Causal Effects Under Self-Selection

Thomas J. Leeper*

November 25, 2013

Abstract

The randomized assignment of units to treatment status provided by an experiment provides unparalleled leverage to understand the average causal effect of an assigned treatment. Yet political communication research — like many areas of social science — has long expressed concern that experimental studies ignore the fact that the particular treatment an individual receives in the real world is often self-selected, not assigned. This paper formally introduces the self-selection problem in the language of potential outcomes and argues that it presents unique methodological challenges and introduces important, often unspecified, assumptions about treatment response. This paper describes those assumptions and their implications for inference. To resolve these challenges, a novel approach is presented that can uncover various effects of self-selected treatments using a combination of the three-group "hybrid" design proposed by Gaines and Kuklinski (2011) and an assumption of conditional ignorability.

*Department of Political Science and Government, Aarhus University. The author graciously acknowledges the financial support of NSF Doctoral Dissertation Improvement Grant SES-1160156. Special thanks are also due to Jamie Druckman and Jim Kuklinski for valuable feedback on earlier drafts of this paper and to Kevin Arceneaux, Rick Lau, and Jason Seawright for helpful discussions on some of the ideas elaborated here.

Social science experiments are meant to emulate reality or some abstraction thereof. Given that reality frequently involves self-selection into treatments, it is unsurprising that researchers (especially in the area of political communication) are starting to design experiments with choice over treatment alternatives (see Arceneaux and Johnson 2012; Levendusky 2012; Gaines and Kuklinski 2011; Druckman, Fein, and Leeper 2012). The two-condition randomized experiment only enables the researcher to readily estimate the sample average treatment effect (SATE) and therefore limits the researcher to answering only one causal question: "What would the effect be if everyone in the sample received the treatment rather than the control?" (Gaines and Kuklinski 2011). This is often interesting but is unhelpful when the researcher's interest lies in understanding the size of a treatment's effects across individuals who have the opportunity to choose their own treatment. What effects do treatments have when self-selected?

The SATE does not necessarily provide particularly helpful answers to this question, unless all individual-level causal effects are homogeneous, individuals always make the same choices, and choosing produces outcomes identical to random assignment. Under those narrow conditions, however, self-selection is causally irrelevant. Gaines and Kuklinski (2011) point out that the researcher is much more often interested in estimating effects of a self-selected treatment on those units that actually choose it.[1] They advocate for a somewhat innovative method for estimating two causal effects (the effect for selectors and the effect for non-selectors) from a "hybrid design" that combines a randomized and observational component.[2] Since Bennett and Iyengar (2008) called for a greater experimental focus on self-selection,[3] hybrid designs have been used in a flurry

[1]That is, research acknowledges that internal validity may be influenced by participants' preferences over treatment alternatives (Corrigan and Salzer 2003; Shadish, Cook, and Campbell 2001; McPherson, Britton, and Wennberg 1997; McPherson and Britton 1999, 2001).

[2]Similar designs have been proposed by Rucker (1989); Shadish, Clark, and Steiner (2008); Cook, Shadish, and Wong (2008); Steiner et al. (2010); Steiner, Cook, and Shadish (2011); Arceneaux and Johnson (2007, 2011).

[3]Reading Hovland (1959), Bennett and Iyengar (2008) rightly point out that "manipulational control actually weakens the ability to generalize to the real world where exposure to politics is typically voluntary. Accordingly, it is important that experimental researchers use designs that combine manipulation with self-selection of exposure" (724).

of recent political science working papers (Boudreau 2011; Arceneaux and Johnson 2011; Levendusky 2011; Joseph 2012; Swigger 2012; Geer, Lau, and Vavreck 2012; Leeper 2012*a*; Johnson and Arceneaux N.d.; Feldman et al. 2013). Yet a thorough methodological discussion of these designs (and of self-selection more broadly) has yet to appear in the literature.[4] These papers correctly focus on self-selection as an important problem, but are vague about what "effect" is of interest in a choice situation and what the "effect" they purportedly estimate says about the consequences of self-selected exposure to a media treatment. This paper introduces and clarifies self-selection as a methodological problem, clarifies two important assumptions underlying the experimental study of self-selected treatments, expands on Gaines and Kuklinski's contribution by specifying the universe of causal estimands associated with self-selected treatments, and explains how to uncover those effects using a combination of randomization and conditional ignorability.

1 Assumptions About Self-Selection and Its Effects

In a hybrid design, individuals are assigned to one of three groups: treatment, control, or self-selection of treatment or control.[5] For example, to understand the effects of televised incivility on political trust (see Mutz and Reeves 2005), individuals in the first group might be forced to watch a rancorous clip of Fox News's *Hannity*, those in

[4]Self-selected treatments have been discussed in the medical literature, but in a generally informal fashion (see, for example, Corrigan and Salzer 2003; Janevic et al. 2003; King et al. 2005; Sidani, Miranda, Epstein, and Fox 2009; Sidani, Miranda, Epstein, Bootzin, Cousins, and Moritz 2009; Swift and Callahan 2009).

[5]More than 50 years ago, Hovland (1959) proposed — in passing — a design for studying self-selection that has independently emerged in the medical literature under the label "patient preference trial." In his words: "It should be possible to assess what demographic and personality factors predispose one to expose oneself to particular communications and then to utilize experimental and control groups having these characteristics. Under some circumstances the evaluation could be made on only those who select themselves, with both experimental and control groups coming from the self-selected audience" (Hovland 1959, 16). The effects of self-selected treatments can be seen in randomized experiments conducted separately on those selecting into and out of treatment. But the design depends heavily on the outcome equivalence assumption described below and may be problematic for those units randomly assigned to a non-selected treatment (see, for example, Floyd and Moyer 2010).

the control are forced to watch a segment from the PBS NewsHour, and the last group are given the choice of Fox or the NewsHour. Random assignment has guaranteed (in expectation) that the three randomly constituted groups have balance across all covariates and comparable distributions of potential outcomes associated with each treatment (Holland 1986).[6]

In papers using such designs, only two effects are frequently reported. The first compares the treatment to control group (i.e., the sample average treatment effect, SATE) to obtain "the effect" of the treatment:

$$\widehat{SATE} = E[Y_{T_E}] - E[Y_{C_E}] \tag{1}$$

where Y_{T_E} and Y_{C_E} are outcomes for units externally (randomly) assigned to the treatment and control groups, respectively. The second effect makes a comparison of mean outcomes between the self-selection group and the control group (i.e., "choice effect," which I denote TE_S):

$$\widehat{TE}_S = E[Y_S] - E[Y_{C_E}] \tag{2}$$

where Y_S represents outcomes for all units in the choice group (regardless of their choice). What is the "choice effect"? And what does it tell us about the influence of the treatment? Without explicit definition of a causal estimand, it is unclear what the "choice effect" means and what assumptions are invoked in interpreting that effect. The remainder of this section defines self-selection in the language of potential outcomes and describes two important assumptions necessary to give clear causal meaning to this "choice effect" comparison: self-selection based on *fixed types* and the equivalence of potential outcomes under random and self-selected assignment.

[6] Implicitly, the same proportion of individuals in each group should also make the same choice between the two treatments.

