

# Information and its Presentation: Treatment Effects in Low-Information vs. High-Information Experiments

David J. Andersen and Tessa Ditonto

Iowa State University, Political Science, 547 Ross Hall, Ames, Iowa 50010, USA. Email: [dander@iastate.edu](mailto:dander@iastate.edu), [tditonto@iastate.edu](mailto:tditonto@iastate.edu)

## Abstract

This article examines how the presentation of information during a laboratory experiment can alter a study's findings. We compare four possible ways to present information about hypothetical candidates in a laboratory experiment. First, we manipulate whether subjects experience a low-information or a high-information campaign. Second, we manipulate whether the information is presented statically or dynamically. We find that the design of a study can produce very different conclusions. Using candidate's gender as our manipulation, we find significant effects on a variety of candidate evaluation measures in low-information conditions, but almost no significant effects in high-information conditions. We also find that subjects in high-information settings tend to seek out more information in dynamic environments than static, though their ultimate candidate evaluations do not differ. Implications and recommendations for future avenues of study are discussed.

*Keywords:* experimental design, laboratory experiment, treatment effects, candidate evaluation, survey experiment, dynamic process-tracing environment, gender cues

Over the past 50 years, one of the major areas of growth within political science has been in political psychology. The increasing use of psychological theories to explain political behavior has revolutionized the discipline, altering how we think about political activity and how we conduct political science research. Along with the advent of new psychological theories, we have also seen the rise of new research methods, particularly experiments that allow us to test those theories (for summaries of the growth of experimental methods, see McDermott 2002; and Druckman *et al.* 2006). Like all methods, experimental research has strengths and weaknesses. Most notably, experiments excel in attributing causality, but typically suffer from questionable external validity. Further, two different types of experiments exist, each of which deals with this tradeoff differently: laboratory studies that maximize control and causal inferences at the expense of external validity, and field studies that increase external validity by weakening control over the research setting (Morton and Williams 2010; Gerber and Green 2012).

In this article, we identify a middle ground and assess whether presenting an experimental treatment in a more realistic, high-information laboratory environment produces different results than those that come from more commonly used, low-information laboratory procedures, and then examine why those differences occur. In particular, we examine whether manipulations of candidate gender have different effects on candidate evaluation when they are embedded within an informationally complex "campaign" than when they are presented in the more traditional low-information survey or "vignette"-style experiment. To do this, we use the Dynamic Process Tracing Environment (DPTE), an online platform that allows researchers to simulate the rich and constantly changing information environment of real-world campaigns.

*Political Analysis* (2018)  
vol. 26:379–398  
DOI: 10.1017/pan.2018.21

Published  
3 August 2018

Corresponding author  
David J. Andersen

Edited by  
R. Michael Alvarez

© The Author(s) 2018. Published  
by Cambridge University Press  
on behalf of the Society for  
Political Methodology.

*Authors' note:* The data, code and any additional materials required to replicate all analyses in this article are available at the *Political Analysis* Dataverse within the Harvard Dataverse Network, at doi:10.7910/DVN/TGFAOH (Andersen 2018).

While this is not the first study to use or discuss DPTE (see Lau and Redlawsk 1997, 2006 for originating work), this is the first attempt to determine whether DPTE studies produce substantively different results from traditional survey experiments, which present subjects only with short vignettes to consider.<sup>1</sup> We use DPTE to examine whether variations in the presentation of information in an experiment create differences in subjects' evaluations of two candidates. We argue here that high-information studies help to correct for exaggerated treatment effects that are often attributed to vignette-style experiments, while still allowing scholars to randomly assign subjects to different conditions and expose them to desired treatments. To do so, we focus upon three simple manipulations: the manner in which information about the candidates is presented (statically or dynamically), the amount of information presented about the candidates (low- vs. high-information) and the gender of the subject's in-party candidate.

## 1 Laboratory Experiments in Political Science

Laboratory experiments have emerged as a leading technique to study topics that are difficult to manipulate in the real world, such as the effects that candidate characteristics like gender have upon voter evaluations of those candidates. Vignette-style experiments are relatively easy to design, low cost and easy to field, and permit clear, strong causal inferences. Use of this design has proliferated in the past several decades, adding a great deal to what we know about political psychology (early paradigm setting examples studying candidate gender include Sigelman and Sigelman 1982; Huddy and Terkildsen 1993a,b). The recent emergence of research centers that provide nationally representative samples online (such as YouGov, Knowledge Networks, and Survey Sampling International), the creation of large national surveys that researchers can join (such as Time-sharing Experiments for the Social Sciences (TESS) and the Cooperative Congressional Election Study (CCES)), as well as the opening of online labor pools like Amazon's Mechanical Turk, have meant that survey experiments can now be delivered inexpensively to huge, representative samples that grant the ability to generalize results onto the broader population (Gilens 2001; Brooks and Geer 2007; Mutz 2011; Berinsky, Huber, and Lenz 2012).

