + 0.2 * c_B + 0.2 * c_C) - (0.5 * c_A + 0.2 * c_B)}{(0.5 * c_A + 0.2 * c_B + 0.3 * c_C) - (0.5 * c_A + 0.5 * c_B)} \\
&= \frac{0.3 * c_C}{0.5 * c_C} \\
&= \frac{2}{3}
\end{aligned}
$$

It is no accident that in the foregoing toy example, the information effect showed stronger persistence than the framing effect. The assumption that changes in weights are less durable than the addition of new considerations is central to the theory of treatment effect persistence employed here.

This model can accommodate the theoretical predictions of treatment effect persistence put forth by Baden and Lecheler (2012). *Accessibility* treatments are hypothesized to operate primarily by increasing the weight given to a particular consideration. Because such treatments only affect outcomes through the weighting scheme, they are hypothesized to be fleeting. As an example of an accessibility treatment, consider the study described in Transue (2007), replicated as Study 4 below. In that experiment, subjects were randomly exposed to a treatment question that asked subjects: “How close do you feel to [your ethnic or racial group]/[other Americans]?” The dependent variable was subjects’ attitudes toward raising taxes for education; the treatment operates by increasing the weight given to ethnic identity or superordinate identity when answering this question.

Applicability frames also operate by changing the weights given to considerations, but do so in a manner distinct from accessibility frames. In this model, subjects hold a set of considerations in mind but the associated weights are attitude-specific. That is, consideration c_A may be given one weight when a subject is offering an attitude on gun control but a different weight when responding to a question on national security. An applicability frame works by linking the two attitudes, so that the considerations in common are given greater weight. As an example of an applicability treatment, consider the frames employed in Levendusky and Malhotra (2015), replicated below as Studies 11 and 12. Subjects are exposed to news stories depicting the electorate as being either moderate or polarized. The dependent variable is subjects' attitudes toward four policies. The ancillary attitude is the value subjects place on moderation in politics; the relevant considerations are given more weight when responding to the policy questions.

Information treatments operate by adding new considerations. These treatments are sometimes referred to as *belief change* treatments, a term I avoid so as to maintain a clear distinction between the cause (information) and the effect (attitude change), although in some instances, information treatments may persuade by changing beliefs. The treatments in the capital punishment study (Study 1 below) are examples of information treatments: subjects were exposed to research reports on the efficacy of capital punishment before registering their support or opposition to the practice.

I have cast framing treatments as operating on the weights and information treatments as introducing new considerations, but there is no reason to imagine that a given framing treatment could never introduce some new consideration or that an information treatment does not influence the weights given to pre-existing considerations. Nevertheless, the distinction is important to maintain in order to highlight the processes by which effects may persist or decay.

Before moving on to the research approach, I wish to underline that the foregoing theory does not draw out subtleties with respect to individual differences across subjects. Chong and Druckman (2010), for example, explore differences according to whether subjects use "memory-based" processing or are high in "need-for-evaluation." My focus here is on differences in the quality of *treatments*, not differences across subjects, though this focus should not be mistaken for an assertion that individual differences do not matter or are unimportant.

2 Research Approach

The theoretical framework outlined above makes predictions about the persistence of survey experimental effects from the characteristics of the treatments themselves. The unit of analysis, therefore, is the experimental treatment, not the experimental subject. In order to obtain a sufficient number of units, a large number of experiments is required.

The database for this investigation is drawn from 18 panel survey experiments, comprising 60 treatment-control comparisons. The majority of the studies (15) were conducted on Amazon's Mechanical Turk, with the remainder (3) conducted on a nationally-representative survey administered by GfK. Four of the studies conducted on Mechanical Turk are original: Study 1 (capital punishment), Study 2 (minimum wage), Study 17 (contentious global warming), and Study 18 (Newspapers). The remaining fourteen experiments are replications of other authors' work. Four (Haider-Markel and Joslyn, 2001; Transue, 2007; Chong and Druckman, 2010; Nicholson, 2012) were chosen for two main reasons. First, as evidenced by their placement in top political science journals, they are well-crafted and theoretically important. Second, because they employ text-based treatments, they are directly replicable. Many other prominent survey experiments (e.g., Nelson, Clawson and Oxley (1997); Mutz and Reeves (2005); Brader (2005)) present subjects with short videos that, for example, depict political debates or advertisements. The original materials for such experiments are sometimes no longer available and further, may appear out-of-date to contemporary audiences.

Seven studies originally conducted by other authors on the Time-Sharing Experiments in the Social Sciences (TESS) platform were also selected for replication. I chose these seven studies because they met the following criteria:

1. The original experiments showed precisely estimated and substantively large immediate average treatment effects. In the absence of so-called "sleeper effects" (Hovland et al., 1949; Hovland and Weiss, 1951; Cook and Flay, 1978), unambiguous immediate effects are required for the study of persistence.

2. The experiments did not employ vignettes or hypothetical candidate scenarios. Asking "Do you prefer Candidate A or Candidate B" after 10 days has no meaning without re-displaying the stimuli, i.e., re-treating subjects.

3. The frames were sufficiently relevant to present-day (2015) politics. Experiments that relied on subjects' presumed familiarity with, for example, the 2008 presidential election, could not be directly replicated.

4. The original authors provided a clear study description or write-up on the TESS website. I did not pursue the few studies lacking this information. In their study of the TESS database, Franco et al. (2014) showed that null results were far less likely to be written up, suggesting that the studies excluded for this reason would probably have been excluded for reason 1 had more information been available in any case.

Three of the TESS studies (Hiscox, 2006; Hopkins and Mummolo, 2015; Levendusky and Malhotra, 2015) were replicated in parallel on Mechanical Turk and GfK. The other four TESS replications (Brader, 2005; Craig and Richeson, 2014; Johnston and Ballard, 2016; McGinty et al., 2013) were conducted on Mechanical Turk only.

The experiments all followed a similar panel survey experimental design.[2] In wave 1, subjects answered a battery of demographic questions, were exposed to their randomly-assigned treatments, and then responded to questions whose answers constituted the outcome variables. Most subjects spent between 5 and 10 minutes on the wave 1 surveys.

