

Misinformation in Experimental Political Science

Matthew Barnfield

The American Political Science Association recently cautioned against the use of *misinformation* (giving research participants false information about the state of the world) in research with human subjects. This recommendation signals a growing recognition, as experimental research itself grows in prevalence in political science, that deceptive practices pose ethical problems. But what is wrong with misinformation in particular? I argue that while this question certainly has an ethical dimension, misinformation is bad for inference too. Misinformation moves us away from answering questions about the political world effectively. I propose a straightforward, intuitive solution to this twofold problem: tell the truth.

When I pretend or engage in make-believe, I close my eyes to the world around me, sometimes literally, the better to imagine a world that isn't actually there.

—Agnes Callard, *Aspiration: The Agency of Becoming* (2018, 84)

Recent decades have seen political scientists increasingly adopt experimental approaches to answering causal questions about politics (Druckman and Green 2021). Such questions are, by their nature, counterfactual: their causal logic depends on the idea that things could be otherwise, if some causal force did or did not operate (Imbens and Rubin 2015). In cases where the causal force takes the form of an informational intervention—which, in political science, it often does—this counterfactual logic can seem to imply that we need to make up some (counterfactual) information and see whether things would be otherwise if that information were true. This intuition can lead researchers down the path of *misinformation*. According to new guidelines, published in 2020 by the American Political Science Association's (APSA) Ad Hoc Committee on Human Subjects Research (ACHSR 2020, 7), misinformation is a type of “deception” in which researchers give participants “false information about the state of the world.” The guidelines advise researchers to “carefully consider” any

use of misinformation. This paper does exactly that. I argue that misinformation raises not only ethical, but also inferential concerns, and researchers would be better off telling the truth.

What Is Misinformation?

Misinformation is the provision of false information about the state of the world (ACHSR 2020, 7). Two criteria must therefore be met for an experiment to involve misinformation: false information must be provided, and that information must be about the state of the world.¹ In principle, researchers could employ misinformation for any purpose, but it is safe to assume that political scientists have little interest in carelessly feeding research subjects random pieces of misinformation for no reason—a practice that would be akin to “bullshit” (Frankfurt 1998). Rather, researchers use misinformation because they are interested in the effect of the misinformation itself. The misinformation is the experimental treatment.

Imagine we are interested in learning about the *bandwagon effect* (Barnfield 2020): the idea that the popularity of a political candidate or party makes people more likely to vote for them. We conduct an experiment: we show a random selection of people the results of a vote intention poll (the treatment) and test whether they are more likely to vote for a more popular candidate than people who do not see the information. This structure is common across political science experiments that “concern how people react to the content and format of information presented to them” (Dafoe, Zhang and Caughey 2018, 400).² We want freedom to manipulate this information—here, the different candidates’ vote shares—to ensure that we target

Matthew Barnfield  (UK) is a Postdoctoral Research Fellow in the Department of Politics at the University of Exeter, UK (m.g.barnfield@gmail.com). He began this project as a Doctoral Candidate in Political Science at Queen Mary University of London, UK.

the effect of interest. The desire to manipulate information encourages us to make up that information.

Examples of experiments that use made-up polling information as their treatment abound (e.g., Dahlgard et al. 2017; Dizney and Roskens 1962; Goitel and Shields 1994; Madson and Hillygus 2019). However, only some fit the bill for misinformation. Others invoke a purely made-up *context* to go with their made-up treatment. For instance, they invite participants to imagine they are voting in a made-up election contested between two made-up candidates. Here, as the “false information” researchers provide is not about the actual “state of the world”, they constitute more of a thought experiment, or science fiction, rather than deception.³ By misinformation, I refer to cases where the experimental context *is* the actual political world, and the treatment is false information about that context.

For example, consider the following experimental design:

A recent poll of vote intentions at the upcoming presidential election, conducted in your state of New York, put Donald Trump and Joe Biden on the following vote shares:

Trump 62%–32% Biden

Please indicate which candidate you intend to vote for.

This example constitutes misinformation because the contextual setup establishes that we are talking about the actual political world, but the treatment information is false—Biden had a large lead in the polls in New York throughout the 2020 campaign.

The Ethics of Misinformation

Misinformation is deceptive. Two common ways of framing why deception matters ethically, and of assessing experimental ethics more broadly, are the Kantian deontological view and the utilitarian view (Berghmans 2007).

The Kantian View

Kant’s ethical theory asserts that people are “ends” in themselves (Kant 1978, 1996; Korsgaard 1996). Treating a person as a “mere means,” by using them to “promote your own ends in a way to which [they] could not possibly consent,” violates their “dignity” (Korsgaard 2018, 77). This violation contravenes the “Formula of Humanity,” which demands that “we avoid all use of force, coercion, and *deception*, that is, all devices that are intended to override or redirect the free and autonomous choices of other people” (Korsgaard 2018, 78, emphasis added).

A predominant concern with deception—especially with “identity,” “motivation,” and “activity” deception (ACHSR 2020, 7)—is that it overrides participants’ autonomy by compromising informed consent (Bok 1995). In

Humphreys’ (2015, 101) terms, informed consent serves both a “diagnostic” and an “effective” function: it provides evidence that respondents were not treated as “mere means” (the diagnostic function) while also actively enhancing their autonomy (the effective function). Since deception typically involves misrepresenting “what you are doing” (activity deception), “who you are” (identity deception), and “the reasons for the research” (motivation deception), it fundamentally prevents participants knowing what they are choosing to do when they consent to participate in an experiment. Under activity deception, respondents are unaware that they are participating in research. Under identity deception, they are unaware of who is conducting that research. Under motivation deception, they are unaware why the research is being conducted. Misinformation does not necessarily involve any of these types of deception. The only deception it necessarily involves is a false piece of information about the state of the world. For we could even, in principle, tell respondents, “you are about to read the result of a completely made-up poll.” If we also state who we are, what we are doing, and why, the diagnostic and effective functions of informed consent should be fulfilled, and respondents’ dignity and autonomy therefore respected.