As they have recently grown in popularity, inevitable methodological counterarguments have also developed (see particularly Gaines, Kuklinski, and Quirk 2007; Kinder 2007; Barabas and Jerit 2010). For all their benefits, experiments—even those that are conducted on a population-based random sample—provide questionable external validity. This has been particularly noted for the vignette-style survey experiments that have become dominant in the discipline. Observed treatment effects in such studies seem to be higher than those observed in the real world via either field or natural experiments (Barabas and Jerit 2010; Jerit, Barabas, and Clifford 2013). This is partially unavoidable. All research that studies a proxy dependent variable (i.e. a vote for hypothetical candidates in a hypothetical election) necessarily lacks the ability to declare a clear connection with the actual dependent variable of interest (i.e. real votes in real-world elections).

Further, all experiments force exposure to a treatment while simultaneously limiting subjects' access to other information. In doing so, they create a tightly controlled information environment in which causal inferences can be easily made. However, this also makes most experimental scenarios decidedly unrealistic (McDermott 2002; Iyengar 2011). For many voters, the bare, minimalistic descriptions available in short vignettes may give little reason at all to vote for, or against, the candidates. Vote decisions, particularly for high-level state and federal offices, are typically much more involved than these minimal information environments allow

1 Please note that by *survey experiments*, we are referring to any experiment that uses survey methods to collect information from subjects before and/or after a treatment where that treatment is a static presentation of a small set of information (Mutz 2011). This includes many experiments conducted in laboratory settings, online, and embedded within nationally representative surveys. This classification depends upon a study's procedure, rather than the nature of the sample. We also use the term *laboratory experiments*, which is any experiment in which the entire information environment is controlled by the researcher.

(Carsey and Wright 1998; Highton 2004; Mcdermott and Jones 2005). Researchers may well find a causal relationship between two variables in a study like this, but what becomes of that relationship in an actual campaign, where candidates present issue stances, make impassioned speeches and launch numerous targeted ads aimed at influencing voters? It is possible, perhaps even likely, that additional information may alter or completely negate that relationship. By restricting the availability of other information, vignette-style experiments create an environment in which the limited information subjects can access may produce outsized effects, simply because it is the only information available.

In addition, by virtue of their design, these experiments immediately measure the response to the treatment, preventing any diminishing of the treatment effect over time (Jerit, Barabas, and Clifford 2013). Treatment effects are not always long lasting, and the influence that any individual piece of information has may decline as time goes on (Lodge, Stroh, and Wahlke 1990; Lodge, Steenbergen, and Brau 1995). A design more concerned with external validity might give subjects more time between accessing a treatment and being asked to evaluate a candidate in order to allow information to be processed for relevance or importance, as happens during a real political campaign. Votes, after all, are still mainly cast on Election Day, permitting voters days, weeks or even months of time to digest campaign information. In the low-information, immediate-reaction scenarios that short, vignette-style survey experiments create, however, treatments are given the “best-case scenario” to produce significant effects.

This is not to say that such experiments are without value—quite the contrary. Low-information vignette experiments seem to exaggerate treatment effects, but they generally do not find results that are out of line with what occurs in more externally valid field experiments or natural experiments (Barabas and Jerit 2010; Jerit, Barabas, and Clifford 2013). They have repeatedly been shown to be very effective at demonstrating that certain treatments *can* have an effect and that a particular independent variable *can* influence a dependent variable. A harder question is determining if treatments *tend* to have effects in the real world, when people have other information to consult, and have time to allow the treatment to dissipate.

For many topics, however, field experiments and natural experiments are not viable possibilities, leaving few alternatives to further test the external validity of observed treatment effects. Many research questions require scenarios that the real world does not frequently present (e.g., races with candidates of various races and genders) or that are difficult to manipulate in the real world (i.e. the conduct of a campaign or the presentation of a candidate). This leaves some form of laboratory or survey experiment as the best option for many research topics.

## 2 Process-Tracing Experiments and Information Processing Theories

While vignette-style experiments are the most commonly used form of laboratory experiments, other options do exist. Process-tracing experiments ask subjects to make a decision between various alternatives by learning about them in a manner that can be observed and followed by the researcher. Rather than restricting subjects to a very limited set of information, process-tracing studies present a much larger universe of information and monitor how subjects opt to learn about the alternatives they are asked to choose between. The goal, rather than providing a small set of information that all subjects view in its entirety, is to provide a larger set of information and allow each subject to choose which information to access. While this may lead subjects to view different information from each other, it better replicates how people make decisions in the real world, by choosing what information they wish to encounter.