Wave 2 occurred approximately 10 days after treatment and consisted only of the dependent variables. The MTurk studies were executed with the assistance of the `MTurkR` package for `R` (Leeper, 2015). In an effort to guard against a possible artifact in which subjects simply remember how they previously answered questions, in some studies, innocuous features of the question presentation were manipulated. For example, in Studies 7-13, whether follow-up response options were displayed vertically or horizontally was randomly varied. Similarly, in one study (Study 18: Newspapers), half the sample was exposed to an additional op-ed treatment in order to defend against a demand effect in which subjects merely remember what they think the experimenters want to hear. The results of these manipulations (not reported here) are negligible, suggesting that these alternative explanations for persistence are not at work.

2.1 Measuring Persistence

When discussing the duration of treatment effects, it is common to dichotomize effects into those that persist and those that do not (Druckman and Nelson, 2003; de Vreese, 2004; Mutz and Reeves, 2005). For example, if a treatment caused a statistically significant shift in outcomes at time 1, but the time 2 estimate is not statistically significant, the treatment effect is said to "not persist." While this categorization may have some

[2]Studies 1 and 2 also included a pre-treatment survey before wave 1.

heuristic value, the approach taken here is to measure the percentage of the treatment effect measured at time 1 still present at time 2.

In the expressions below, $Y_{T1,i}$ and $Y_{T2,i}$ represent the outcome for subject i at time 1 and 2, respectively. Z_i is subject i's treatment assignment; the average treatment effect (ATE) at time 1 is represented by α_1 and the ATE at time 2 by β_1. The error terms ϵ_i and η_i represent idiosyncratic variation in the outcome variables not accounted for by Z_i.

$$Y_{T1,i} = \alpha_0 + \alpha_1 Z_i + \epsilon_i$$
$$Y_{T2,i} = \beta_0 + \beta_1 Z_i + \eta_i$$

The ratio of β_1 to α_1 is our metric of treatment effect persistence, γ, that is, the percentage of the ATE at time 1 still evident at time 2.

$$\gamma = \frac{\beta_1}{\alpha_1}$$

I estimate α_1 and β_1 separately by an ordinary least squares (OLS) regression of outcomes on the treatment indicator. I estimate γ by calculating the ratio of $\hat{\beta_1}$ to $\hat{\alpha_1}$, and obtain standard errors and confidence intervals via the nonparametric bootstrap. Because γ is a ratio, when $\hat{\alpha_1}$ is very close to zero, persistence estimates may be correspondingly very large, and further, are very unstable, swinging from extremely large positive values to extremely large negative values. To combat this problem, I trim off the 0.5$^{\text{th}}$ and 99.5$^{\text{th}}$ percentiles of the bootstrap samples before computing standard errors. This operation has a very small influence on the resulting confidence interval estimates, but constrains estimates of standard errors to a reasonable range of values. In practice, these standard error estimates are slightly larger than robust standard errors resulting from two-stage least squares, a parametric model that corresponds to this setup.

Another complication attending to the estimation of treatment effect persistence is panel attrition. If not all subjects respond to the survey at time 2, then estimates of β_1 (and therefore γ) may be biased. I address this problem with an assumption about the sample distribution of subject types. Subjects have a set of potential outcomes R that may differ depending on their treatment assignment. For example $R_i(Z_i = 1)$ describes the subject i's reporting status when i is in treatment ($Z_i = 1$). If this quantity is equal to 1, i

reports in the second round and not otherwise. Subjects who never report in the second round, regardless of treatment status ($R_i(Z_i) = 0$) are called "Never-reporters." Subjects who always report, regardless of treatment status ($R_i(Z_i) = 1$) are called "Always-reporters." Similarly, types such as "If-treated reporters" and "If-untreated reporters" are theoretically possible. These experiments, however, all have more than two treatment arms, so the possible number of subject types quickly becomes unwieldy. In the analyses that follow, subjects are assumed to be either Never-reporters or Always-reporters. The estimates of over-time persistence only pertain to the Always-reporters. For this reason, α_1 (the immediate effect of treatment) will be estimated among Always-reporters only as well.

2.2 Theoretical Predictions

In this section, I use the theoretical framework described above to generate persistence predictions for this set of experiments.[3] The descriptions below are limited to the bare-bones justifications for the associated predictions. See the appendix for full descriptions of study design, including treatment stimuli, number of subjects, and wording of dependent variables. Finally, scholars may reasonably disagree about which set of mechanisms is primarily responsible for treatment effects, so I will make the case for my own categorization while recognizing that others are possible as well.

- *Study 1: Capital Punishment*
  Treatments: Pro, con, or null social scientific evidence about the efficacy of capital punishment in deterring crime.
  Mechanisms: *Information*. The evidence was presented in graphs with accompanying explanatory text.
  Prediction: Strong persistence.

- *Study 2: Minimum Wage*
  Treatments: Short web videos, two in favor of raising the minimum wave, two against it, and two placebos.
  Mechanisms: *Information*. The non-placebo treatments offered very detailed information about the economics and politics of raising the minimum wage.
  Prediction: Strong persistence.

- *Study 3: Gun Control*
  Original study: Haider-Markel and Joslyn (2001)
  Treatments: Gun control question framing in terms of public safety or individual rights.
  Mechanisms: *Accessibility*. Safety and rights considerations are triggered by the frame. *Applicability*.

[3]The predictions for the 10 TESS replication studies (Studies 7 - 16) were pre-registered and posted at egap.org

For some subjects, these considerations may be newly applicable to the gun control debate. *Information* The questions also contained a small dose information about a concealed carry gun law.
Prediction: Moderate persistence. In hindsight, I would change this prediction to weak persistence, since the applicability and information channels are so thin. I came to this realization with the benefit of having seen the experimental results, so I nevertheless group this study with the "moderate" predictions in the analyses that follow.