However, admitting the treatment information is false would compromise the “integrity” of the research design (ACHSR 2020, 8). Ideally, if we are studying what people would do if some false information were true, we should not forewarn them that it is false. But when omitting this disclaimer, we cannot assume that respondents are consenting to being deceived. They might not consent if they knew they were about to be exposed to misinformation. By participating, they are not acting in accordance with their (would-be) “self-chosen plan” (Beauchamp and Childress 2019; Phillips 2021).

Moreover, under misinformation, participants cannot autonomously make the decision they are asked to make in the experiment itself. In the bandwagon example, respondents cast a vote. Misinformation interferes with their ability to make *this* choice autonomously or rationally. And importantly, this experimental choice often mimics a real-world choice respondents will go on to make in their real lives. In experiments that ask people to express an opinion as an experimental outcome, this measure might reflect the opinion they have formed about the real-world matter at hand. By extension then, misinformation could lead people away from making autonomous, rational decisions in their lives outside of the experimental context (Phillips 2021, 284). When we treat research participants with misinformation, we deceive them into making the choice in a way that achieves our ends—say, the falsification or verification of some hypothesis—and not theirs. On the Kantian view, misinformation therefore violates respondents’ dignity by disrespecting their autonomy.

The Utilitarian View

The utilitarian view, meanwhile, emphasizes a trade-off “between the wrong or harm of deception and the benefits resulting from the research” (Berghmans 2007, 14). What is the “wrong or harm” of misinformation?

Firstly, the loss of decision-making autonomy just discussed is assimilable, in a utilitarian framework, to what Humphreys (2011) calls a “participation cost”—a direct cost to the participant incurred by taking part in the experiment. Insofar as people value their capacity for autonomy and rationality, denying them of this capacity is harmful.

Secondly though, the most obvious harms of misinformation are “process costs”—those “broader” costs “associated with the implementation of the research, including the immediate social effects of the intervention” (Humphreys 2011). One process cost of misinformation is the harm it may do to science by compromising participants’ trust in researchers. As Hertwig and Ortmann (2008, 63; see also Wilson and Hunter 2010) note, “a likely outcome of deceptive practices is participants’ future resistance to other research efforts.” Given rising rates of survey non-response and the “overextraction” (Leeper 2019) of human respondents by social research—which increasingly relies on limited groups of regular survey-takers (Krupnikov, Nam, and Style 2021)—it is wise to avoid deterring participants. Even if respondents go on to participate in future experiments, they may be distrustful of the information they receive in these experiments (Humphreys 2014; Svorenčík 2016), compromising research quality.⁴

Beyond this, the treatment effect measured in the experiment represents a potential process cost if it is based on a falsehood. In the bandwagon example, the treatment is designed to change people’s voting behavior. Participants might end up voting for Trump where they would have voted for Biden. Influencing the vote through misinformation is detrimental to democracy (Brown 2018; De Ridder 2021). Participating in misleading experiments measuring preferences for war crimes “predisposes citizens toward behavior that would violate international humanitarian law but also can undermine their understanding of what is permissible under the laws of war” (Carpenter, Montgomery, and Nylen 2021, 8). In other words, misinformation can increase confusion about, and boost support for, war crimes. Zimmerman (2016, 185–187) argues that political science experiments “often” cause “great” and “lasting” harm to “at least one person or group.”

Indeed, political science applications of misinformation seem especially likely to incur such harmful process costs, compared to other disciplines. In canonical cases of deception discussed in the psychological sciences (see, e.g., Kelman 1967), the effects only reach the individuals who participated in the experiment—that is, they tend to take the form of participation costs. By contrast,

changing someone’s vote intention alters the result of an election and thereby the livelihoods of the members of a polity (Desposato 2016, 277). Making people more supportive of war crimes shifts public opinion about whether potentially thousands of citizens should undergo atrocities (Carpenter, Montgomery, and Nylen 2021). The questions political scientists address have collective implications. Correspondingly, there is a greater “collective wrong” (Whitfield 2019, 532–534) that the benefits of research must outweigh.

The False Promise of Debriefing

Some, including APSA (ACHSR 2020, 8), propose “debriefing”—providing participants “information about the study [and] their role in it” (Desposato 2016, 282) after the fact—as a means of ameliorating the ethical pitfalls of deception (Berghmans 2007). From a Kantian perspective, when employing misinformation, researchers should debrief respondents. Debriefing is a way of showing a person respect by telling them what happened to them, even if it will change nothing in terms of concrete participation or process costs.

From a utilitarian perspective, debriefing may offset the process costs of misinformation discussed earlier. First, debriefing may help “to increase public trust and understanding of what we are doing” (Desposato 2016, 283). Arguably, confessing to having misinformed participants makes scientists seem more trustworthy than they would seem if they did not admit it. But it seems equally plausible that debriefing will *compromise* trust. Participants sometimes lack the political knowledge required to recognize that an experiment is deceiving them. If they do not learn after the experiment that the treatment was untrue, they will see no reason to believe it was. The debrief guarantees that they will find out. Debriefing, itself, could therefore incur its own process cost by damaging future research efforts (Hertwig and Ortmann 2008, 63).