The first process-tracing experiments asked people to use a static information board to learn relevant information about each possible alternative, typically by flipping over notecards tacked to a board (Payne 1976; Ericsson and Simon 1980), while the researcher observed the subjects’ behavior (Jacoby, Speller, and Kohn 1974; Carroll and Johnson 1990). In order to better mimic

the dynamic nature of a political campaign, Richard R. Lau and David Redlawsk developed the Dynamic Process Tracing Environment (DPTE), which recreates the basic premise of static information boards in a dynamic, computer-based platform. DPTE places all bits of available information into a single, randomized column of “information boxes” that scroll down a computer screen, giving subjects the ability to choose the information they would like to click on and learn more about. The dynamic nature of this design better resembles a real-world campaign, where a great deal of information exists and its presentation and availability are largely out of our control, but where we ultimately choose much of what we see. (For a discussion of how dynamic environments more closely mimic campaigns, see Lau 1995, Lau and Redlawsk 1997 and Lau and Redlawsk 2006).

Dynamic process-tracing techniques have been used to analyze voter decision-making (e.g. Redlawsk 2004; Ditonto, Hamilton, and Redlawsk 2014; Ditonto 2017), and have been demonstrated to produce replicable results using American National Election Study data (see Lau and Redlawsk 1997; Lau, Andersen, and Redlawsk 2008). They have not yet, however, been compared to similar vignette experiments. We posit that dynamic process-tracing studies may serve as a middle ground between shorter vignette-style experiments and the real world—allowing researchers the ability to examine causal relationships while also providing a sufficiently realistic information environment to produce more externally valid results. More specifically, we believe that the design of high-information dynamic process-tracing studies attenuate treatment effects in ways that more closely mimic the real world.

We ground our beliefs in information processing theories, which suggest that the manner in which people encounter and process information matters to how they use that information in making evaluations and decisions (Simon 1979; Anderson 1983; Hastie 1986; Lau and Redlawsk 2006). In limiting the availability and presentation of information, short vignette-style experiments may exaggerate the role played by the treatments presented. This alters the research question being addressed from “*does* this information have an effect?” to “*can* this information have an effect?” or perhaps more specifically “can this information have an effect in *isolation*?”

Information processing theory suggests that bytes of information do not have constant, persistent effects, but are used to update beliefs relative to what other considerations a person has in short-term and working memory (Lodge and Hamill 1986; McGraw, Lodge, and Stroh 1990; Zaller 1992; Zaller and Feldman 1992; Lau and Redlawsk 2006). An information item may be influential on an opinion or not, depending on what other information is immediately available. Over time, the effects of new information also tend to dissipate, and may disappear altogether (Lodge, Stroh, and Wahlke 1990; Lodge, Steenbergen, and Brau 1995). Thus, there may be great differences on the effects of learning a new item of information depending on whether alternative information is readily available, or whether the measurement of opinion change occurs immediately or some time later.

All laboratory experiments constrain the universe of information a subject has available. While this strengthens causal inferences and makes for a more parsimonious design, it also makes whatever information subjects are presented with more likely to be influential. Each individual piece of information represents a larger share of the total information available when that universe is smaller. In the real world however, all information is encountered among a milieu of other considerations and balanced for relevance and importance. A more effective way to assess if a treatment actually has an effect in the real world might be to simply present that treatment alongside a larger set of other information in the laboratory, in a manner similar to how such decisions are typically made, but do so in a manner that allows the researcher to track how much and which information is being accessed.

### 3 Test Case—Gender Stereotypes and Candidate Evaluations

In order to test our theory, we examine the role of a candidate's gender in influencing his or her evaluations and electoral fortunes. This is a topic that has received much attention from political psychologists over the past 20 years, and about which there is still much contention. A great deal of experimental evidence suggests that a candidate's gender can affect the way voters judge him or her, and that women candidates are often subject to a number of stereotypes. For example, women candidates are often assumed to have more feminine and communal characteristics—they are seen as more compassionate, gentle, warm, cautious and emotional, for example (Leeper 1991; Huddy and Terkildsen 1993a,b; Kahn 1996). Also, they are often seen as more trustworthy and honest than male candidates (Kahn 1996). At the same time, they are stereotyped as less agentic—less competent, less able to handle the emotional demands of high office, and lacking in masculine traits like “toughness” (Huddy and Terkildsen 1993a,b; Carroll and Dittmar 2010). Stemming from these assumptions about women's personality traits, voters often assume that women have different areas of policy expertise than men, with particular proficiency in “compassion issues” like education, healthcare, poverty, and child care often attributed to women candidates. At the same time, more “masculine” issues like crime, the military, and the economy are seen as the arena of male politicians (Leeper 1991; Alexander and Andersen 1993; Cook, Thomas, and Wilcox 1994; Dolan 2004). Finally, women candidates are stereotyped as more liberal than male candidates (McDermott 1997, 1998; Koch 2000, Koch 2002).