- *Study 4: Superordinate Identity*
  Original study: Transue (2007)
  Treatments: Primes that activate either particularistic (racial/ethnic) or superordinate (national) identity.
  Mechanisms: *Accessibility* The treatment makes one identity more salient or accessible, but does not create new associative links or offer new information.
  Prediction: Weak persistence.

- *Study 5: Patriot Act*
  Original study: Chong and Druckman (2010)
  Treatment: Pro or con information about the Patriot Act.
  Mechanisms: *Information*. Subjects are directly given facts about the Patriot Act. *Accessibility*. The information framed to activate some pre-existing considerations such as individual freedom. *Applicability*. The information may also create links between considerations such as surveillance and national security.
  Prediction: Strong persistence. Of the replications, this was the only original study to measure outcomes at multiple points in time – the strong persistence prediction, therefore, is made with the benefit of knowing beforehand that the effects did indeed persist for 10 days originally.

- *Study 6: Elite Endorsements*
  Original study: Nicholson (2012)
  Treatments: Policy proposals with partisan endorsement cues.
  Mechanisms: *Accessibility*. These cues may make considerations of partisan loyalty more accessible but do not apply, for example, Democratic party values to new issue areas. *Information*. This treatment may also operate by giving subjects the novel information that someone they respect (or despise) endorses a proposal. This sort of information is qualitatively different from information that, for example, informs subjects of the policy's expected benefits.
  Prediction: Weak persistence.

- *Studies 7 and 8: Free Trade A and B*
  Original study: Hiscox (2006)
  Treatments: Valence frames, which describe why some Americans do or do not support free trade, and an expert frame, which states that professional economists are united in their support for free trade.
  Mechanisms: *Accessibility*. The valence frames prompt subjects to give negative or positive economic considerations extra weight. *Information* Subjects learn that professionals in this complicated area appear united on this question.
  Prediction: Moderate persistence.

- *Studies 9 and 10: Frame Breadth A and B*
  Original study: (Hopkins and Mummolo, 2015)

Treatments: Pro-conservative statements by U.S. senators.
Mechanisms: *Accessibility*. These treatments operate by making terrorism, crime, health care costs, or wasteful spending more salient when answering budget questions, though the senators do provide information in their arguments.
Prediction: Weak persistence. In hindsight (but also with the benefit of having seen the results) I would now predict moderate persistence for these treatments in view of their information content, though a weak prediction was pre-registered. In the analysis below, the pre-registered predictions will be used to evaluate the theoretical framework.

- *Studies 11 and 12: Polarization A and B*
Original study: Levendusky and Malhotra (2015)
Treatments: Articles depicting politics as polarized or moderate.
Mechanisms: *Applicability*. If subjects place a value on being "moderate" or "reasonable" members of the electorate, and this value is made more salient via the polarized treatment, then this consideration will be made newly applicable to the policy questions.
Prediction: Moderate persistence.

- *Study 13: Immigration*
Original study: Brader (2005)
Treatments: Positive or negative newspaper stories on immigration featuring either Latino or European immigrants.
Mechanisms: *Accessibility*. The articles are intended to heighten racial considerations when providing immigration policy preferences. *Information*. The accompanying text also includes some information about, variously, the positive or negative economic effects of immigration.
Prediction: Moderate persistence.

- *Study 14: System Threat*
Original study: (Craig and Richeson, 2014)
Treatments: A frame highlighting the increasing population share of minorities.
Mechanisms: *Accessibility*. The treatment may trigger respondents' racial anxieties.
Prediction: Weak persistence.

- *Study 15: Expert Economists*
Original study: Johnston and Ballard (2016)
Treatments: Subjects are provided the consensus view among economists for various policy proposals.
Mechanisms: *Information*. Subjects are given information about expert views on these five policies. *Accessibility*. Subjects are more or less told which answer is the "correct" one. If the first mechanism is at work, the treatment ought to have enduring effects. If effects are primarily due to the second mechanism, effects should be fleeting. Overall, I view this treatment as operating through the accessibility channel: the "correct" answer is more accessible at the moment of providing a survey response.
Prediction: Weak persistence.

- *Study 16: Mental Illness*
Original study: McGinty et al. (2013)
Treatments: Newspaper vignettes depicting a mass shooting and providing policy proposals.
Mechanisms: *Accessibility*. One treatment (the story depicting the shooting) operates by highlighting

considerations about gun violence. *Information*. The second treatment (the policy proposals) operates by providing information about various gun policies.
Prediction: Moderate persistence.

- *Study 17: Contentious Global Warming*
  Treatments: Graphs that showed a scientific consensus about warming trends or evidence of a warming "hiatus".
  Mechanisms: *Information*. The treatment explicitly furnishes subjects with new information about warming trends.
  Prediction: Strong persistence.

- *Study 18: Newspapers*
  Treatments: Newspaper op-ed pieces on five policy areas.
  Mechanisms: *Information*. The op-eds make extended arguments in favor of a particular policy.
  Prediction: Strong persistence.
  In this experiment, three waves of post-treatment measurement were conducted, so I will provide a more detailed explanation of this study below. I also note that the design of this study was strongly influenced by (Hopkins and Mummolo, 2015), replicated here as Studies 9 and 10.

3 Results

In this section, I present summaries of the persistence of effects for all studies. To ease interpretation, all outcome variables were standardized by subtracting off the control group mean and dividing by the control group standard deviation. In Study 6 (Elite Endorsements), the ordinal dependent variable is rescaled to take numeric values between -1 and 1 in order to apply this standardization procedure. See appendix tables A.1 and A.2 for the complete results of all studies.

Figure 1 plots standardized time 1 treatment effect estimates on the horizontal axis and time 2 estimates on the vertical axis, separately for each study. Points lying on the 45 degree line represent perfect persistence: the time 2 estimate is the same as the time 1 estimate. Points above the 45 degree line indicate that the effect strengthened over time and points below indicate that the treatment effect decayed over time. Figure 1 shows a great deal of variation in persistence from study to study. Study 1 (Capital Punishment) shows a very strong degree of persistence: effects that are strongly positive in time 1 are strongly positive in time 2. Study 3 (Gun Control) featured a very robust time 1 estimate; this effect evaporates completely by time 2.