Second, debriefing may offset process costs by overturning treatment effects. Suppose respondents are told after the bandwagon experiment that the poll result they saw was fake. People will see that they should not have been persuaded by the treatment, cueing them to reverse that change. This argument only holds if the treatment *independently* produces the behavior observed in the experiment. Yet when people process new or surprising information, they draw on the wider information environment, to understand or rationalize it (see, e.g., Lodge and Taber 2013).⁵ In research on the bandwagon effect, this tendency is called the “cognitive response mechanism” (Mutz 1998, 212): people change their mind in response to poll results because learning that a candidate is popular causes them to rationalize the information, by reflecting on other arguments supportive of that candidate.⁶ These other

arguments convince them to jump on the bandwagon. The behavior observed in the bandwagon experiment, then, would be the product of reflection on arguments supportive of candidates other than just their popularity, brought about by false information about their popularity. Debriefing cannot guarantee that these unknown additional arguments will cease to be relevant to the respondent. Indeed, evidence suggests that factual corrections alter people's factual beliefs, but not their political attitudes (Nyhan et al. 2020).

If treatments are "transformative" (Paul and Healy 2018), the inability to overturn treatment effects through debriefing becomes an ontological problem. Experimental treatments can effectively replace participants' "psychological selves" (Paul and Healy 2018, 326). Under this framework, the people who receive a debriefing message at T₂ are a fundamentally different population from the group who entered the experiment at T₁. Then, a debrief is not really a debrief, but rather a new message delivered to a group of people who are different from the group we misinformed. The difference is theoretically salient because the changes were brought about by a theory-derived intervention. The cognitive response problem is a (perhaps weaker) version of this challenge. The debrief is an attempt to persuade a group, more of whom will now be Trump supporters (according to the bandwagon effect), not to vote for Trump based on the information we previously provided to a group that was then composed of more Biden supporters—an importantly different group.

My point is not to reject the idea of debriefing outright, but rather that debriefing does not eliminate the ethical problems with misinformation. On a Kantian view, we should debrief, but we should still not misinform—there is no clear sense in which the former injunction overrides the latter. On a utilitarian view, debriefing cannot be expected to offset the process costs of misinformation and may even produce process costs of its own. Research subjects and scholars share the view that debriefing is not a solution to deception: absent informed consent, neither group judges experiments with debriefing more acceptable than those without (Desposato 2018, 745).

Inference under Misinformation

A simple response to these ethical challenges would be that misinformation, as a type of deception, is a necessary way of acquiring important knowledge (Penslar 1995). As Bok (1995, 2) puts it, researchers see a "trade-off between minor violations of the truth for the sake of access to far greater truths." With misinformation, the unique combination of a randomized treatment and a realistic context should enable researchers to get informative estimates of how people would really react if the treatment came true in the actual world—knowledge that can be applied to the benefit of society. In Humphreys' (2011) terms, though

misinformation incurs "participation" and "process" costs, it produces "outcome benefits"—the benefits of the knowledge produced by the findings and how it can be applied. So, in utilitarian terms, it may be worth it if these outcome benefits outweigh the costs discussed earlier.

There is, however, debate over whether the outcome benefits from experiments are valuable enough to justify their process and participation costs (e.g., Carlson 2020; Phillips 2021) and, indeed, whether such a calculation is even possible (Zimmerman 2016). Findley and Nielson (2016, 152) propose one approach, the "half-doubled rule" in which researchers "sincerely" estimate the benefits to human knowledge of a given experiment, and its costs to subjects, then reduce the estimated benefits by half and double the estimated costs. Such an adjustment may be necessary because researchers deceive themselves into believing their research is more valuable than it is (Desposato 2020). An extreme case of this logic, raised and rejected by Teele (2014, 116), is that ethics "are beside the point" when "experimental research is the only way to ensure valid causal inference."

The trouble with misinformation is that it tends to compromise the validity of such causal inference. So its putative outcome benefits are largely illusive. It therefore looks unlikely that the general calculus of weighing up outcome benefits against process and participation costs will favor employing misinformation in many cases. This argument can be seen at three levels of abstraction: the research design level, the causal inferential level, and the metaphysical level.

Research Design

It is a well-recognized tenet of good research design that effects should be externally valid—it should be possible to generalize them to the "population of interest" (Toshkov 2016, 173). The real contexts invoked by misinformation experiments would appear to make them more likely to find an externally valid effect; the realistic situation means that the sample should behave more like the population out in the actual world. But when seeking to maximize external validity, it not only matters that the *context* reflects the actual world; it also matters that *treatments* do. In the bandwagon example, there is a population of polls out in the world. The experimental design must maximize the representativeness of the sample of polls used in the experiment with respect to this population, if its results are to be generalizable to the relevant context. In my example, that context is the 2020 U.S. presidential election. Misinformation inherently falls short here, because it is not sampled from a population, but rather made up. If the design and analysis of the experiment implicitly treat Trump's lead in New York as no less likely than Biden being in the lead (which the polls would really have

suggested), then the effect of this scenario can be seriously miscalculated when mapping it onto the actual world (see de la Cuesta, Egami, and Imai 2021).

Realism also affects internal validity—whether “the independent variable produced the observed effect” (Halperin and Heath 2017, 149). Misinformation compromises internal validity by contravening the tenet of “experimental realism”:

to the extent that subjects become psychologically engaged in the process they confront, internal validity intensifies. Similarly, internal validity diminishes in the face of subject disengagement . . . If subjects approach a task with skepticism or detachment, then genuine responses fade . . . This raises the possibility that measures obtained do not accurately reflect the process being manipulated, but rather manifest a different underlying construct altogether. (McDermott 2011, 28)

In the face of misinformation, skeptical respondents might not look at the experimental task as something to take seriously. Rather than thinking about the election and how they would behave, instead the false treatment leads them to behave in arbitrary ways. They might also see the false treatment as deliberately pushing them in a certain direction, and then do what they think the researcher “wants” them to do (Jimenez-Buedo and Guala 2016; Zizzo 2010).