Despite the plethora of experimental evidence that female candidates are subject to gender-based stereotypes, other scholars have found that, in real-world scenarios, “when women run, women win.” In other words, women are not generally disadvantaged in real elections and often win their races as often as men do (Burrell 1997, Darcy, Welch, and Clark 1997, Seltzer, Newman, and Leighton 1997, Woods 2000, Dolan 2004). Further, several studies have found that expressly political factors, such as partisanship matter much more than candidate gender in real-world elections (Sanbonmatsu and Dolan 2009; Hayes 2011; Dolan 2014).

What accounts for this disconnect between findings that stereotypes exist and those that find that gender does not seem to influence electoral outcomes? It has been suggested that part of this discrepancy may be methodological in nature (e.g. Brooks 2013, Dolan 2014). The bulk of the evidence suggesting that female candidates are evaluated differently from men comes from experimental studies, and vignette experiments in particular. At the same time, many of the findings that seem to demonstrate that candidate gender does not matter are the results of nationally representative survey research (though see Brooks 2013 for a prominent example of experimental evidence that candidate gender is not relevant). Dolan (2014), for example, uses survey data to show that voters generally do not use stereotypes to evaluate female candidates, and even if they do, political party matters much more than gender in determining vote decisions.

Most relevant for our purposes, several studies have found that gender matters specifically in low-information elections (McDermott 1997, 1998; Sapiro 1981; Higgle *et al.* 1997; Matson and Fine 2006; Banducci *et al.* 2008). This is not surprising since psychologists have found that the existence of individuating information (that is, substantive information about a particular individual) has the ability to minimize the use of stereotypes in person evaluations (Fiske and Neuberg 1990). Voters in low-information elections have little individuating information to go on, so gender becomes an important cue. However, it is possible that candidate gender would matter less, or not at all, if other individuating information was available, like a real-world campaign.

To our knowledge, though, no one has yet explicitly compared the effects of candidate gender in low- vs. high-information scenarios. It is our contention that most vignette-style experiments are essentially simulating low-information elections, whether they intend to or not, and that the presentation of a gender manipulation with minimal individuating information will lead to very different evaluations than the presentation of that same manipulation along with other kinds of

	Low Information	High Information
Static Presentation	News Articles	Static Information Board
Dynamic Presentation	Low-Information Dynamic Board	High-Information Dynamic Board

**Figure 1.** Information Groups.

information that are generally available during most high-level political campaigns (i.e. Federal and most statewide offices). If we find that gender matters in low-information conditions but not in high-information environments, that may be evidence that the lack of clarity about the role of gender in elections has to do with the methods being used by researchers and that the information environment in a particular experiment matters a great deal. If gender influences candidate evaluations across the board, though (or not at all) that may be evidence that other factors are at play, such as the changing nature of gender roles and expectations within society.

## 4 Data and Method

To test whether different styles of experiments create significantly different experiences for subjects, leading to substantively different results, we fielded a  $2 \times 2 \times 2$  experiment<sup>2</sup> in the summer of 2015 to approximately 800 subjects<sup>3</sup> recruited through Amazon’s Mechanical Turk. We used the Dynamic Process Tracing Environment (DPTE) to create four different methods of delivering information to our subjects. Each subject proceeded through four “stages” in the experiment. They first answered some basic demographic and political questions, then participated in a “practice round” to learn how the program worked, then met the candidates in a “campaign” and finally cast a vote and evaluated the candidates.