Figure 2 plots estimates of γ, the percentage of the time 1 effect still present at time 2 separately for each treatment/dependent variable combination. When the estimate of α, the effect of treatment at time 1, is very small (or imprecisely estimated), estimates of γ become very unstable. In some cases, the estimate is far outside the -1.5 to 1.5 range displayed on the plot. The vertical line represents the precision weighted average of the effects in each study, and the vertical shading displays the 95% confidence interval implied by the standard error[4] of the weighted average (Gerber and Green, 2012, p. 356). In all 18 studies, the weighted average of the persistence estimates is positive and the 95% confidence interval does not cross zero in 10 of these studies.

Figures 1 and 2 display the same persistence estimates in two related ways. The points in Figure 2 can be thought of as the ratio of the y-axis to the x-axis in Figure 1. The slope of the points in Figure 1 is a summary of the overall persistence within a study, similar to the weighted average summary represented by the vertical lines in Figure 2.

Table 1 presents persistence estimates of γ and associated standard errors for each treatment effect. These estimates are identical to the estimates that would be obtained by dividing the y-axis by the x-axis in Figure 1. As indicated by the sometimes large standard errors, these estimates can be quite imprecise. A meta-analytic summary of all studies is obtained by taking a weighted average of the estimates, where the weights are the inverse of the squared standard errors (Borenstein et al., 2009, p. 65). The overall estimate is 44% with a standard error of 2%.

To what extent are the predictions of the theoretical framework presented above borne out? Using the study-specific summaries presented in Table 1, I take the average of studies in each prediction category (weak, moderate, strong), weighting each study's average persistence by the inverse of its squared standard error. The table shows that on average, studies predicted to have weak persistence were at 18% of their original strength after 10 days. Those predicted to have moderate persistence were at 34% the original magnitude, while studies with a strong persistence prediction were at 58% of the time 1 effect at time 2. The differences across the prediction groups are all statistically significant at $p < 0.01$. Broadly speaking, these results confirm the predictions generated by the theoretical framework described above.

[4]The calculation of this standard error relies on the assumption that the observations are independent of one another; they are not. In the case of multiple treatment arms, effects are all assessed relative to the same control group. In the case of multiple outcome variables, estimates compare the same subjects on multiple dimensions, necessarily introducing dependence among them. Both of these features will produce a positive covariance, implying that the variance estimates presented in the graph are slightly anti-conservative.