Such concerns can be traced back to the influential work of psychologist Sidney Siegel (1961; see also Siegel and Goldstein 1959), in whose view deception

undermined the relationship between the experimenter and the subjects, who, in consequence, are often too preoccupied with the true agenda of the experiment or doubt the announced relation between actions and rewards, which therefore ultimately jeopardizes experimental control. (Svorenčík 2016, 279)

Arguably, this concern is only valid if participants *know* they are being misinformed. If they are deceived, then they do not realize that the treatment is unrealistic. But, in this case, the ethical problems discussed earlier presumably must be taken more seriously. Essentially, if a treatment is deceptive, it must pass a threshold of realism. If it is not realistic, it must not be deceptive. Misinformation depends on being deceptive, with all the attendant ethical concerns, to maximize its scientific value.

Causal Inference

In causal inferential terms, these problems arise because misinformation is “ill-defined” or “ambiguous” (VanderWeele 2018, 1).⁷ Misinformational treatments are a “hypothetical intervention,” defined as “the specification, possibly contrary to fact, of the event or state $X = x$ ”—for instance, a hypothetical situation in the world where the polls (X) show Trump winning in New York (x) (VanderWeele 2018, 1). Such a hypothetical intervention is “well-defined” if “for each individual i in population P , there is a unique . . . distribution of values

. . . such that the event or state $X = x$, along with the state of the universe and the laws of nature, jointly entail” the observed outcome—otherwise “the hypothetical intervention is said to be ill-defined or ambiguous” (VanderWeele 2018, 1).

Misinformation is ambiguous by design. As VanderWeele (2018, 4) goes on to explain,

we could speak of a well-defined hypothetical intervention on a genetic variant even if we cannot at present alter it . . . it is relatively straightforward to think of just that one variant being other than it was and the rest of the universe being the same. In contrast, if we were to consider the temperature being 30 degrees versus 40 degrees in a specific town, there may have to be numerous things that would have to be different for the temperature to be different, such as the wind speed and air pressure in that town, along with the temperature in spatially contiguous towns, etc. . . and these different states of the universe may have very different implications for the outcome.

For Donald Trump to poll at 62% in New York, lots of other things would have to be different too. Importantly, many of these differences are direct implications of the theories that we (for the most part) subscribe to as political scientists. For example, the economic vote (e.g., Pickup and Evans 2013) would predict that, had Trump’s presidency brought about considerable economic growth and prosperity in the United States, he may have performed better at the 2020 election. If we lend strong credence to economic voting theory, we might argue that the economy would *have* to be in a better state for Trump’s counterfactual polling to be well-defined. But of course, the theory also implies that if the economy were doing better our research subjects would already be more likely to vote for Trump anyway. This is just one possible scenario, and there are many others that could give rise to the counterfactual we set up in our treatment. All those scenarios, if they truly came to pass, would produce a different distribution of responses to the experiment. The hypothetical intervention is ill-defined.

Such ambiguity can give rise to “information equivalence”:

manipulating information about a particular attribute will generally alter respondents’ beliefs about background attributes in the scenario as well . . . Manipulating whether a country is described as “a democracy” or “not a democracy,” for example, is likely to affect subjects’ beliefs about such background features as the country’s geographic location or demographic composition. If it does, then any differences between experimental groups cannot be reliably attributed to the effects of the beliefs of interest. (Dafoe, Zhang and Caughey 2018, 400–401)

Here then, the “things that would have to be different” themselves directly affect participants’ behavior in the experiment because the participants assume those things are different in order to make sense of the ambiguous intervention. In the bandwagon example, participants assume that the economy must be doing better, and this assumption affects their vote choice. As a result, the experiment does not uniquely isolate the effect of interest,

but rather any number of other effects of the background beliefs that respondents form to reduce the ambiguity of the misinformation (Landgrave and Weller 2022).

Metaphysics

In metaphysical terms, such problems arise because misinformation assumes the existence of impossible worlds. On a prominent view, counterfactuals are grounded in distinctions between the *actual* world and *possible* worlds (Lewis 1973): if something is false in our world, then counterfactual reasoning assumes there is a possible world where this thing is true—essentially, an alternate universe. Misinformation experiments attempt to create a possible world that is a “perfect duplicate” (Paul and Healy 2018, 321) of the actual world in all respects until the treatment is delivered. The truth of the treatment is the only thing that differs between the actual world and the possible world created by the experiment.

To see why we cannot do this, consider with David Lewis (1973, 9, emphasis original) the statement “if kangaroos had no tails, they would topple over”:

We might think it best to confine our attention to worlds where kangaroos have no tails and *everything* else is as it actually is; but there are no such worlds. Are we to suppose that kangaroos have no tails but that their tracks in the sand are as they actually are? Then we shall have to suppose that these tracks are produced in a way quite different from the actual way. Are we to suppose that kangaroos have no tails but that their genetic makeup is as it actually is? Then we shall have to suppose that genes control growth in a way quite different from the actual way (or else that there is something, unlike anything there actually is, that removes the tails). And so it goes; respects of similarity and difference trade off. If we try too hard for exact similarity to the actual world in one respect, we will get excessive differences in some other respect.

We cannot assume that there is a possible world whose only difference from our actual world is that the misinformation is true. There are no such worlds. There is no possible world where everything was the same right up to the point that 62% of voters in New York decided to support Donald Trump. There are possible worlds where Trump polls at 62% in New York; those worlds just differ from ours in other ways too.

Our respondents are not in any of those other worlds and have not undergone the processes that produced the differences in those worlds relative to the actual world. So misinformation cannot speak to the worlds where the treatment is true *and* there are other necessary differences from our own, because the sample on which it bases its findings, and the population to which it generalizes its effects, do not come from those worlds. They live in the actual world.