### 4.1 Information presentation manipulations

Subjects were randomly sorted into four conditions that altered how they learned about the two candidates, classified across two axes of information presentation. First, each subject was randomly assigned to either a “low” information or “high” information condition. In the low-information condition, subjects could only learn five facts about each candidate—their education, family, prior political experience, religion, and an evaluation of them by the state newspaper’s editorial page. The low-information conditions were designed to be similar to previous vignette experiments and so present the types of background information often found in such studies (in particular, we use the information included in Huddy and Terkildsen’s highly influential 1993 articles). In the high-information condition, subjects could learn the five factors presented in the minimal conditions along with 15 additional attributes about each candidate, making them reasonably well-defined.<sup>4</sup>

Subjects were also randomly sorted to learn about candidates either statically or dynamically. In the static conditions, information was presented in a manner in which subjects were easily able to access all of the information that would be available to them. They had complete access to available information without limitation. In the dynamic condition, the information was presented randomly in a dynamic information board, presenting them with six available information boxes at a time. The boxes slowly scrolled down the screen, and for each box that scrolled off the bottom of the screen, a new information item replaced it at the top until each item had appeared twice. This created a  $2 \times 2$  set of conditions as displayed below in Figure 1.

<sup>2</sup> The archived experiment can be accessed by going to: <http://bit.ly/2o7cvws>.

<sup>3</sup> Demographics of the full sample and of the individual groups can be found in Table X1 in the Supplementary Appendix.

<sup>4</sup> The full list of available information is available in Table 5.

In the News Articles condition, subjects were asked to view two news articles, one dedicated to each candidate. Again, this condition, in particular, was designed to mimic commonly used survey experiments. Each news article conveyed five attributes of a candidate using the same wording available in the other conditions. The articles were both about 200 words and were viewable by clicking on a box with the respective candidate's name and picture. Both boxes appeared simultaneously on the screen, and the order of the boxes was randomized between subjects.

The Static Board condition created a computerized version of the classic “notecards on a board” process-tracing design used in marketing research. It listed the two candidates' names along the top, creating two columns, and then listed the 20 available attributes about the candidates along the side of the screen in rows. Below each candidate's name were a series of codes that could be entered that would reveal the relevant attribute about the candidate. Subjects would enter a code, view the information, and then return to the static board where they could input a new code.

The dynamic presentation conditions (both low- and high-information) entered subjects into a dynamic information board loaded with the available information. Each information box listed the candidate's name and picture, as well as the attribute the box contained. Each information item was available two times and the order of items was randomized for each subject. The boxes slowly scrolled down the screen and continued to scroll while subjects clicked on boxes and read the information inside. All information about the candidates was identical between the presentation conditions and differed only in presentation style and availability.

We propose that the high-information condition is more realistic in mimicking what voters face during most federal and statewide campaigns—whether they choose to learn it or not, there is a wealth of information available. Similarly, we believe that, by design, the dynamic conditions are more realistic than the static conditions, making information available to subjects without giving them complete control over the information environment and also ultimately allowing them to choose the information they access. We can rank the information environments in these conditions, then, from simplest to most complex and from least to most realistic: News Articles, Low-information Dynamic Board, Static Board, High-Information Dynamic Board.

## 4.2 Candidate gender manipulation

We manipulated the gender of the subjects' in-party candidate<sup>5</sup> so that half of the subjects viewed a man and half viewed a woman running for their party. We presented this information to subjects in three ways. First, we gave the candidates gendered names (Patrick/Patricia Martin for the Democrats and James/Jamie Anderson for the Republicans). Second, we associated pictures with the candidates and introduced subjects to the two candidates by presenting these pictures and the candidate names in an opening campaign synopsis page. We then used those pictures on all the information boxes to identify which candidate the box pertained to. Third, we used gendered pronouns (he/she, his/her, himself/herself) in the information items to refer to the candidates.

## 4.3 Hypotheses

We expect that the presentation (dynamic vs. static) and amount of information (low vs. high) will have significant and substantive effects on how subjects experience the study, evaluate the candidates, and react to the candidate gender manipulation. Our expectations are as follows:

<sup>5</sup> We only varied the sex of the in-party candidate because we believe that subjects devote more time to considering the in-party candidate regardless of which information search strategy they adopt (see Lau and Redlawsk 2006 for a fuller explanation). We determined the in-party candidate by asking the standard series of party identification questions, where those who identified as partisans were sorted into their respective parties along with those who identified themselves as “leaning” toward one party. For pure independents, we determined the in-party by comparing feeling thermometer ratings for “most Republicans” and “most Democrats.” The higher rating determined independent subjects' in-party candidate. All of our subjects were successfully sorted in this way, avoiding the use of any further tie-breaker criteria.

H1: Subjects in high-information conditions will be less likely to exhibit treatment effects than the low-information conditions. With more information available, we believe that the influence of the gender manipulation will be counterbalanced by individuating information about the candidates, decreasing or eliminating treatment effects.