1.1 The Fixed Types Assumption

In practice, experimentalists often talk about the SATE as "the effect" — as in the effect on trust of watching uncivil political news rather than something else is x units of trust — but what we actually mean is that the central tendency of an unobservable distribution of individual-level differences between potential outcomes is x units of the outcome scale. Thus, we actually want to make statements about individual-level effects, but we are constrained in our ability to observe them. Interpreting effects from the hybrid design therefore requires attention to the meaning of individual-level potential outcomes for each of the three experimental groups (treatment, control, and choice).

In Equation 2, the researcher appears to be interested in the unobservable individual-level effect underlying the $E[Y_S] - E[Y_{C_E}]$ comparison (i.e., for person i, $Y_{iS} - Y_{iC_E}$). The individual's potential outcome revealed by the choice compared to the individual's potential outcome when exogenously assigned to control. For the individual-level equation to be well-defined, each unit must always choose the same alternative from the choice set (and/or, much less interestingly, each unit must have the same potential outcome regardless of which treatment they choose from that choice set, thus making all treatments placebos). This *fixed types assumption* requires that units have strict preferences over alternatives that are perfectly deterministic of choice. An individual who chooses Fox News always chooses Fox News. For this assumption to be plausible, choices can never vary by any individual or contextual factor (e.g., age, time constraints, prior choices, etc.).[7]

Because the individual would always choose Fox from a choice set of {Fox, NewsHour}, the choice set only presents one potential outcome, Y_{iS}. Yet, research on human decision making tells us that "people can, and frequently do, use different choice strate-

[7] A less strict (but even less plausible) variant of this assumption is that units manifest the same outcome regardless of whether they happened to choose one treatment or another from the choice set. This would mean that when presented with the choice between Fox and the NewsHour, an individual would have the same level of trust regardless of which alternative they chose. Such an assumption negates the possibility of any causal effect of anything when treatments are self-selected. This would only make sense if treatments were all placebos whose effect was due to units' anticipation of influence on an outcome rather than having actual influence in a causal process.

gies — and, thus, different information-acquisition patterns — in different situations" (Lau 1995, 188). Indeed, the idea of variable choice strategies — and thus varying choices — has withstood considerable empirical scrutiny (see, for a prominent example, Payne, Bettman, and Johnson 1993). Individuals' choice(s) among alternatives might easily be influenced by time pressures, the importance of particular criteria in making the decision among alternatives, framing of the choice and alternatives, the number of choices available, and so forth. Which alternative is most attractive to an individual can easily vary across contexts, and people are rarely constrained (except in highly controlled settings) to choose only one alternative from a set.[8] The fixed types assumption comes in strong and weak forms: The strong form requires that individuals have stable, fixed types (i.e., that their choices are always the same) such that an individual will always choose Fox News whenever it is available regardless of all other factors, whereas the weak form only requires that types are conditionally fixed (i.e., conditional on choiceset and context) such that if we were able to counterfactually re-run an experiment involving choice, each individual's choice(s) would be the same but choices in other contexts would differ.

The strong form of the fixed types assumption is generally implausible and merely satisfying the weak form disallows inference beyond the narrow scope of the choiceset used in the current experiment. Without a fixed types assumption, it becomes necessary to see how choices vary across contextual factors. Thus each unit has a large set of potential outcomes, $Y_{iS_{k|Z}}$, one for each treatment, k, in each choiceset, Z. Accordingly, the "choice effect" in Equation (2) is not well-defined because the choices and outcomes revealed by the self-selected treatment group are context-dependent and thus the array of causal effects that might be of interest increases dramatically because each unit has a causal effect associated with the potential outcome observable with each selectable treatment in the choice set. Establishing how units choose treatments from the choice set is

[8] Indeed, many self-selection designs allow research participants to alternate among the available treatments (television stations) to varying degrees (Arceneaux and Johnson 2012; Geer, Lau, and Vavreck 2012).

therefore critical to defining and estimating the effects of treatments.

1.2 The Outcome Equivalence Assumption

A second assumption in analyzing choice-based experiments relates to assignment mechanisms. According to Rubin (1991), the "assignment mechanism" is a function, possibly random or a function of covariates, that determines how a given unit receives a particular treatment. The assignment mechanism reveals a manifest outcome for each unit and produces missing data for the unobserved potential outcomes. Regardless of the specification of the assignment function, each unit typically has a single potential outcome that would manifest if it was assigned to the respective value of the treatment. For individual i, their level of trust after watching Fox News is the same regardless of the form of the assignment mechanism.[9] Each unit i has only the same two potential outcomes to be revealed: (1) Y_{iT}, regardless of how they receive the treatment, and (2) Y_{iC}, regardless of how they receive the control.[10]

Contrary to this *potential outcomes equivalence* assumption, when self-selection (as opposed to an external assignment agent) determines treatment assignment, the assignment mechanism could also reveal different potential outcomes for the same unit exposed to the same treatment.[11] Being given a choice of treatment, choosing that treat-

[9]Of course, the manifest outcome revealed by an external agent enforcing a particular (random or nonrandom) assignment mechanism may be different for each unit (i.e., random assignment may not expose the same manifest outcome for unit i as another random assignment mechanism or a non-random assignment mechanism).

[10]All methodological writing about the estimation of causal effects from nonrandom assignment mechanisms has implicitly invoked this assumption (Rosenbaum and Rubin 1983; Rosenbaum 2009).

[11]Following standard methodological writing on causal inference to define causal effects of self-selection — as Gaines and Kuklinski (2011) do — confounds the unique challenges of self-selection with those of other nonrandom assignment mechanisms. While research on noncompliance offers some parallels to the problem of self-selection, theorizing problems of self-selection focuses on experimental units that are capable of making choices as opposed to units that are not presented with an explicit choice among treatments (e.g., agricultural plots or passive recipients of phone calls). Indeed, this distinction should make clear that self-selection is a problem unique to the social sciences and possibly a narrow problem at that. Most experiments are intended to understand processes where units are nonrandomly assigned to treatment by outside agents (e.g., Nature, government, other people). Situations where a human actor assigns a given unit to treatment falls outside the scope of self-selection, except to the extent that agent's choice affects their ongoing engagement with the treated unit (e.g., a physician allocating a preferred

ment from among a choice set of alternatives (or merely believing that one has made a choice of a given treatment) might lead a self-selecting assignment mechanism to not only produce a different pattern of missing manifest outcomes data across units but also reveal entirely distinct potential outcomes from those that the units would have displayed if forcibly assigned to the treatment by an outside agent. (I examine this more formally in the next section.) When people can make choices, the act of choosing and the set of alternatives from which one can choose might affect what outcome people experience. Indeed, if we think people have the same outcomes when randomly assigned to treatment versus when they choose that treatment, then there is little reason to incorporate choice into an experimental design at all.

Consider it this way: when an individual exogenously receives a treatment (e.g., a clip from Fox News) randomly versus nonrandomly in an experimental setting, the potential outcome they demonstrate post-treatment should be the same (provided they are blind to the particular treatment assignment mechanism) because they are ignorant to how they were assigned either way and are ignorant of the broader choice set. The experiment looks the same to them regardless of assignment mechanism. But being assigned to watch Fox News and choosing to watch Fox News are quite different subjective experiences. Thus even though the treatment appears the same (in each case, the same Fox News clip is watched), the outcomes associated with the self-selected and externally assigned experience of that treatment cab be different, potentially leading to differences in the individual's level of trust. The self-selected clip and the externally assigned clip can then essentially be seen as different treatments even though their content is identical.