Figure 1: Standardized Average Treatment Effect Estimates: Time 1 versus Time 2

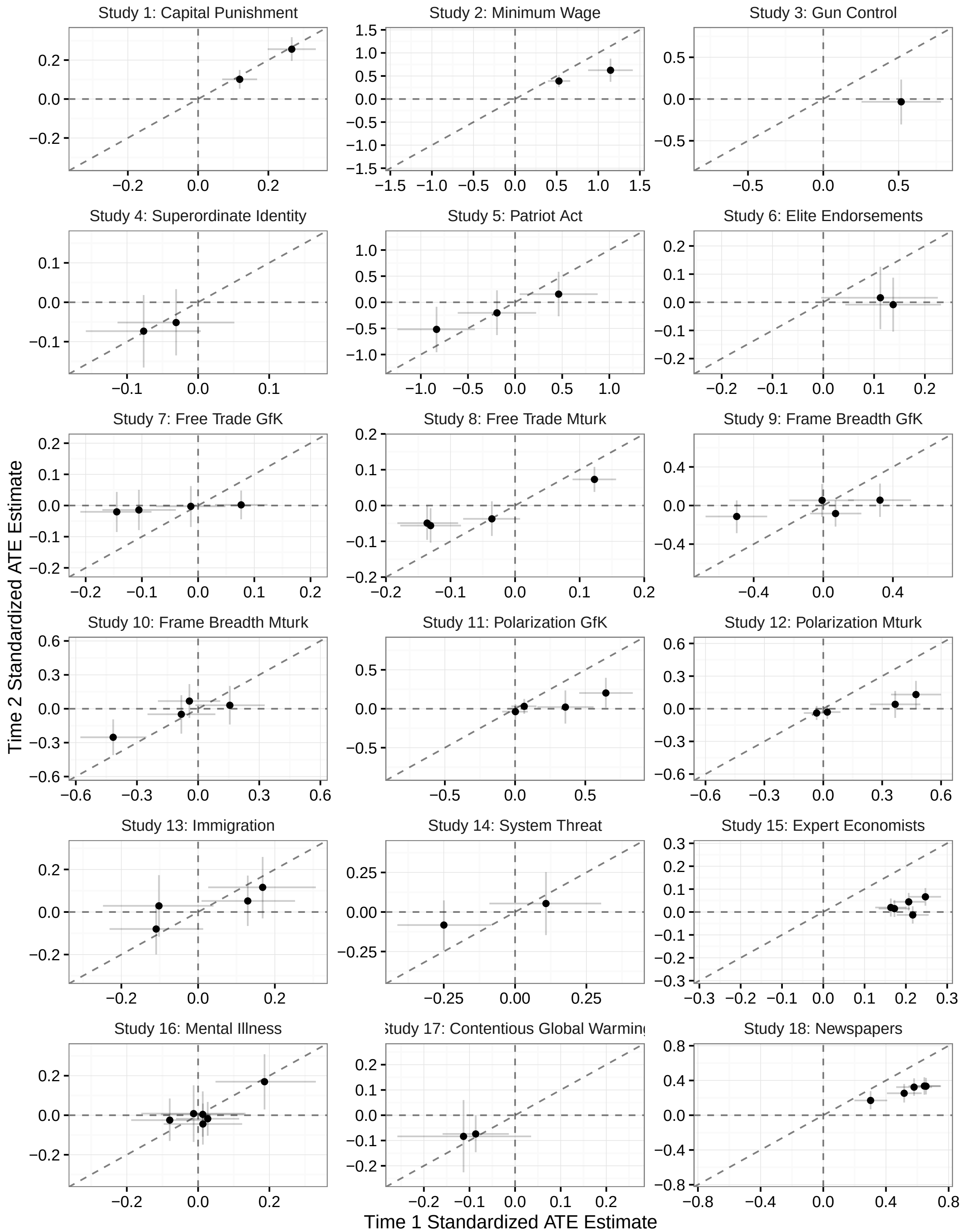


Figure 2: Persistence Estimates for 18 Studies

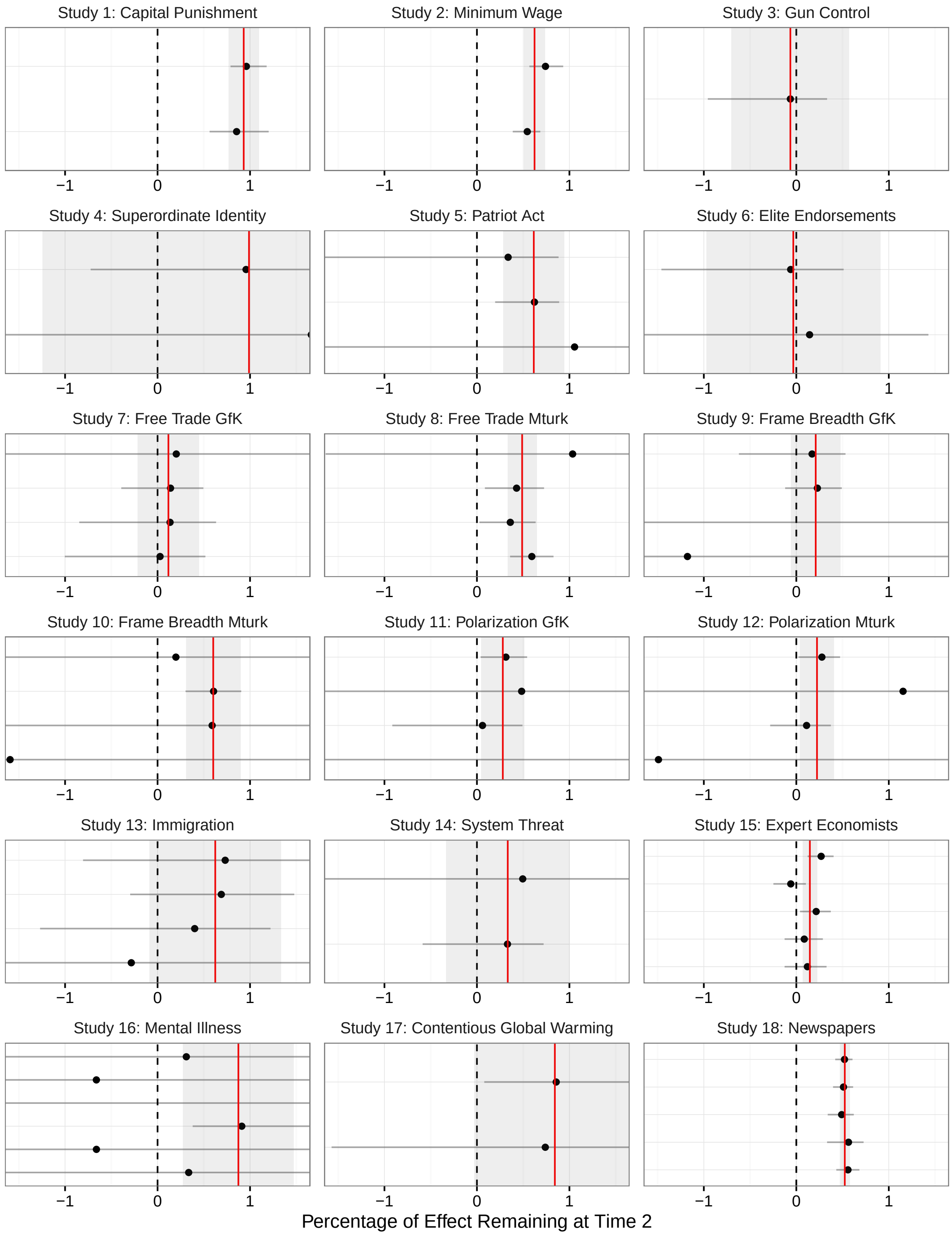


Table 1: Average Persistence of 18 Studies

	Study	Average Persistence	SE	Prediction
Study 6	Elite Endorsements	-0.03	0.48	Weak
Study 15	Expert Economists	0.15	0.04	Weak
Study 9	Frame Breadth GfK	0.21	0.14	Weak
Study 14	System Threat	0.33	0.34	Weak
Study 10	Frame Breadth Mturk	0.60	0.15	Weak
Study 4	Superordinate Identity	0.99	1.14	Weak
Study 3	Gun Control	-0.06	0.32	Moderate
Study 7	Free Trade GfK	0.12	0.17	Moderate
Study 12	Polarization Mturk	0.22	0.09	Moderate
Study 11	Polarization GfK	0.28	0.12	Moderate
Study 8	Free Trade Mturk	0.49	0.08	Moderate
Study 13	Immigration	0.62	0.36	Moderate
Study 16	Mental Illness	0.87	0.31	Moderate
Study 18	Newspapers	0.52	0.03	Strong
Study 5	Patriot Act	0.61	0.17	Strong
Study 2	Minimum Wage	0.62	0.06	Strong
Study 17	Contentious Global Warming	0.84	0.44	Strong
Study 1	Capital Punishment	0.93	0.08	Strong
	All Studies	0.44	0.02	