H2: Further, we expect that the dynamic conditions will be less likely to exhibit treatment effects than static conditions. Because of the design of the dynamic boards, subjects are required to stay in the “campaign” for longer, leaving more time for information effects to dissipate. We expect that dynamically presented information will accordingly decrease the effects of the gender manipulation because they will take more time to complete, allowing the influence of any initial gender treatment effect to attenuate.

H3: Finally, we expect that the level of information available will be more influential than the style of presentation. Of the two information presentation manipulations, we believe that the availability of information will prove more important than how it is presented.

Thus, taking the above hypotheses together, we expect that treatment effects will be strongest in the News Articles group (low-information, static) condition, followed by the Low-Information Dynamic group, then the Static Board group, while the High-Information Dynamic group should produce the weakest treatment effects.

## 5 Results

We split our analysis into two sections: the treatment effects found from the candidate gender manipulation, and the behavioral differences observed between groups in the various conditions. We present the gender manipulation results first, then explain those differences with a more detailed explanation of what subjects experienced during the study.

### 5.1 Gender cues

We examine the role of the in-party candidate’s sex by using 10 dependent variables commonly used to evaluate candidates, particularly when examining the role of candidate sex. First, we use the more general way of measuring affect toward the candidates with the in-party candidate’s feeling thermometer score. We also assess subjects’ ratings of their in-party candidate on the 7-point liberal-conservative scale, looking at Republicans and Democrats separately. Then, we include the subject’s rating of the candidate on four trait assessments covering the in-party candidate’s compassion, competence, leadership and trustworthiness. Next, we use subject ratings of the in-party candidate’s ability to handle four types of issues; economic issues, military issues, helping the poor and closing the wage gap between men and women.<sup>6</sup> In all, this gives us 11 dependent variables to examine.

We treat each information group as an independent sample, as our interests are in how researchers conducting similar studies using different methods would view their results. Given the nature of our samples and dependent variables and to match previously published results, we calculate treatment effects using the `ttest` command in Stata.<sup>7</sup> The specific wording of the questions used can be viewed in the Supplementary Appendix.

We have three substantive findings in Table 1 (below). First, we find that our women candidates largely outperform the men, scoring higher in most of our candidate evaluation ratings, regardless of treatment group. We discuss this further in the conclusion of this section.

6 Like the background information available in the low-information conditions, these dependent variables were also taken from Huddy and Terkildsen (1993a,b).

7 We additionally calculate our results using difference in proportion tests using the `ranksum` command in Stata. Our traits and issues questions have only four levels, making difference in proportion tests more appropriate. However, previously published results (particularly Huddy and Terkildsen 1993a,b) have relied mainly upon difference in means tests, so we report those here for consistency. The results are similar and can be viewed in the Supplementary Appendix in Table X2.