Why might choice affect individuals' outcomes? Making choices among alternatives alters humans' perceptions of those choices and affects broader psychological factors such as affect and self-regulation (see, for a review, Iyengar and Lepper 2002).

medical treatment; see, for a review, King et al. 2005). Similarly, even choices that appear to be made by one individual might easily be made by another — with only one television, all members of a household might be constrained by the media choices of whomever holds the remote.

Making choices alters individuals' preferences over the alternatives in the original choice set, and those changes are durable over time (Sharot et al. 2012). Making choices can both increase one's confidence in the chosen alternative (Brownstein, Read, and Simon 2004) and bolster one's satisfaction (Iyengar and Lepper 2000). Certain choice situations can also produce counterfactual thinking (Botti, Orfali, and Iyengar 2009; Hafner, White, and Handley 2012), reduced self-control (Vohs et al. 2008), and post-decisional regret (Zeelenberg et al. 1998). Perceptions of anticipated regret can alter what choices are initially made and reactions of regret and disappointment with a given choice can alter how people cope with their choices (Zeelenberg and van Dijk 2010; Carmon, Wertenbroch, and Zeelenberg 2003). Different choice sets can further alter peoples' perceptions of available alternatives and their choices (see Iyengar 2010; Greifeneder, Scheibehenne, and Kleber 2010). These issues — including the factors that drive choices — have been extensively discussed in research on patients' preferences over alternative medical interventions but have not had a significant impact on social science literatures (Sidani, Miranda, Epstein, and Fox 2009; Sidani, Miranda, Epstein, Bootzin, Cousins, and Moritz 2009; Burke et al. 2008; Clark et al. 2008; Floyd and Moyer 2010).[12] In short, choices seem to impact various outcomes, so there is good reason to believe that outcome equivalence is not always a plausible assumption.

2 Defining and Estimating Self-Selection Effects

If the choice set and the act of choosing matter (that is, if the fixed types and/or equivalence assumptions are violated), then the researcher must explicitly consider what po-

[12]One possible reaction to this argument about choices and their effects is that most units for most treatments are not going to vary in their potential outcomes depending only on whether treatment was externally assigned or self-selected. In response, I say maybe not for all treatments and all units, but the psychological literature has shown effects of choice on gambling, consumer behavior, medical decisions, and in other areas (see Botti and Iyengar 2006; Iyengar 2010). Political phenomena should not be immune to these moderating effects of assignment mechanism. And understanding the effects of choices have important policy and normative implications to the extent that individuals' may not be inclined to choose a treatment that produces the "best" outcome(s) for them or others (Botti and Iyengar 2006; Burke et al. 2008).

Table 1: Assumptions and their Implications

		Outcome Equivalence	
		Yes	*No*
Fixed Types	*Yes*	*Goal*: Estimate single value of α *Strategy*: Use a hybrid design	*Goal*: Estimate "choice bonus" *Strategy*: Use a hybrid design *and* matching
	No	*Goal*: Identify variation in α across contexts *Strategy*: Use an encouragement design	*Goal*: Identify variation in α and outcomes *Strategy*: Use encouragement *and* matching

tential outcomes and related causal estimands are of interest, how those outcomes might be observed experimentally, and what analytic techniques are necessary for estimating of causal effects. If individuals do not have fixed types and choosing modifies potential outcomes, the number of potential outcomes for each individual can be quite large for even a relatively small set of possible treatments. And where violations of the fixed types assumption principally have implications for research design, violations of outcome equivalence have implications for whether it is even possible to study a reality defined by self-selection within the confines of captive, random-assignment experiments.

Table 1 summarizes the implications of the fixed types an outcome equivalence assumptions. The remainder of the paper elaborates this table in detail, connecting potential outcomes under each combination of assumptions to a set of causal estimands and an experimental technique capable of estimating them.

2.1 Notation

Scholarship since Holland (1986) tends to define causal effects of a given treatment variable, x, in terms of two potential outcomes that might manifest for a given unit, i: (1) Y_{iT}, if the unit receives treatment ($x_i = T$), and (2) Y_{iC}, if the unit receives control ($x_i = C$).

In the simplest case of self-selection (where individuals choose only between treatment and control), an expansion of this notation is necessary to account for different potential outcomes under external and self-selecting assignment mechanisms. I refer, hereafter, to the canonical outcomes as Y_{iT_E} and Y_{iC_E} for outcomes under external assignment to T or C. I then refer to $Y_{iT_{S|Z}}$ and $Y_{iC_{S|Z}}$ as the potential outcomes for Y for self-selecting into T and C, respectively, from a choice set, Z, composed of $\{T, C\}$. Under the fixed types assumption, only one of $Y_{iT_{S|Z}} or Y_{iC_{S|Z}}$ exists meaning outcomes could be labeled simply $Y_{iS|Z}$, but if any individual factors or contextual variations affect choices — thus violating fixed types — that notation would be too elementary. Under the outcome equivalence assumption, $Y_{iT_E} = Y_{iT_{S|Z}}$ and $Y_{iC_E} = Y_{iC_{S|Z}}$ for all individuals i, such that any heterogeneity in causal effects among those selecting T and C could be seen in a randomized experiment without any self-selection component, eliminating any need for the hybrid design proposed by Gaines and Kuklinski (2011).[13]

When accounting for self-selection, effects beyond the SATE might be defined. Notably, one might answer the classic observational research question: "What is the difference in the outcome between those units that select the treatment and those that choose the control?" by drawing a naive estimate of the causal effect, estimated from groups self-selecting into treatment and control:

$$\widehat{SATE}_{naive} = E[Y_{T_{S|Z}}] - E[Y_{C_{S|Z}}] \quad (3)$$

As Holland (1986) has made clear, this is not a causal effect at all, except under fairly strict assumptions. Simply observing the difference in levels of trust among those that self-select Fox and self-select the NewsHour is, obviously, not a causal effect. And then

[13]Furthermore, if potential outcomes depend on the contents of the *treatment choice set*, then each level of the putatively causal variable, x, may have multiple $Y_{T_{S|Z}}$'s depending on the alternative treatments selected and not selected. That is, self-selecting A from {A,B} may produce a different potential outcome from self-selecting A from {A,B,C} or {A,B} from either {A,B,} or {A,B,C}. Thus, in order to make sense of the manifest outcomes $Y_{T_{S|Z}}$ and $Y_{C_{S|Z}}$, one needs to understand not only the self-selection mechanism, but also the treatment alternatives available in each unit's treatment choice set.

there is the “choice effect” from Equation (2), but that effect only has individual-level meaning when the fixed types assumption holds. There is more that can be said about the effects of self-selected treatments.

3 Causal Effects of Self-Selected Treatments

Our ability to draw inferences about causal effects of self-selected treatments depends on the plausibility of the fixed types and outcome equivalence assumptions. The researcher's objectives, and thus the research design developed, depend quite directly on the combination of assumptions one is willing to make. While any randomized experiment allows the researcher to estimate an SATE, the meaning of effect estimates from choice-based experiments depends on the assumptions invoked. Indeed, all of the typically estimated effects in choice-based experiments — those directly comparing a self-selection condition to a randomized condition — only have individual-level interpretations when both assumptions hold. Without fixed types or outcome equivalence, such comparisons do not have direct, individual-level causal interpretations. When those assumptions are relaxed, however, we can alternatively invoke the conditional ignorability of choices in order to understand variation in choices across contexts (while retaining an outcome equivalence assumption), variation in outcomes when receiving a (non-)preferred treatment (while retaining the fixed types assumption), or both variation in choices and variations in potential outcomes across those choices (while relaxing both assumptions). These implications are laid out in Table 1.