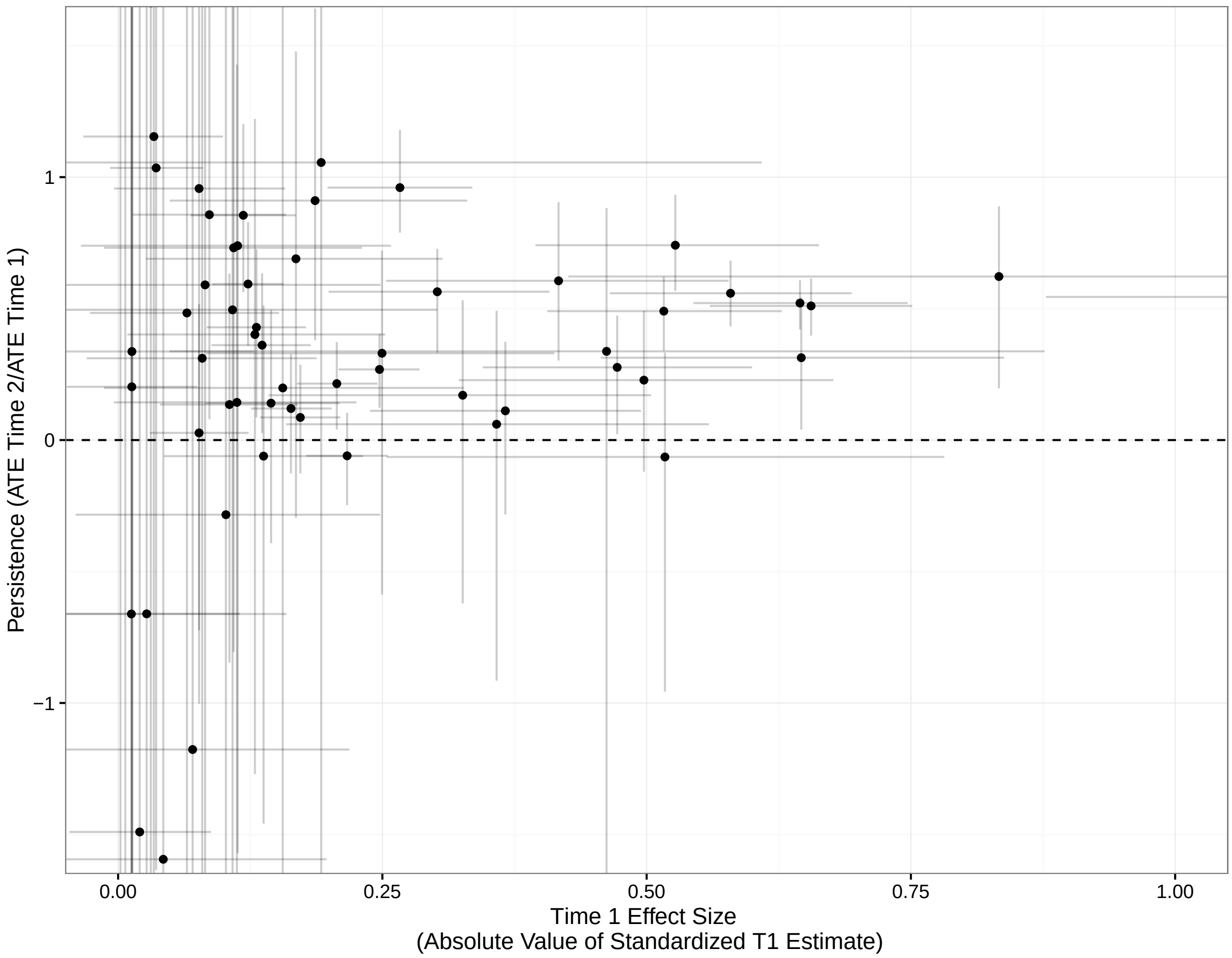
Table 2: Average Persistence, Pooled by Prediction

Prediction	Average Estimate	SE
Weak	0.18	0.04
Moderate	0.34	0.05
Strong	0.58	0.02

3.1 Alternative Explanation: Large Effects Persist?

To what extent can these results be explained by the more parsimonious theory that “large effects persist; small ones do not?” Figure 3 shows that this theory fails to provide meaningful insight into why some effects persist and others do not. On the x-axis, the effect size estimated at time 1 (in absolute value) is presented; the persistence estimate is on the y-axis. If the “large effects” theory of persistence were true, we would observe a positive correlation in these data. The relationship between these two variables, however, is essentially flat. The deviations from this flat relationship are highly heteroskedastic – the variance of the persistence estimates gets larger as the time 1 estimate gets weaker. Stated another way, the variance of a ratio estimate increases as the denominator gets closer to zero.

Figure 3: Magnitude of T1 ATE Versus Persistence



4 Long Term Stability

The persistence results presented thus far suggest that treatment effects decline to approximately half their original strength after 10 days. If effects continue to decay at the same rate, we might project that they dissipate entirely after 20 days. Alternatively, effects might exhibit some measure of proportional decay, resulting in 25% strength after 20 days, 12.5% strength after 30 days, and so on. In order to trace a fuller picture of the rate of decay, we need to measure outcomes at more points in time.

One of the studies (Study 18: Newspapers) included in the batch of 18 analyzed above included a third wave of post-treatment measurement after 30 days. That study employed a 6-group design: 5 groups that read an op-ed on a particular policy area and one control group. All subjects answered a series of questions in each issue area. These questions were combined into a composite scale using principal components analysis. Question wordings and details of scale construction are available in the appendix.

Figure 4 displays the results of this experiment. In each panel, the average outcomes for a single policy area are shown on the vertical axis. The horizontal axis is the number of days since treatment, and displays three measurements for each of the six randomly-formed groups: immediately after reading the op-eds, 10 days after treatment, and 30 days after treatment.