But what if, when cast into a possible world by misinformation, participants imagine they are in that world and then act accordingly? A fake poll result might make

them assume broader differences in the election context—in effect, projecting themselves into some possible world where the treatment makes sense. Of course, again, the trouble with this solution is precisely that people are likely to care about those differences and respond accordingly. Even if no unobserved factors affected responses in this way, people would have to be able to project themselves into infinitely many possible worlds to resolve the problem fully, because otherwise the question remains unanswered in many possible worlds—all those other worlds into which they do not “project themselves”.

Approximating the Truth

Arguably, these inferential concerns are less pressing the more misinformation approximates the truth. Consider the bandwagon case again. Polls have a margin of error. If we were interested in studying what would happen if Trump had the lead in state polls, we could choose a state where this situation is within the margin of error of current polling. We could then make up a poll within this margin and use this as our treatment. For example, we could ask:

A recent poll of vote intentions at the upcoming presidential election, conducted in your state of Arizona, put Donald Trump and Joe Biden on the following vote shares:

Trump 49%–45% Biden

Please indicate which candidate you intend to vote for.

These made-up vote shares would be within the margin of error of a Morning Consult poll conducted in late October 2020 suggesting Biden was in the lead on approximately 48% to Trump’s 46%.⁸ Because the margin of error is largely brought about by random noise, we could claim that nothing else would have to change to produce this alternative outcome—we would just have to get lucky. But as the treatment is still made-up, this is still a case of misinformation—just one we can analyze without worrying about inferential problems.

A first issue with this response is that such a quantitative idea of approximation is not always applicable. Countless potential applications of misinformation are not as malleable as the bandwagon example. As soon as we move beyond conveniently blurry cases, all the ethical and inferential issues raised above rear their heads. A second issue is that going to the effort of approximating the truth begs the question—why not just *tell* the truth?

True Treatments

That is, why not take data from the actual world and deploy it as a treatment in an experiment that is firmly grounded in the actual world—as Desposato (2016, 282) puts it, “tell the whole truth”? Call this the “true treatments” approach. Indeed, in the Arizona example,

the very same reason that the workaround makes sense—the fact that it could be produced randomly—also means that a poll could actually be conducted that comes up with the situation whose effects we want to test. A true treatments approach would involve taking such actual polling data and exposing respondents in the treatment group to it (e.g., van der Meer, Hakhverdian, and Aaldering 2016). This enables researchers to intervene in the political world without facing the charge of unduly influencing it by false pretenses.

For example, we could ask:⁹

A recent poll of vote intentions at the upcoming presidential election, conducted in your state of South Carolina, put Donald Trump and Joe Biden on the following vote shares:

Trump 51%—45% Biden

Please indicate which candidate you intend to vote for.

This approach might be criticized on the grounds that it does not do what experiments are supposed to do. Experiments allow us to answer counterfactual questions by drawing on possible political worlds; keeping the treatment within the range of values it takes within the actual world misses the potential to observe these counterfactuals. But this is the wrong way to think about what experiments do. The question is not strictly about what would happen if the political world were different. It is about what would happen if people's *exposure* to the political world were different—this is what we actually manipulate. We ask what people would do upon receiving some information about the political world. This is what the theories tested by informational interventions are ultimately about (Dafoe, Zhang and Caughey 2018, 400). The earlier arguments demonstrate that answering such questions raises considerable ethical and inferential challenges when the treatment information is not true. It works perfectly well with true treatments.

The Ethics of True Treatments

True treatments are not deceptive. As Humphreys (2014) claims, “interventions that seek to inject expert information into public debate around elections can sometimes be implemented using an entirely transparent design, with consent and no deception.” On Kant’s deontological view, this means that true treatments better respect the autonomy of the participant. On the utilitarian view, it means that they reduce the process and participation costs discussed earlier. Of course, they still produce consequences: if I report a real poll in a bandwagon experiment, the point is still to change people’s opinions or behavioral intentions. But there are crucial differences. First, because the information is true, any effect it has could have happened

anyway. Second, because the information is true, any effect it has may be normatively and politically desirable in some cases. Arendt (2005), for example, argues that truth-telling is essential to political life precisely because it informs people’s beliefs and promotes a legitimate plurality of opinions. Legislating between the cases where Arendt’s claim applies may require stronger assumptions about what is politically desirable that are beyond the scope of this paper—and on which political science lacks any strong consensus (Whitfield 2019, 530). Third though, a commitment to true treatments could shift debates away from the ethics of deception and towards—potentially more productive—discussions about how to resolve such “normative ambiguity” (Phillips 2021, 281). Perhaps there are political domains into which experimental political scientists should not intervene, but regardless, we are likely to be better off telling the truth when we do so.

Inference under True Treatments

Correspondingly, true treatments produce the outcome benefits that misinformation fails to produce, because they provide valid causal inferences. At the research design level, true treatments are representative of the information out in the actual world, so their effects are more externally valid than those of misinformation. Ideally, to maximize external validity, where different versions of the same information are available (as is often the case with polls) the versions should also be presented to respondents in proportion to their likelihood of exposure to those versions in their lives.¹⁰ As for internal validity, true treatments are realistic, so should minimize skepticism and other sources of disengagement.

At the causal inferential level, the hypothetical intervention is well-defined in the case of true treatments because the experiment makes the intervention non-hypothetical. We want to estimate the effect of information exposure as a counterfactual state, and the experiment directly intervenes to bring about this state—i.e., to make it no longer counterfactual for some people. As this exposure is determined purely randomly, it does not rely on anything else changing to produce it. That is, in VanderWeele’s (2018, 4) terms, the “state of the universe” is held constant. There is no ambiguity in the intervention. We directly isolate and manipulate the intervention of interest and observe the effect. There is, then, no intrinsic reason for participants to “update their beliefs about background attributes,” reducing concerns about “information equivalence” (Dafoe, Zhang and Caughey 2018, 400–401).