In the remainder of the article, I discuss design choices and estimable effects that follow from these assumptions. First, I discuss effects that can be estimated solely from the hybrid design (two of which we have already discussed). I then discuss violations of fixed types and introduce several encouragement techniques for gaining leverage on variations in α. Next I discuss violations of outcome equivalence, through which I pro-

Table 2: Causal Effects of Self-Selection

Causal Effect	Estimator	Assumptions: Fixed Types	Assumptions: Outcome Equivalence	Assumptions: Conditional Ignorability
SATE	$E[Y_{T_E}] - E[Y_{C_E}]$			
Naive	$E[Y_{T_S}] - E[Y_{C_S}]$			
Gaines–Kuklinski's $\hat{t}_s$	$\frac{1}{\hat{\alpha}}(E[Y_S] - E[Y_{C_E}])$	✓	✓	
Gaines–Kuklinski's $\hat{t}_n$	$\frac{1}{1-\hat{\alpha}}(E[Y_{T_E}] - E[Y_S])$	✓	✓	
"Availability"	$E[Y_S] - E[Y_{C_E}]$	?	✓	
"Opt-Out"	$E[Y_S] - E[Y_{T_E}]$	?	✓	
"Treatment Removal"	$E[Y_{C_E}] - E[Y_S]$	?	✓	
"Status Quo"	$E[Y_{T_E}] - E[Y_S]$	?	✓	
Treatment Equivalence	$E[Y_{T_{S\|Z}} \| \mathbf{W}_{T_{S\|Z}}] - E[Y_{T_E} \| \mathbf{W}_{T_{S\|Z}}]$			✓
	$E[Y_{T_{S\|Z_1}} \| \mathbf{W}_{T_{S\|Z_1}}] - E[Y_{T_{S\|Z_2}} \| \mathbf{W}_{T_{S\|Z_2}}]$			✓
Control Equivalence	$E[Y_{C_{S\|Z}} \| \mathbf{W}_{C_{S\|Z}}] - E[Y_{C_E} \| \mathbf{W}_{C_{S\|Z}}]$			✓
	$E[Y_{C_{S\|Z_1}} \| \mathbf{W}_{C_{S\|Z_1}}] - E[Y_{C_{S\|Z_2}} \| \mathbf{W}_{C_{S\|Z_2}}]$			✓
Matched SATT	$E[Y_{T_{S\|Z}} \| \mathbf{W}_{T_{S\|Z}}] - E[Y_{C_{S\|Z}} \| \mathbf{W}_{T_{S\|Z}}]$		✓	✓
	$E[Y_{T_{S\|Z}} \| \mathbf{W}_{T_{S\|Z}}] - E[Y_{C_E} \| \mathbf{W}_{T_{S\|Z}}]$		✓	✓
Matched SATC	$E[Y_{T_{S\|Z}} \| \mathbf{W}_{C_{S\|Z}}] - E[Y_{C_{S\|Z}} \| \mathbf{W}_{C_{S\|Z}}]$		✓	✓
	$E[Y_{T_E} \| \mathbf{W}_{C_{S\|Z}}] - E[Y_{C_{S\|Z}} \| \mathbf{W}_{C_{S\|Z}}]$		✓	✓

pose to trade-off outcome equivalence for the conditional ignorability of choices. I finally discuss those effects that might be of interest when both fixed types and outcome equivalence are violated. Table 2 reports the universe of these causal effects and the assumptions necessary to estimate them.

3.1 Causal Effects with Satisfied Assumptions

Looking at Table 2, the first two effects (the SATE and the naive effect) have already been introduced above, the former requiring a simple two-condition experimental design and the latter requiring only observation of choice behavior and its effects. The next two effects, described by Gaines and Kuklinski (2011), leverage the hybrid design but invoke the fixed types and outcome equivalence assumptions. They show that the hybrid design's choice condition reveals the proportion of units, α, that self-select into the treatment and the remaining proportion that do not, $1 - \alpha$. Given the proportion of would-be self-selectors who opt-in to the treatment is expected to be identical in both the observational and experimental sub-experiments (due to randomization-induced balance and the fixed types assumption), the experimental ATE can be reconfigured as a weighted average of two effects for the two types of people (those who self-select the treatment, t_T, and those who do not, $t_{\neg T}$), such that: $SATE = \alpha t_T + (1 - \alpha) t_{\neg T}$. Therefore, t_s (also called the sample average treatment effect among the treated, SATT) estimated from experimental data is defined as:

$$\hat{t}_T = \frac{E[Y_S] - E[Y_{C_E}]}{\hat{\alpha}} \tag{4}$$

Partitioning the sample in this way examines any heterogeneity in the causal effect of treatment among individuals of fixed types T and T, essentially two subsample ATEs. The estimator of this effects, for example, take the proportion of individuals choosing Fox News from the choice condition in order to make sense of the results from the randomized conditions. Estimating $\hat{\alpha}$ requires the fixed types assumption and using it to estimate a choice effect requires potential outcome equivalence (because these two effects are only well-defined at the individual level if the potential outcomes for an individual choosing treatment (control) and being assignment treatment (control) are identical.

Indeed, we can see that the numerator of Equation (4) is the "choice effect" introduced in Equation (2). More accurately, we might call this an *availability effect* or intention-to-treat effect. This effect is meaningful in situations of unobserved (or unobservable) self-selection because it is informative about how offering but not inducing a treatment produces an aggregate (sample-level) change in an outcome. The availability effect is interesting from the perspective of self-selection to the extent that many interventions (in domains of politics, health, education, etc.) can rarely be forced upon individuals. But, it is a sample-level aggregate effect that has no individual-level meaning without the outcome equivalence assumption, which we can see by expressing the availability effect (which is also the numerator of Equation (4)) as a weighted sum of effects among fixed types that would choose treatment, T, and those who would not, $\neg T$:

$$\begin{aligned}\widehat{TE}_S &= E[Y_S] - E[Y_{C_E}] \\ &= \alpha(E[Y_{T_S}|T] - E[Y_{C_E}|T]) + \\ &\qquad (1-\alpha)(E[Y_{C_S}|\neg T] - E[Y_{C_E}|\neg T])\end{aligned}$$

Under outcome equivalence, the final line showing the difference in outcomes between non-choosers in the self-selection and control conditions, $(E[Y_{C_S}|\neg T] - E[Y_{C_E}|\neg T])$, must be zero because their outcomes are equivalent under both self-selection and external assignment and they are equally represented in both conditions due to random assignment. Thus, under outcome equivalence, the availability effect (and, by extension, Equation (4)) is simply a comparison of T-type people receiving treatment and T-type people receiving control.[14] This means that the physical act of self-selection is irrelevant to revealing appropriate potential outcomes, such that the researcher simply needs to estimate the proportion of T-type units, $\hat{\alpha}$, and randomly assign those individuals to treatment and control (either by using the hybrid design to estimate $\hat{\alpha}$ or estimating $\hat{\alpha}$ directly by asking participants for their choice). But if outcome equivalence is violated, the availability effect (and by extension, Equation (4)) reflects a combination of causal effect for individuals of type T and any difference in potential outcomes for those individuals (and for individuals of type $\neg T$) under the choice and external assignment conditions (i.e., a "choice bonus" due to receiving a preferred treatment).