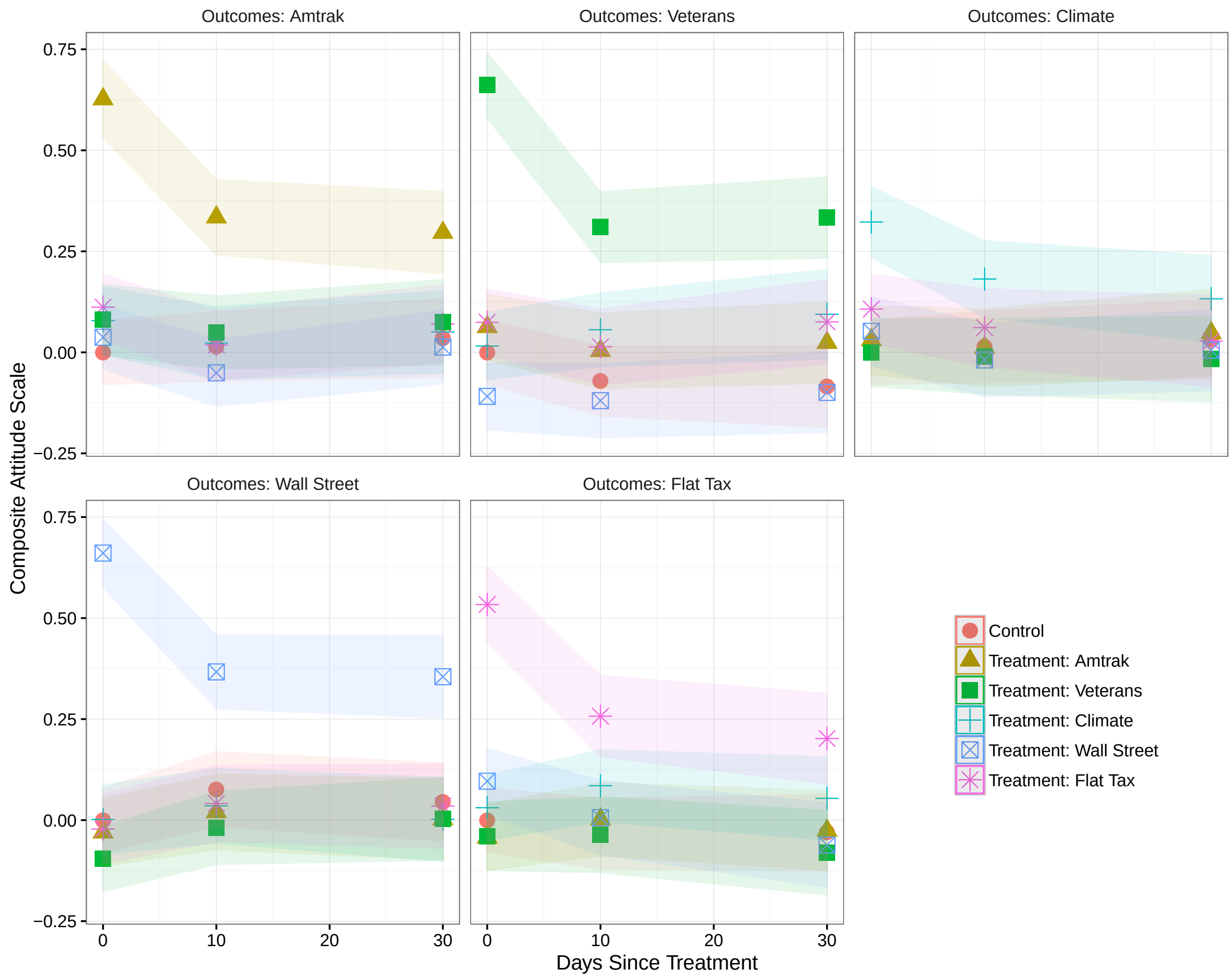
The pattern of results is extraordinarily consistent across all five issue areas. The group that was treated with the issue-specific op-ed has much higher outcomes immediately after zero; the remaining groups cannot be distinguished from control. After 10 days, as we saw above, treatment effects decay to 52.1% (SE=2.4%) their original magnitudes. Remarkably, very little further decay appears to take place between 10 days and 30 days after treatment. After 30 days, average persistence declines to 50.0% (SE = 2.4%), indicating that the change in treatment effect persistence between day 10 and day 30 is not statistically significant.

These results suggest a "hockey stick" pattern of decay: after an initial decline, subsequent decreases are smaller. This pattern can be reconciled with the larger theoretical setup described above: when new considerations are introduced, they arrive with artificially high weights attached. Over time, these weights "settle," but the consideration itself remains in the mind of the subject. This conjecture should be subjected to further empirical testing in experiments that track subjects' responses to treatment over longer periods of time.

5 Discussion

The results from this diverse set of studies yields a striking summary conclusion: on average, survey experimental treatment effects are approximately half their immediate size after 10 days. The challenge set by Gaines, Kuklinski and Quirk (2007) was to demonstrate that survey experimental estimates of effects are politically relevant by showing that such effects endure beyond the few minutes it takes to complete a public

Figure 4: Long Term Effects of Newspaper Op-eds



opinion survey. By that standard, I have demonstrated that persuasive treatments of the kind studied here likely do have important and long-lasting impacts on political discourse.

The theoretical framework above predicted that treatments that operate primarily by making considerations more accessible would have fleeting effects, and treatments that operate by creating new associations or providing new information would last longer. This pattern was confirmed: the treatments in studies 1, 2, 5, 17 and 18 provided subjects with new information and all had strongly persistent effects. Studies 7 and 15 were predicted to increase the accessibility of certain considerations; the effects of these treatments appear not to have persisted after 10 days.

These results also offer an opportunity to revisit the large literature in psychology on the so-called "sleeper effect" (Hovland et al., 1949; Hovland and Weiss, 1951; Cook and Flay, 1978), in which initially null effects would blossom into strong effects in time. The proposed mechanism is that subjects would forget why they initially discounted some new piece of information; when the discounting falls away, the information would exert some persuasive influence. Attempts to document the existence of the sleeper effect have usually failed. Consistent with this line of evidence, none of the studies reported here saw an initially null result that became statistically significant. I do grant, however, that a determined advocate of the sleeper effect would rightly point out that the required theoretical conditions are probably not present in these experiments.

These results indicate that the psychological mechanisms of *accessibility*, *applicability*, and *new information* may indeed play a role in the persistence of treatment effects; they also may play a role in the strength of treatment effects measured immediately. However, the treatments in this set of studies do not only vary along these dimensions. For example, each one is concerned with a different issue area. Further, some deliver treatments as questions and others as statements. In order to determine the extent to which each of these mechanisms influences persistence, it would be preferable to hold all other features of the treatment constant and randomly assign the mechanisms. I leave this project to future research.