At the metaphysical level, true treatments do not require the existence of worlds that could not possibly exist. Instead, the treatment *is* true in the actual world. The possible world we are interested in is one where people are

exposed to this information. We know that this possible world exists because we create it by exposing them to the information. And randomization into treatment and control groups guarantees that there is a possible world in which each respondent was exposed to the information. In these possible worlds, *everything else* was the same up to the point that they were exposed to the information, because the process by which this exposure is determined is random. The counterfactual makes sense. Causal inference is possible.

Conclusion

In response to APSA's call to "carefully consider" any use of misinformation in experimental design, this paper has made three contributions. First, it has clarified how we can think about the ethics of misinformation, as a specific type of deception. Second, it has argued that misinformation tends to compromise inference. Third, it has invited researchers to adopt "true treatments" as a solution to these problems.

By providing these reflections, I raise what Paul and Healy (2018, 334) refer to as a "methodological challenge" to experimental political science. While such challenges often require that scholars look for "innovations" (Paul and Healy 2018, 334) to continue doing experimental research, I have argued that no innovation is necessary. If political scientists still see misinformation as the only way to tackle their research questions, then in addition to APSA's (ACHSR 2020, 7) suggestion to disclose, justify, and explain steps taken to mitigate the ethical cost of deception, researchers should also address whether and how their design overcomes the inferential problems set out here. But a better solution is right under our noses. By simply *telling the truth*, political scientists not only do more ethical research, but also do causal inference that makes more sense.

Acknowledgments

The author would like to thank Peter Allen, Philip Cowley, Javier Sajuria, Mirko Palestrino, Jack Bailey, James Barnfield, Alejandro Fernández-Roldán Díaz, María Jiménez-Buedo, and Salomé Letter for insightful comments on this paper. He also thanks Michael Bernhard and the two anonymous reviewers for their helpful advice and suggestions.

Notes

¹ APSA (ACHSR 2020, 8) notes that deception can be an "act of omission" in that researchers can also provide "formulations intended to mislead participants (whether or not the information is literally true)," somewhat blurring a distinction made by psychologists between "lying by omission" (the "passive omission of relevant information") and "paltering" (the "use of

truthful statements to convey a misleading impression") (Rogers et al. 2017, 456). Experimental treatments can, in principle, be deceptive in all these ways. For expositional clarity, I focus on cases of "lying by commission" (the "active use of false statements") (Rogers et al. 2017, 456), as this form of deception fits best with APSA's original definition of misinformation. Resolving debates about whether lies of omission and paltering pose the same ethical and inferential problems as lies of commission is beyond the scope of this paper. Future work may fruitfully address these nuances—indeed, these are exactly the kinds of debate I hope to spark.

² Zimmerman (2016, 184–185) provides an extensive list of such experiments, on a wide range of topics.

³ Science fiction writer Ursula K Le Guin (2017, xiv) highlighted this analogy when explaining that "you can read ... a lot of ... science fiction, as a thought experiment Let's say this or that is such and so, and see what happens In a story so conceived, the moral complexity proper of the modern novel need not be sacrificed Thought and intuition can move freely within bounds set only by the terms of the experiment." She concludes, elsewhere, that "fiction is invention, but it is not lies"—however "if you begin to fake [the facts], to pretend things happened in a way that makes a nice neat story, you're misusing imagination. You're passing invention off as fact, which is, among children at least, called lying" (2019, 108).

⁴ Concerns about subject disengagement and overextraction are particularly pertinent in the case of "elite field experiments" or "audit experiments," where potential participants are savvier to the deceptive practices used by researchers (Landgrave 2020, 490)—not to mention, there are fewer of them. Misinformation is a common feature of such experiments (Bischof et al. 2021) and may produce unique process costs that extend beyond the effects on the individual taking the survey (Desposato 2022).

⁵ Some theories emphasize "hot" or otherwise "affectionately charged" cognition (Lodge and Taber 2013). In Affective Intelligence Theory (Marcus, Neuman, and MacKuen 2000), attitude-incongruent information induces "anxiety," motivating a "search" for information. Debriefing may reassure people that they do not need to have been anxious, but it cannot undo the experience. Kelman (1967) questions at length the researcher's right to affect the psychology of participants in such ways.

⁶ This mechanism may underlie the problem, discussed later, of "information equivalence" (Dafoe, Zhang and Caughey, 2018).

⁷ Really, external and internal validity are aspects of causal inference. I keep them separate recognizing that even those who are not willing to commit to more

explicitly causal language, or the dominant frameworks of causal inference, are nonetheless likely to endorse the importance of validity.

⁸ See <https://morningconsult.com/form/2020-u-s-election-tracker/>.

⁹ This is the result of a Morning Consult poll conducted in late October 2020: <https://morningconsult.com/form/2020-u-s-election-tracker/>.

¹⁰ This is also a clear way to limit the presence of “lies of omission” and “paltering” (Rogers et al. 2017, 456), which true treatments do not fully rule out *per se*.