Gaines and Kuklinski (2011) also propose an effect analogous to Equation (4) for

[14] An implication is that under outcome equivalence, Equation (4) (Gaines and Kuklinski's effect for type-T units) is identical to Equation (2) (the availability effect).

individuals of type $\neg T$:

$$\hat{t}_{\neg T} = \frac{E[Y_{T_E}] - E[Y_S]}{(1-\hat{\alpha})} \tag{5}$$

For the same reasons already stated about Equation (4), this is a well-defined individual-level only under fixed types and outcome equivalence. The numerator of this effect has interpretation quite similar to the availability effect: representing the sample-level change in the outcome were units allowed to not take a treatment they would otherwise be forced to experience (i.e., an aggregate *opt-out effect*).[15] This effect is useful for understanding the difference in an outcome between a universal mandate to take some treatment compared to an opportunity to self-select out of that treatment, e.g., the effect on disease rates of loosening childhood vaccination requirements. Reversing the order of terms in the availability and opt-out effects have similar, though opposite, causal interpretations. For example, the *treatment removal effect* simply reverses the order of the terms in the availability effect. Quite similarly, Manski and Nagin (1998) describe an aggregate effect — termed the *status quo effect* — which is simply the reverse of the opt-out effect. All four of these effects are listed in the third panel of Table 2.

Any of these four effects could be seen as a "choice effect" (i.e., a comparison between a universe where treatments are assigned and a universe where some choice is provided). Deciding what precisely we want to know about the effects of self-selection, and what assumptions we are willing to make, is critical before jumping into any particular analysis. It seems presumptuous to assume that the availability effect is always what we want to know, particularly because this set of effects only have an individual-level meaning when the fixed types and outcome equivalence assumptions hold. If we further relax the fixed types assumption, it becomes clear that these effects are not just four effects but a universe of four families of sample-level effects that will differ accord-

[15]Under outcome equivalence, Equation (5) (Gaines and Kuklinski's effect for type-$\neg T$ units) is identical to this opt-out effect.

ing to the proportion of individuals, $\hat{\alpha}$, choosing T. We can observe this variation in effects using encouragement (see, for example, Albertson and Lawrence 2009).

3.2 Testing the Fixed Types Assumption with Encouragement

If fixed types is violated but we retain a potential outcome equivalence assumption (i.e., we find ourselves in the lower-left cell of Table 1), we aim to find different estimates of $\hat{\alpha}$ by which to understand how inducing different (numbers of) individuals to choose T impacts the aggregate effect of treatment provision. In other words, we simply need to identify choices but we can assume response to an assigned and preferred treatment is the same as response to choosing that preferred treatment. At least three types of encouragement technique can become essential tools for causal inference.

The first, "pretreatment encouragement," should be familiar to experimentalists and involves manipulation of a particular pretreatment factor thought to affect treatment self-selection. This approach produces a randomized instrument for individuals' self-selections, thereby enabling comparisons not only between self-selection within a given choice set and assignment to any of that set's treatments, but also comparisons within the same choice set across different levels of that initial encouragement instrument. For example, Leeper (2012*a*) manipulates attitude importance and finds that information choices (and downstream outcomes) differ considerably after receiving the high and low importance manipulations. Treatment effects can be estimated among each of the initial treatment groups, ignoring the intervening treatment selections or, if the sequential ignorability assumption described by Imai et al. (2011) is met, the intervening treatment self-selections can also be modeled as mediators between the encouragement and the outcomes. This approach can enable further tests of outcome equivalence and, most obviously, will demonstrate whether fixed types assumptions are at all plausible in that choice context.

The second technique, "choice set encouragement," entails manipulation of which treatment alternatives are made available in a choice set. Here, multiple choice set environments serve as treatments. For example, environments that are composed of ideologically biased sets of information might encourage participants to choose liberal or conservative messages, respectively, due to the higher proportion of liberal (conservative) messages in the choice set (see, for example, Leeper 2012*b*; Sheagley 2012). This technique helps explain how treatments are self-selected and what effects particular choices have under alternative conditions, but requires the specification of additional potential outcomes (due to a larger number of possible choices) and tests of outcome equivalence for the same treatment chosen from different choice sets.

A third technique, "choice value encouragement," involves changing the apparent value, desirability, or cost of alternative treatments within the choice set (see, for two different approaches, Botti and Iyengar 2004; Boudreau 2011). Making particular choices more desirable or beneficial than others will help to explore who among selectors (or non-selectors) alters their choices when the incentive structures change. Like the Hovland design, this approach looks at effects of treatment among self-selecting audiences, but the motive to select a particular treatment is additionally manipulated in order to facilitate causal inference. The altered value of the alternative treatments makes those assigned to a particular treatment choice set more or less likely to select in a given way.

3.3 Testing for Outcome Equivalence

The three groups of the hybrid design involve exogenous assignment to treatment, control, or choice among the two, but the design itself does not provide the leverage necessary to estimate effects when the outcome equivalence assumption is violated. This is because all of the comparisons between the choice condition and the random assignment condition require that units manifest the same outcome regardless of how they are assigned to a given treatment (as we saw above in the discussion of the availability effect).

If outcome equivalence is violated,[16] we might find ourselves in the upper-right cell of Table 1 and be interested in whether there is some kind of "choice bonus" associated with an individual receiving their preferred treatment (essentially a form of placebo effect associated with the feeling of having made a choice). But we also relax fixed types, in order to suggest that who experiences this bonus (i.e., $\hat{\alpha}$) can also vary contextually, then we might find ourselves in the lower-right cell of Table 1. Either way, to gain any causal leverage at all, we need to tradeoff the outcome equivalence assumption for an assumption of conditional ignorability of self-selected treatments.

Because self-selection involves choices, a choice situation is unlikely to ever satisfy the "independence" assumption underlying experimental causal inference (Holland 1986, 948-949). This assumption requires that treatment assignment be independent of potential outcomes. Choices are, by definition, not exogenous assignments to treatment, so potential outcomes of self-selection into versus out of treatment are never going to be independent of treatment assignment. The only available approach is to attempt to make treatment assignment ignorable so that individuals who made one choice can be compared to similar others who made a different choice or who randomly received a treatment.[17] For those unconvinced by extant methods for circumventing independence

[16]Outcome nonequivalence could take many forms. First, it is possible that self-selecting has a "bonus"-type effect, with a constant difference between outcomes for any treatment under choice versus external assignment. This bonus might be positive or negative, depending on whether making a choice is beneficial or detrimental. Second, there could also be a preferred treatment effect, whereby choosing one's preferred treatment has an effect over being assigned that treatment but choosing a nonpreferred treatment is no different from simply being assigned to it. Finally, we might expect some kind of heterogeneity, whereby choice magnifies an effect with those choosing a beneficial outcome obtaining more positive scores than those assigned to the same treatment and those choosing a less beneficial outcome becoming more negative than those assigned to the same treatment. There are likely other possibilities. The medical literature has extensively discussed the "preference effect" — that is, the additional (positive) effect of receiving a preferred treatment or the (negative) effect of receiving a non-preferred treatment (Rucker 1989; Halpern 2003; McCaffery et al. 2011).