**Table 1.** T-tests and treatment effects of In-Party evaluations, by Information Group.

In-Party evaluation on:	Cand Sex	News Articles		Low-Info Dyn Brd		Static Board		High-Info Dyn Brd	
		Mean	TE	Mean	TE	Mean	TE	Mean	TE
Feeling Thermometer	Man	63.68 (1.12)	-3.45 (2.41)	63.11 (1.60)	-4.36 (2.27)	68.18 (1.56)	0.15 (3.13)	64.89 (2.09)	-5.96 (3.05)
	Woman	67.13 (1.81)		67.47 (1.60)		68.02 (2.33)		70.85 (2.19)	
Lib-Con (Republicans)	Man	4.67 (0.25)	-0.42 (0.28)	4.68 (0.30)	-0.47 (0.38)	6.00 (0.23)	0.88 (0.39)	5.19 (0.24)	-0.41 (0.33)
	Woman	5.09 (0.16)		5.16 (0.24)		5.12 (0.23)		5.60 (0.21)	
Lib-Con (Democrats)	Man	3.22 (0.12)	0.03 (0.18)	3.20 (0.13)	0.12 (0.19)	2.62 (0.14)	0.14 (0.19)	2.82 (0.19)	0.28 (0.24)
	Woman	3.19 (0.13)		3.07 (0.14)		2.48 (0.10)		2.55 (0.14)	
Compassion	Man	3.20 (0.06)	-0.14 (0.08)	3.10 (0.05)	-0.22 (0.09)	3.26 (0.07)	0.00 (0.11)	3.17 (0.07)	-0.13 (0.10)
	Woman	3.34 (0.06)		3.32 (0.07)		3.26 (0.08)		3.30 (0.08)	
Competence	Man	3.34 (0.06)	-0.15 (0.08)	3.28 (0.07)	-0.16 (0.09)	3.38 (0.07)	0.06 (0.10)	3.32 (0.07)	-0.10 (0.10)
	Woman	3.49 (0.05)		3.44 (0.06)		3.32 (0.07)		3.43 (0.08)	
Leadership	Man	3.27 (0.06)	-0.00 (0.09)	3.17 (0.06)	-0.07 (0.09)	3.21 (0.06)	0.11 (0.09)	3.18 (0.07)	-0.10 (0.11)
	Woman	3.28 (0.06)		3.24 (0.06)		3.10 (0.07)		3.28 (0.08)	
Trustworthiness	Man	3.04 (0.06)	-0.30 (0.05)	3.08 (0.06)	-0.25 (0.09)	3.11 (0.07)	0.04 (0.10)	3.17 (0.07)	-0.02 (0.11)
	Woman	3.34 (0.06)		3.33 (0.07)		3.07 (0.07)		3.19 (0.08)	
Economic Issues	Man	2.98 (0.06)	-0.23 (0.09)	2.92 (0.07)	-0.24 (0.10)	3.13 (0.07)	0.07 (0.10)	3.05 (0.08)	-0.07 (0.11)
	Woman	3.21 (0.06)		3.16 (0.07)		3.05 (0.07)		3.12 (0.08)	
Military Issues	Man	2.74 (0.07)	0.02 (0.11)	2.80 (0.07)	-0.05 (0.11)	2.87 (0.07)	-0.02 (0.11)	2.95 (0.08)	-0.05 (0.12)
	Woman	2.72 (0.08)		2.85 (0.08)		2.89 (0.08)		3.00 (0.09)	
Helping the Poor	Man	3.14 (0.08)	-0.17 (0.10)	3.06 (0.07)	-0.14 (0.10)	3.16 (0.09)	-0.02 (0.12)	3.08 (0.09)	0.04 (0.13)
	Woman	3.31 (0.07)		3.20 (0.07)		3.17 (0.08)		3.04 (0.09)	
Gender Wage Gap	Man	2.96 (0.07)	-0.47 (0.10)	2.90 (0.07)	-0.45 (0.10)	3.17 (0.07)	-0.02 (0.10)	2.91 (0.08)	-0.27 (0.12)
	Woman	3.43 (0.07)		3.36 (0.07)		3.18 (0.07)		3.18 (0.08)	
Significant Findings		(n = 200)	5	(n = 189)	6	(n = 187)	1	(n = 200)	2

Another main finding in Table 1 is that the two low-information groups produce many more significant findings than do the two high-information groups. The News Articles group finds five significant differences in how men and women candidates are evaluated (on compassion, competence, trustworthiness, economic issues and the gender wage gap) while the Low-Information Dynamic condition produces six significant differences (feeling thermometer, compassion, competence, trustworthiness, economic issues and the gender wage gap). In contrast, the two maximum information groups barely produce any findings. The Static Board has only one significant result (lib-con for Republicans) while the High-Information Dynamic Board has two (feeling thermometer and gender wage gap).

Contrary to our expectations however, the dynamic conditions were not less likely to produce significant differences, and in fact produced slightly more. The Low-Information Board produced 6 significant findings, while the News Articles produced only 5. Similarly, the High-Information Dynamic Board produced 2 significant findings, compared with only 1 for the Static Board. While these are not large differences, they are contrary to our expectations.

Now, imagine that you were a researcher and conducted this study using only one of these groups, remembering that, when using the 0.05 significance level as the cutoff value, we would expect to produce about one false positive per 20 tests. Over these 11 tests, there is thus nearly an even 50–50 chance of producing at least one spurious significant result for each group.<sup>8</sup> Had we only run either the News Articles or Low-Information Dynamic Board group, we might easily reject the possibility that our findings were spurious, because approximately half were significant—far more than the expected error rate. However, had we only run the Static Board Group or High-Information Dynamic Board, our lackluster findings may lead us to believe that candidate sex played no substantial role in candidate evaluation. Notice that the general pattern of results does not change much between the four information groups (though the Static Board produces seven results, including the one significant finding, that are against the direction of the other groups). In the low-information groups, the differences are strong enough to produce significant results, while in the high-information groups this is not the case.

One could argue that this pattern of results was caused by a relatively small sample size (although 200 cases per group is hardly small), which would be corrected if only the sample had been larger. Perhaps the low-information groups are producing marginally stronger effects and with a larger sample size the Static Board and High-Information Dynamic Board groups would also produce similar significant results. Given that the general pattern of results we have seen thus far has demonstrated minimal differences between the static and dynamic groups, we can address this claim by pooling our groups into a binary classification solely based upon the level of information subjects were given access to. Doing so doubles the sample size in each group, and permits us to test the claim that these differences are simply a result of sample size.