References

- Ad Hoc Committee on Human Subjects Research ACHSR). 2020. “Principles and Guidance for Human Subjects Research.” *Principles and Guidance*, April. (; https://www.apsanet.org/Portals/54/diversity%20and%20inclusion%20prgms/Ethics/Final_Principles%20with%20Guidance%20with%20intro.pdf?ver=2020-04-20-211740-153).
- Arendt, Hannah. 2005 [1967]. “Truth and Politics.” In *Truth: Engagements across Philosophical Traditions*, ed. José Medina and David Wood, 295–314. Oxford: Blackwell Publishing.
- Barnfield, Matthew. 2020. “Think Twice before Jumping on the Bandwagon: Clarifying Concepts in Research on the Bandwagon Effect.” *Political Studies Review* 18(4): 553–74.
- Beauchamp, Tom L., and James F. Childress. 2019. *Principles of Biomedical Ethics*. 8th ed. New York: Oxford University Press.
- Berghmans, Ron L.P. 2007. “Misleading Research Participants: Moral Aspects of Deception in Psychological Research.” *Netherlands Journal of Psychology* 63(1): 12–17.
- Bischof, Daniel, Gidon Cohen, Sarah Cohen, Florian Foos, Patrick Michael Kuhn, Kyriaki Nanou, Neil Visalvanich, and Nick Vivyan. 2021. “Advantages, Challenges and Limitations of Audit Experiments with Constituents.” *Political Studies Review*, OnlineFirst. <https://doi.org/10.1177/14789299211037865>
- Bok, Sissela. 1995. “Shading the Truth in Seeking Informed Consent for Research Purposes.” *Kennedy Institute of Ethics Journal* 5(1): 1–17.
- Brown, Étienne. 2018. “Propaganda, Misinformation, and the Epistemic Value of Democracy.” *Critical Review*, 30(3–4): 194–218.
- Callard, Agnes. 2018. *Aspiration: The Agency of Becoming*. Oxford: Oxford University Press.
- Carlson, Elizabeth. 2020. “Field Experiments and Behavioral Theories: Science and Ethics.” *PS: Political Science & Politics* 53(1): 89–93.
- Carpenter, Charli, Alexander H. Montgomery, and Alexandria Nylen. 2021. “Breaking Bad? How Survey Experiments Prime Americans for War Crimes.” *Perspectives on Politics* 19(3): 912–24.
- Dafoe, Allan, Baobao Zhang, and Devin Caughey. 2018. “Information Equivalence in Survey Experiments.” *Political Analysis* 26(4): 399–416.
- Dahlgaard, Jens Olav, Jonas Hedegaard Hansen, Kasper M. Hansen, and Martin V. Larsen. 2017. “How Election Polls Shape Voting Behaviour.” *Scandinavian Political Studies* 40(3): 330–43.
- de la Cuesta, Brandon, Naoki Egami, and Kosuke Imai. 2021. “Improving the External Validity of Conjoint Analysis: The Essential Role of Profile Distribution.” *Political Analysis* 30(1): 19–45.
- De Ridder, Jeroen. 2021. “What’s So Bad about Misinformation?” *Inquiry*. <https://doi.org/10.1080/0020174X.2021.2002187>
- Desposato, Scott. 2016. “Conclusion and Recommendations.” In *Ethics and Experiments: Problems and Solutions for Social Scientists and Policy Professionals*, ed. Scott Desposato, 267–289. New York: Routledge.
- . 2018. “Subjects and Scholars’ Views on the Ethics of Political Science Field Experiments.” *Perspectives on Politics*. 16(3): 739–50.
- . 2020. “The Ethical Challenges of Political Science Field Experiments.” In *The Oxford Handbook of Research Ethics*, ed. Ana S. Iltis. and Douglas P. MacKay, 1–34. Oxford: Oxford University Press.
- . 2022. “Public Impacts from Elite Audit Experiments: Aggregate and Response Delay Harms.” *Political Studies Review*, OnlineFirst. <https://doi.org/10.1177/14789299211059657>
- Dizney, Henry F., and Ronald W. Roskens. 1962. “An Investigation of the “Bandwagon Effect” in a College Straw Election.” *Journal of Educational Sociology* 36(3): 108–14.
- Druckman, James N., and Donald P. Green. 2021. “A New Era of Experimental Political Science.” In *Advances in Experimental Political Science*, ed. James N. Druckman and Donald P. Green, 1–18. Cambridge: Cambridge University Press.
- Findley, Michael, and Daniel Nielson. 2016. “Obligated to Deceive? Aliases, Confederates, and the Common Rule in International Field Experiments.” In *Ethics and Experiments: Problems and Solutions for Social Scientists and Policy Professionals*, ed. Scott Desposato, 151–70. New York: Routledge.
- Frankfurt, H. 1998. *The Importance of What We Care About*. Cambridge: Cambridge University Press.
- Goidel, Robert K. and Todd G. Shields. 1994. “The Vanishing Marginals, the Bandwagon, and the Mass Media.” *Journal of Politics* 56(3): 802–10.