[17]Essentially, self-selection problems require considerable information about the factors driving treatment choice. Rubin (1978) points out the difficulty of obtaining ignorability when treatments are self-selected: "If the assignment mechanism depends on (X,Y) values, then those values must be recorded by the recording mechanism if the assignment mechanism is to be ignorable [. . .] if patients select the treatments themselves on the basis of their unrecorded opinions of their health, the assignment mechanism is not ignorable [. . .] for a particular assignment mechanism, one can choose a recording mechanism that makes the assignment mechanism not ignorable [. . .] The more involved the assignment mechanism (in the sense of depending on more values), the more complete must be the recording mechanism if the

(namely, matching or regression, which invoke conditional ignorability, see Rosenbaum and Rubin 1983), then causal inference about self-selected treatments is only possible when the outcome equivalence assumption holds.

Given the need for covariate data with which to attempt to satisfy this strong ignorability assumption (Rosenbaum and Rubin 1983), matching methods seem an obvious solution to identifying causal effects of self-selection. While matching as a general approach to causal inference has raised concerns when applied to the study of observational data (LaLonde 1986; Arceneaux, Gerber, and Green 2006, 2010),[18] applying matching to the hybrid design is plausibly more robust. In expectation, random assignment guarantees covariate balance across the three hybrid design conditions so matching between choice subgroups (those choosing one treatment or the other) and comparable others in the random assignment conditions can be assured (in expectation) to find reasonable comparison units. Introducing the hybrid design with a conditional ignorability assumption (and associated analytic technique, such as matching) further allows one to compare the manifest outcomes for those self-selecting one treatment alternative to the outcomes of similar units externally assigned to that same treatment. Indeed, if we are willing to temporarily accept conditional ignorability, we can actually test whether outcome equivalence is plausible by comparing self-selected and exogenously assigned groups, conditional on observable covariates.

The next panel of Table 2 highlights two comparisons that test the equivalence of outcomes for treatment selectors between external assignment and self-selection and between self-selection from different choice sets, as well as two analogous comparisons of outcomes for treatment non-selectors. As an example, potential outcome equivalence (between external assignment and self-selection of treatment) can then be tested by estimating the comparison $E[Y_{T_{S|Z}}|\mathbf{W}] - E[Y_{T_E}|\mathbf{W}]$, where $\mathbf{W}$ is some matrix of unit

assignment mechanism is to be ignorable" (42).

[18] Also, a large number of alternative matching techniques exist, each with its own merits, advocates, and critics (see Sekhon 2009) and I focus here on estimands rather than particular matching estimators.

characteristics used to match individuals across conditions.[19] This *treatment equivalence* effect is analogous to a *control equivalence* effect. When zero, there is no individual-level "choice bonus," meaning that outcome equivalence holds and there is thus no analytic advantage to actually having individuals self-select into treatment because their outcomes in response to the externally assigned form of the treatment is the same as having self-selected that treatment. The randomized design alone thus reveals all the necessary potential outcomes and the researcher could simply ask which treatment each individual would choose (to estimate $\hat{\alpha}$, as above).

If outcome equivalence holds, it is also possible to use the self-selection condition of the hybrid design to directly estimate a sample average treatment effect for the treated (SATT) and sample average treatment effect for the control (SATC). Indeed, both can be estimated in two ways. For the SATT, outcomes for those self-selecting treatment could be compared either to comparable portion of the the group self-selecting control or to the comparable portion of the group externally assigned to control. Similarly, for the SATC, outcomes for those self-selecting control could be compared either to the comparable portion of the group self-selecting treatment or to the comparable portion of the group externally assigned to treatment. When outcome equivalence is violated, the two different estimates (of each of the SATT and SATC) should not be equal because, as in any violation of outcome equivalence, the estimates reflect both a causal effect and a

[19]There are three ways to think about **W**. The first is as a matrix of pretreatment covariates, as in a typical matching procedure. Of course, common covariate support, a sufficiently large sample size, and a large number of selection-relevant covariates are all necessary to obtaining ignorability (see Shadish, Clark, and Steiner 2008; Cook, Shadish, and Wong 2008; Steiner et al. 2010; Steiner, Cook, and Shadish 2011). The second is as individuals' stated preferences over treatment alternatives, as in a patient preference trial. In these cases, stated choice is expected to encompass all choice-relevant information. If the fixed types assumption does not hold, this may be a less plausible way of defining **W** because choices are functions of covariates that vary within individuals across contexts. The third way is to think of **W** as a continuous measure of stated or revealed preference that places individuals on a spectrum of preference for treatment or control, either as directly reported by experimental units or as estimated from a propensity score model. Encouragement, discussed in the next section, also has the effect of reducing the number of relevant covariates to condition on in order to satisfy conditional ignorability. As research on the self-selection of any particular treatment evolves, evidence of strong relationships between a particular covariate and self-selection of a particular treatment can be used both as measured covariates but also as blocking factors.

"choice bonus."[20]

4 Conclusion

Historically, observational research has dominated political science (see Druckman et al. 2006). Yet the growth of experimental research has also brought with it a rapid increase in the study of self-selected treatments, particularly in the area of political communication. Choice-based experimental designs have face validity because they assign units to treatment as they actually are in the real world (via self-selection) rather than through the artificiality of captive and random treatment assignment. This paper has argued that the study of self-selected treatments requires novel experimental methods but also requires researchers to clearly articulate the causal estimand(s) of interest to their research and carefully design experimental studies to satisfy the assumptions involved. While the assumptions necessary to draw inferences from choice-based experiments can be quite strong, an array of causal questions can be answered with relatively simple experimental design and analysis even when those assumptions do not hold.

It is also worth noting that the importance of the issues raised here reaches beyond the domain of political communication, where self-selection is prevalent and concerns about its implications are common. For example, what effect does choosing public versus private school have on the effect of attending those schools among the general population and the populations likely to select one versus the other? How do differences in the available choices affect which school is chosen and the effects that choice has on outcomes in the chosen school? Does choosing a particular brand of product alter the effect of using the brand on satisfaction and future willingness to purchase?

[20] A final set of effects that could be estimated involve comparisons of units' outcomes after receiving a treatment different from the one that they would have self-selected. For example, one might be interested to know the effect of (forcibly) assigning the treatment to those who are inclined to self-select into control, like forcing Fox News viewers to watch the PBS NewsHour. I leave thorough consideration of these effects to future research, but note that the definition and estimation thereof would follow logically from the effects just described. Again, if potential outcome equivalence holds, all of these effects can simply be estimated with a randomized experiment and there is no need to rely on a choice-based design.

Do various features of group deliberations change when groups are randomly constituted versus self-selected, controlling for characteristics that predict selection? These are only a few of the social research questions that ask about the comparative effects of self-selected versus externally assigned treatments. Answering these questions requires clear definition of the causal effects of interest, clever research design to uncover the requisite potential outcomes, and analyses that respect the underlying causal processes at work. The hybrid design plus matching approach provides additional tools in the political and social researcher's toolkit that can be used together with existing techniques — such as encouragement, blocking, instrumental variables, and mediation analysis — to clarify causal processes prone to self-selection better than "gold standard" randomized experiments.