References

Ajzen, Icek and Martin Fishbein. 1980. *Understanding Attitudes and Predicting Social Behavior*. Vol. 278 Englewood Cliffs, NJ: Prentice Hall.

Baden, Christian and Sophie Lecheler. 2012. "Fleeting, Fading, or Far-Reaching? A Knowledge-Based Model of the Persistence of Framing Effects." *Communication Theory* 22(4):359–382.

Borenstein, Michael, Larry V. Hedges, Julian P.T. Higgins and Hannah R. Rothstein. 2009. *Introduction to Meta-Analysis*. Hoboken, NJ: John Wiley & Sons.

Brader, Ted. 2005. "Striking a Responsive Chord: How Political Ads Motivate and Persuade Voters by Appealing to Emotions." *American Journal of Political Science* 49(2):388–405.

Chong, Dennis and James N. Druckman. 2010. "Dynamic Public Opinion: Communication Effects over Time." *American Political Science Review* 104(04):663–680.

Chong, Dennis and James N. Druckman. 2012. Dynamics in Mass Communication Effects Research. In *The Sage Handbook of Political Communication*, ed. Holli Semetko and Maggie Scammell. Los Angeles, CA: Sage Publications pp. 307–323.

Cook, Thomas D and Brian R Flay. 1978. "The Persistence of Experimentally induced Attitude Change." *Advances in Experimental Social Psychology* 11:1–57.

Craig, Maureen A. and Jennifer A. Richeson. 2014. "More Diverse Yet Less Tolerant? How the Increasingly Diverse Racial Landscape Affects White Americans' Racial Attitudes." *Personality and Social Psychology Bulletin* 40(6):750–761.

de Vreese, Claes. 2004. "The Effects of Strategic News on Political Cynicism, Issue Evaluations, and Policy Support: A Two-Wave Experiment." *Mass Communication and Society* 7(2):191–214.

Druckman, James N. 2001. "The Implications of Framing Effects for Citizen Competence." *Political Behavior* 23(3):225–256.

Druckman, James N. and Kjersten R. Nelson. 2003. "Framing and Deliberation: How Conversations Limit Elite Influence." *American Journal of Political Science* 47(4):729–745.

Franco, Annie, Neil Malhotra and Gabor Simonovits. 2014. "Publication Bias in the Social Sciences: Unlocking the File Drawer." *Science* 345(6203):1502–1505.

Gaines, Brian J., James H. Kuklinski and Paul J. Quirk. 2007. "The Logic of the Survey Experiment Reexamined." *Political Analysis* 15:1–20.

Gerber, Alan S. and Donald P. Green. 2012. *Field Experiments: Design, Analysis, and Interpretation*. New York: W.W. Norton.

Haider-Markel, Donald P. and Mark R. Joslyn. 2001. "Gun Policy, Opinion, Tragedy, and Blame Attribution: The Conditional Influence of Issue Frames." *The Journal of Politics* 63(2):520–543.

Hiscox, Michael J. 2006. "Through a Glass and Darkly: Attitudes Toward International Trade and the Curious Effects of Issue Framing." *International Organization* 60(03):755–780.

Hopkins, Daniel J. and Jonathan Mummolo. 2015. "Assessing the Breadth of Framing Effects." *Unpublished Manuscript* .
URL: `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2863930`

Hovland, Carl I., Arthur A. Lumsdaine and Fred D. Sheffield. 1949. *Experiments on Mass Communication, Vol. 3.* Princeton, NJ: Princeton University Press.

Hovland, Carl I and Walter Weiss. 1951. "The Influence of Source Credibility on Communication Effectiveness." *Public Opinion Quarterly* 15(4):635–650.

Johnston, Christopher D. and Andrew O. Ballard. 2016. "Economists and Public Opinion: Expert Consensus and Economic Policy Judgments." *The Journal of Politics* 78(2):443–456.

Lecheler, Sophie and Claes H. de Vreese. 2011. "Getting Real: The Duration of Framing Effects." *Journal of Communication* 61(5):959–983.

Leeper, Thomas J. 2015. *MTurkR: Access to Amazon Mechanical Turk Requester API via R*. R package version 0.6.5.1.

Levendusky, Matthew and Neil Malhotra. 2015. "Does Media Coverage of Partisan Polarization Affect Political Attitudes ?" *Political Communication* .

McGinty, Emma E., Daniel W. Webster and Colleen L. Barry. 2013. "Effects of News Media Messages about Mass Shootings on Attitudes Toward Persons with Serious Mental Illness and Public Support for Gun Control Policies." *American Journal of Psychiatry* 170(5):494–501.

Mutz, Diana C. and Byron Reeves. 2005. "The New Videomalaise: Effects of Televised Incivility on Political Trust." *American Political Science Review* 99(1):1–15.

Nelson, Thomas E., Rosalee A. Clawson and Zoe M. Oxley. 1997. "Media Framing of a Civil Liberties Conflict and Its Effect on Tolerance." *American Political Science Review* 91(3):567–583.

Nelson, Thomas E., Zoe M. Oxley and Rosalee A. Clawson. 1997. "Toward a Psychology of Framing Effects." *Political Behavior* 19(3):221–246.

Nicholson, Stephen P. 2012. "Polarizing Cues." *American Journal of Political Science* 56(1):52–66.

Tewksbury, David H., J. Jones, M. W. Peske, A. Raymond and W. Vig. 2000. "The Interaction of News and Advocate Frames: Manipulating Audience Perceptions of a Local Public Policy Issue." *Journalism & Mass Communication Quarterly* 77(4):804–829.

Transue, John E. 2007. "Identity Salience, Identity Acceptance, and Racial Policy Attitudes: American National Identity as a Uniting Force." *American Journal of Political Science* 51(1):78–91.

Zaller, John R. 1992. *The Nature and Origins of Mass Opinion*. New York: Cambridge University Press.