-
- Halperin, Sandra, and Oliver Heath. 2017. *Political Research: Methods and Practical Skills*. 2nd ed. Oxford: Oxford University Press.
- Hertwig, Ralph, and Andreas Ortmann. 2008. "Deception in Experiments: Revisiting the Arguments in Its Defense." *Ethics & Behavior* 18(1): 59–92.
- Humphreys, Macartan. 2011. "Ethical Challenges of Embedded Experimentation." (<http://www.columbia.edu/~mh2245/papers1/20110912Ethics.pdf>)
- . 2014. "How to Make Field Experiments More Ethical." *Monkey Cage—Washington Post*, November 2. (<https://www.washingtonpost.com/news/monkey-cage/wp/2014/11/02/how-to-make-field-experiments-more-ethical/>).
- . 2015. "Reflections on the Ethics of Social Experimentation." *Journal of Globalization and Development* 6(1): 87–112.
- Imbens, Guido W., and Donald B. Rubin. 2015. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. New York: Cambridge University Press.
- Jimenez-Buedo, Maria, and Francesco Guala. 2016. "Artificiality, Reactivity, and Demand Effects in Experimental Economics." *Philosophy of the Social Sciences* 46(1): 3–23.
- Kant, Immanuel. 1978 [1798]. *Anthropology from a Pragmatic Point of View*. Trans. Victor Lyle Dowdell. London: Fffer and Simons, Inc.
- . 1996 [1797]. *The Metaphysics of Morals*. Trans. Mary Gregor. Cambridge: Cambridge University Press.
- Kelman, Herbert C. 1967. "Human Use of Human Subjects: The Problem of Deception in Social Psychological Experiments." *Psychological Bulletin* 67 (1): 1–11.
- Korsgaard, Christine. 1996. *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press.
- . 2018. *Fellow Creatures: Our Obligations to the Other Animals*. Oxford: Oxford University Press.
- Krupnikov, Yanna, H. Hannah Nam, and Hillary Style. 2021. "Convenience Samples in Political Science Experiments." In *Advances in Experimental Political Science*, ed. James N. Druckman and Donald P. Green, 165–183. Cambridge: Cambridge University Press.
- Landgrave, Michelangelo. 2020. "Can We Reduce Deception in Elite Field Experiments? Evidence from a Field Experiment with State Legislative Offices." *State Politics & Policy Quarterly* 20(4): 489–507.
- Landgrave, Michelangelo, and Nicholas Weller. 2022. "Do Name-Based Treatments Violate Information Equivalence? Evidence from a Correspondence Audit Experiment." *Political Analysis* 30(1): 142–48.
- Le Guin, Ursula K. 2017. "Introduction." In *The Left Hand of Darkness*, xiii–xvii. London: Gollancz.
- . 2019. *Words Are My Matter: Writings on Life and Books*. New York: Mariner Books.
- Leeper, Thomas J. 2019. "Where Have the Respondents Gone? Perhaps We Ate Them All." *Public Opinion Quarterly* 83(S1): 280–88.
- Lewis, David. 1973. *Counterfactuals*. Oxford: Blackwell Publishing.
- Lodge, Milton, and Charles S. Taber. 2013. *The Rationalizing Voter*. Cambridge: Cambridge University Press.
- Madson, Gabriel J., and D. Sunshine Hillygus. 2019. "All the Best Polls Agree with Me: Bias in Evaluations of Political Polling." *Political Behavior* 42:1055–72.
- Marcus, George E., W. Russell Neuman, and Michael MacKuen. 2000. *Affective Intelligence and Political Judgement*. Chicago: University of Chicago Press.
- McDermott, Rose. 2011. "Internal and External Validity." In *Cambridge Handbook of Experimental Political Science*, ed. James N. Druckman, Donald P. Green, James H. Kuklinski, and Arthur Lupia, 27–40. New York: Cambridge University Press.
- Mutz, Diana C. 1998. *Impersonal Influence: How Perceptions of Mass Collectives Affect Political Attitudes*. Cambridge: Cambridge University Press.
- Nyhan, Brendan, Ethan Porter, Jason Reifler, and Thomas J. Wood. 2020. "Taking Fact-Checks Literally but Not Seriously? The Effects of Journalistic Fact-Checking on Factual Beliefs and Candidate Favorability." *Political Behavior* 42(3): 939–60.
- Paul, L. A., and Kieran Healy. 2018. "Transformative Treatments." *Noûs* 52(2): 320–35.
- Penslar, Robin L. 1995. *Research Ethics: Cases and Materials*. Indianapolis: Indiana University Press.
- Phillips, Trisha. 2021. "Ethics of Field Experiments." *Annual Review of Political Science* 24(1): 277–300.
- Pickup, Mark, and Geoffrey Evans. 2013. "Addressing the Endogeneity of Economic Evaluations in Models of Political Choice." *Public Opinion Quarterly* 77(3): 735–54.
- Rogers, Todd, Richard Zeckhauser, Francesca Gino, Michael I Norton, and Maurice E Schweitzer. 2017. "Artful Paltering: The Risks and Rewards of Using Truthful Statements to Mislead Others." *Journal of Personality and Social Psychology* 112(3): 456–73.
- Siegel, Sidney. 1961. "Decision Making and Learning under Varying Conditions of Reinforcement." *Annals of the New York Academy of Sciences* 89(5): 766–83.
- Siegel, Sidney, and Goldstein, Donald A. 1959. "Decision-Making Behavior in a Two-Choice Uncertain Outcome Situation." *Journal of Experimental Psychology* 57(1): 37–42.
- Svorenčík, Andrej. 2016. "The Sidney Siegel Tradition: The Divergence of Behavioral and Experimental Economics at the End of the 1980s." *History of Political Economy* 48:270–94.

- Teele, Dawn L. 2014. "Reflections on the Ethics of Field Experiments." In *Field Experiments and Their Critics: Essays on the Uses and Abuses of Experimentation in the Social Sciences*, ed. Dawn L. Teele, 115–140. New Haven, CT: Yale University Press.
- Toshkov, Dimiter. 2016. *Research Design in Political Science*. London: Bloomsbury Publishing.
- van der Meer, Tom W.G., Armen Hakverdian, and Loes Aaldering. 2016. "Off the Fence, Onto the Bandwagon? A Large-Scale Survey Experiment on Effect of Real-Life Poll Outcomes on Subsequent Vote Intentions." *International Journal of Public Opinion Research* 28(1): 46–72.
- VanderWeele, Tyler J. 2018. "On Well-Defined Hypothetical Interventions in the Potential Outcomes Framework." *Epidemiology* 29(4): e24–25.
- Whitfield, Gregory. 2019. "TRENDS: Toward a Separate Ethics of Political Field Experiments." *Political Research Quarterly* 72(3): 527–38.
- Wilson, James, and David Hunter. 2010. "Research Exceptionalism." *American Journal of Bioethics* 10(8): 45–54.
- Zimmerman, B. 2016. "Information and Power: Ethical Considerations of Political Information Experiments." In *Ethics and Experiments: Problems and Solutions for Social Scientists and Policy Professionals*, ed. Scott Desposato, 183–197. New York: Routledge.
- Zizzo, Daniel John. 2010. "Experimenter Demand Effects in Economic Experiments." *Experimental Economics* 13(1): 75–98.