References

Albertson, Bethany L., and Adria Lawrence. 2009. "After the Credits Roll: The Long-Term Effects of Educational Television on Public Knowledge and Attitudes." *American Politics Research* 37(2): 275–300.

Arceneaux, Kevin, Alan S. Gerber, and Donald P. Green. 2006. "Comparing Experimental and Matching Methods Using a Large-Scale Field Experiment on Voter Mobilization." *Political Analysis* 14(1): 37–62.

Arceneaux, Kevin, Alan S. Gerber, and Donald P. Green. 2010. "A Cautionary Note on the Use of Matching to Estimate Causal Effects: An Empirical Example Comparing Matching Estimates to an Experimental Benchmark." *Sociological Methods & Research* 39(2): 256–282.

Arceneaux, Kevin, and Martin Johnson. 2007. "Does Media Choice Minimize Videomalaise? Selective Exposure and the Effect of Talk Shows on Political Trust." Paper presented at the Annual Meeting of the Midwest Political Science Association, Chicago, IL.

Arceneaux, Kevin, and Martin Johnson. 2011. "Does Media Fragmentation Produce Mass Polarization? Selective Exposure and a New Era of Minimal Effects." Unpublished paper, Temple University.

Arceneaux, Kevin, and Martin Johnson. 2012. *Changing Minds or Changing Channels? Media Effects in the Era of Viewer Choice*.

Bennett, W. Lance, and Shanto Iyengar. 2008. "A New Era of Minimal Effects? The

Changing Foundations of Political Communication." *Journal of Communication* 58(4): 707–731.

Botti, Simona, and Sheena S. Iyengar. 2004. "The Psychological Pleasure and Pain of Choosing: When People Prefer Choosing at the Cost of Subsequent Outcome Satisfaction." *Journal of Personality and Social Psychology* 87(3): 312–326.

Botti, Simona, and Sheena S. Iyengar. 2006. "The Dark Side of Choice: When Choice Impairs Social Welfare." *Journal of Public Policy & Marketing* 25(1): 24–38.

Botti, Simona, Kristina Orfali, and Sheena S. Iyengar. 2009. "Tragic Choices: Autonomy and Emotional Responses to Medical Decisions." *Journal of Consumer Research* 36(3): 337–352.

Boudreau, Cheryl. 2011. "The Market for Political Information: How the Consumption of Information Affects Citizens' Decisions." Unpublished paper, University of California, Davis.

Brownstein, Aaron L., Stephen J. Read, and Dan. Simon. 2004. "Bias at the Racetrack: Effects of Individual Expertise and Task Importance on Predecision Reevaluation of Alternatives." *Personality & Social Psychology Bulletin* 30(7): 891–904.

Burke, L.E., M. Warziski, M.A. Styn, E. Music, A.G. Hudson, and S.M. Sereika. 2008. "A Randomized Clinical Trial of a Standard Versus Vegetarian Diet for Weight Loss: The Impact of Treatment Preference." *International Journal of Obesity* 32(1): 166–176.

Carmon, Ziv, Klaus Wertenbroch, and Marcel Zeelenberg. 2003. "Option Attachment: When Deliberating Makes Choosing Feel Like Losing." *Journal of Consumer Research* 30(1): 15–29.

Clark, Noreen M., Nancy K. Janz, Julia A. Dodge, Lori Mosca, Xihong Lin, Qi Long, Roderick J.A. Little, John R.C. Wheeler, Steven Keteyian, and Jersey Liang. 2008. "The Effect of Patient Choice of Intervention on Health Outcomes." *Contemporary Cinical Trials* 29(5): 679–686.

Cook, Thomas D., William R. Shadish, and Vivian C. Wong. 2008. "Three Conditions Under Which Experiments and Observational Studies Produce Comparable Causal Estimates: New Findings from Within-Study Comparisons." *Journal of Policy Analysis and Management* 27(4): 724–750.

Corrigan, Patrick W., and Mark S. Salzer. 2003. "The Conflict Between Random Assignment and Treatment Preference: Implications for Internal Validity." *Evaluation and Program Planning* 26(2): 109–121.

Druckman, James N., Donald P. Green, James H. Kuklinski, and Arthur Lupia. 2006. "The Growth and Development of Experimental Research in Political Science." *American Political Science Review* 100(4): 627–635.

Druckman, James N., Jordan Fein, and Thomas J. Leeper. 2012. "A Source of Bias in Public Opinion Stability." *American Political Science Review* 106(2): 430–454.

Feldman, Lauren, Natalie Jomini Stroud, Bruce Bimber, and Magdalena Wojcieszak. 2013. "Assessing Selective Exposure in Experiments: The Implications of Different Methodological Choices." *Communication Methods and Measures* (August): 37–41.

Floyd, Anna H L, and Anne Moyer. 2010. "Effects of Participant Preferences in Unblinded Randomized Controlled Trials." *Journal of Empirical Research on Human Research Ethics* 5(2): 81–93.

Gaines, Brian J., and James H. Kuklinski. 2011. "Experimental Estimation of Heterogeneous Treatment Effects Related to Self-Selection." *American Journal of Political Science* 55(3): 724–736.

Geer, John G., Richard R. Lau, and Lynn Vavreck. 2012. "Changing Times: Political Advertising and Information Seeking in an Era of Viewer Choice." Unpublished paper, Vanderbilt University.

Greifeneder, Rainer, Benjamin Scheibehenne, and Nina Kleber. 2010. "Less May Be More When Choosing Is Difficult: Choice Complexity and Too Much Choice." *Acta Psychologica* 133(1): 45–50.

Hafner, Rebecca J., Mathew P. White, and Simon J. Handley. 2012. "Spoilt for Choice: The Role of Counterfactual Thinking in the Excess Choice and Reversibility Paradoxes." *Journal of Experimental Social Psychology* 48(1): 28–36.

Halpern, Scott D. 2003. "Evaluating Preference Effects in Partially Unblinded, Randomized Clinical Trials." *Journal of Clinical Epidemiology* 56(2): 109–115.

Holland, Paul W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81(396): 945–960.

Hovland, Carl I. 1959. "Reconciling Conflicting Results Derived from Experimental and Survey Studies of Attitude Change." *American Psychologist* 14(1): 8–17.

Imai, Kosuke, Luke Keele, Dustin Tingley, and Teppei Yamamoto. 2011. "Unpacking the Black Box: Learning about Causal Mechanisms from Experimental and Observational Studies." *American Political Science Review* 105(4): 765–789.

Iyengar, Sheena S. 2010. *The Art of Choosing*. New York: Grand Central Publishing.

Iyengar, Sheena S., and Mark R. Lepper. 2000. "When Choice is Demotivating: Can One Desire Too Much of a Good Thing?" *Journal of Personality and Social Psychology* 79(6): 995–1006.

Iyengar, Sheena S., and Mark R. Lepper. 2002. "Choice and Its Consequences: On the Costs and Benefits of Self-Determination." In *Self and Motivation: Emerging Psychological Perspectives*, ed. Abraham Tesser. Washington, DC: American Psychological Association, 71–96.

Janevic, Mary R., Nancy K. Janz, Julia A. Dodge, Xihong Lin, Wenqin Pan, Brandy R. Sinco, and Noreen M. Clark. 2003. "The Role of Choice in Health Education Intervention Trials: A Review and Case Study." *Social Science & Medicine* 56(7): 1581–

1594.

Johnson, Martin, and Kevin Arceneaux. N.d. "Who Watches Political Talk? Revisiting Political Television Reception with Behavioral Measures." Unpublished paper, University of California, Riverside.

Joseph, Patricia Hallam. 2012. "Value Preference, the Framing Process, and Information Choices: The Route from Values to News Frame Selection." Paper presented at the Annual Meeting of the Midwest Political Science Association, Chicago, IL.

King, Michael, Irwin Nazareth, Fiona Lampe, Peter Bower, Martin Chandler, Maria Morou, Bonnie Sibbald, and Rosalind Lai. 2005. "Impact of Participant and Physician Intervention Preferences on Randomized Trials: A Systematic Review." *Journal of the American Medical Association* 293(9): 1089–1099.

LaLonde, Robert J. 1986. "Evaluating the Econometric Evaluations of Training Programs with Experimental Data." *American Economic Review* 76(4): 604–620.

Lau, Richard R. 1995. "Information Search during an Election Campaign: Introducing a Processing-Tracing Methodology for Political Scientists." In *Political Judgment: Structure and Process*, eds. Milton Lodge, and Kathleen M. McGraw. Ann Arbor, MI: University of Michigan Press.

Leeper, Thomas J. 2012*a*. "Strong Attitudes and Biased Political Engagement." Paper presented at the Annual Meeting of the Midwest Political Science Association, Chicago, IL.

Leeper, Thomas J. 2012*b*. "The Informational Basis of Mass Polarization." Paper presented at the Annual Scientific Meeting of the International Society for Political Psychology, Chicago, IL.

Levendusky, Matthew S. 2011. "Do Partisan Media Polarize Voters?" Unpublished paper, University of Pennsylvania.

Levendusky, Matthew S. 2012. *Partisan News that Matters: How Partisan News Impacts American Politics*. Chicago, IL: The University of Chicago Press.

Manski, Charles F., and Daniel S. Nagin. 1998. "Bounding Disagreements About Treatment Effectcs: A Case Study of Sentencing and Recidivism." *Sociological Methodology* 28(1): 99–137.

McCaffery, Kirsten J., Robin Turner, Petra Macaskill, Stephen D. Walter, Siew Foong Chan, and Les Irwig. 2011. "Determining the Impact of Informed Choice: Separating Treatment Effects from the Effects of Choice and Selection in Randomized Trials." *Medical Decision Making* 31(2): 229–236.

McPherson, Klim, and A. Britton. 1999. "The Impact of Patient Treatment Preferences on the Interpretation of Randomised Controlled Trials." *European Journal of Cancer* 35(11): 1598–1602.

McPherson, Klim, and A. Britton. 2001. "Preferences and Understanding Their Effects on Health." *Quality in health care : QHC* 10: i61–i66.

McPherson, Klim, Annie R. Britton, and John E. Wennberg. 1997. "Are Randomized Controlled Trials Controlled? Patient Preferences and Unblind Trials." *Journal of the Royal Society of Medicine* 90(12): 652–656.

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.

Payne, John W., James R. Bettman, and Eric J. Johnson. 1993. *The Adaptive Decision Maker*. New York: Cambridge University Press.

Rosenbaum, Paul R. 2009. *Design of Observational Studies*. New York: Springer.

Rosenbaum, Paul R., and Donald B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70(1): 41–55.

Rubin, Donald B. 1978. "Bayesian Inference for Causal Effects: The Role of Randomization." *Annals of Statistics* 6(1): 34–58.

Rubin, Donald B. 1991. "Practical Implications of Modes of Statistical Inference for Causal Effects and the Critical Role of the Assignment Mechanism." *Biometrics* 47(4): 1213–1234.

Rucker, Greta. 1989. "A Two-Stage Trial Design for Testing Treatment, Self-Selection, and Treatment Preference Effects." *Statistics in Medicine* 8: 477–485.

Sekhon, Jasjeet S. 2009. "Opiates for the Matches: Matching Methods for Causal Inference." *Annual Review of Political Science* 12: 487–508.

Shadish, William R., M.H. Clark, and Peter M. Steiner. 2008. "Can Nonrandomized Experiments Yield Accurate Answers? A Randomized Experiment Comparing Random and Nonrandom Assignments." *Journal of the American Statistical Association* 103(484): 1334–1344.

Shadish, William R., Thomas D. Cook, and Donald T. Campbell. 2001. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton-Mifflin.

Sharot, Tali, Stephen M Fleming, Xiaoyu Yu, Raphael Koster, and Raymond J Dolan. 2012. "Is Choice-Induced Preference Change Long Lasting?" *Psychological science* (August).

Sheagley, Geoffrey D. 2012. "Political Decision Making, Information Search, and Disagreeable Information." Paper presented at the Annual Meeting of the Midwest Political Science Association, Chicago, IL.

Sidani, Souraya, Joyal Miranda, Dana Epstein, and Mary Fox. 2009. "Influence of Treatment Preferences on Validity: A Review." *Canadian Journal of Nursing Research* 41(4): 52–67.

Sidani, Souraya, Joyal Miranda, Dana Epstein, Richard R. Bootzin, Jennifer Cousins, and Patricia Moritz. 2009. “Relationships Between Personal Beliefs and Treatment Acceptability, and Preferences for Behavioral Treatments.” *Behaviour Research and Therapy* 47(10): 823–829.

Steiner, Peter M., Thomas D. Cook, and William R. Shadish. 2011. “On the Importance of Reliable Covariate Measurement in Selection Bias Adjustments Using Propensity Scores.” *Journal of Educational and Behavioral Statistics* 36(2): 213–236.

Steiner, Peter M., Thomas D. Cook, William R. Shadish, and M.H. Clark. 2010. “The Importance of Covariate Selection in Controlling for Selection Bias in Observational Studies.” *Psychological Methods* 15(3): 250–267.

Swift, Joshua K., and Jennifer L. Callahan. 2009. “The Impact of Client Treatment Preferences on Outcome: A Meta-Analysis.” *Journal of Clinical Psychology* 65(4): 368–381.

Swigger, Nathaniel. 2012. “Self-Selection and the Utility of Viral Video.” Paper presented at the Annual Meeting of the Midwest Political Science Association, Chicago, IL.

Vohs, Kathleen D., Roy F. Baumeister, Brandon J. Schmeichel, Jean M. Twenge, Noelle M. Nelson, and Dianne M. Tice. 2008. “Making Choices Impairs Subsequent Self-Control: A Limited-Resource Account of Decision Making, Self-Regulation, and Active Initiative.” *Journal of Personality and Social Psychology* 94(5): 883–98.

Zeelenberg, Marcel, and Wilco W. van Dijk. 2010. “On Bad Decisions and Disconfirmed Expectancies: The Psychology of Regret and Disappointment.” *Cognition & Emotion* 14(4): 521–541.

Zeelenberg, Marcel, Wilco W. van Dijk, Joop van der Pligt, Anthony S. R. Manstead, van Empelen Pepijn, and Dimitri Reinderman. 1998. “Emotional Reactions to the Outcomes of Decisions: The Role of Counterfactual Thought in the Experience of Regret and Disappointment.” *Organizational Behavior and Human Decision Processes* 75(2): 117–41.