Table 2 (above) replicates the previous t-tests, this time pooling the samples between the levels of information subjects had access to. With only two groups to compare, we can also now easily show difference-in-difference scores between the various treatment groups. In these tests, eight of the low-information group's tests produce significant differences, compared to only *one* of the high-information group's. This is a clear indication that the level of information subjects have access to drives the results that are produced in experiments. Interestingly, only three of the dependent variables produce significantly different treatment effects according to

8 Using multiple dependent variables in this manner necessitates the use of multiple-hypothesis correction to account for the increasing likelihood of false positives when running more tests. However, our intent here is to view this from the stance of a researcher conducting an initial analysis, as opposed to conducting appropriate statistical corrections when reporting results. We do conduct and report Holms–Bonferroni corrections (see Holm 1979; Gaetano 2013) on all of our difference in means and proportions tables in the Supplementary Appendix (Tables X4, X5, X6 and X7). The pattern of results remain the same.

**Table 2.** T-tests and DIDs on 10 In-Party evaluation measures, by Level of Information manipulation.

In-Party evaluation on:	In-Party Candidate		Low Information		High Information			
	Sex	N	Mean	TE	N	Mean	TE	DID
Feeling Thermometer	Man	199	63.39 (1.12)	-3.90 (1.65)	206	66.41 (1.48)	-3.01 (2.18)	-0.89 (2.74)
	Woman	190	67.28 (1.22)		181	69.41 (1.60)		
Lib-Con (Republicans)	Man	34	4.68 (0.20)	-0.45 (0.24)	43	5.40 (0.20)	0.04 (0.25)	-0.49 (0.34)
	Woman	41	5.12 (0.14)		51	5.35 (0.16)		
Lib-Con (Democrats)	Man	129	3.21 (0.09)	0.07 (0.13)	123	2.72 (0.12)	0.20 (0.15)	-0.13 (0.20)
	Woman	127	3.14 (0.10)		103	2.51 (0.09)		
Compassion	Man	199	3.15 (0.04)	-0.19 (0.06)	206	3.21 (0.05)	-0.07 (0.07)	-0.12 (0.10)
	Woman	190	3.33 (0.04)		181	3.28 (0.05)		
Competence	Man	199	3.31 (0.04)	-0.16 (0.06)	206	3.35 (0.05)	-0.02 (0.07)	-0.14 (0.09)
	Woman	190	3.47 (0.04)		181	3.37 (0.05)		
Leadership	Man	199	3.22 (0.04)	-0.04 (0.06)	206	3.19 (0.05)	-0.01 (0.07)	-0.05 (0.09)
	Woman	190	3.26 (0.04)		181	3.19 (0.05)		
Trustworthiness	Man	199	3.06 (0.04)	-0.28 (0.06)	206	3.14 (0.05)	0.01 (0.07)	-0.29 (0.10)
	Woman	190	3.34 (0.05)		181	3.13 (0.05)		
Economic Issues	Man	199	2.95 (0.05)	-0.24 (0.07)	206	3.09 (0.05)	0.00 (0.07)	-0.24 (0.10)
	Woman	190	3.19 (0.05)		181	3.09 (0.05)		
Military Issues	Man	199	2.77 (0.05)	-0.01 (0.08)	206	2.92 (0.05)	-0.03 (0.08)	-0.02 (0.11)
	Woman	190	2.78 (0.06)		181	2.94 (0.06)		
Helping the Poor	Man	199	3.10 (0.05)	-0.15 (0.07)	206	3.12 (0.06)	-0.01 (0.09)	-0.16 (0.11)
	Woman	190	3.26 (0.05)		181	3.11 (0.06)		
Gender Wage Gap	Man	199	2.93 (0.05)	-0.47 (0.07)	206	3.03 (0.05)	-0.15 (0.08)	-0.31 (0.11)
	Woman	190	3.39 (0.05)		181	3.18 (0.06)		
Significant Findings				8			1	3

the difference-in-difference tests.<sup>9</sup> This indicates that both types of studies are producing similar treatment effects, but that larger variance in the higher-information groups is preventing significant results from emerging. This in turn suggests that greater information may be causing some, but perhaps not all, subjects within these groups to alter their behavior.

## 5.2 Conclusions—gender analysis

These findings strongly suggest that the manner in which experiments allow subjects to learn about political candidates have serious repercussions for how those candidates are evaluated, and what conclusions we draw from the study. The two manipulations in information presentation we examine here—the level of information and the presentation style—are not equally influential. Supporting our first and third hypotheses, the level of information seems to produce much stronger differences and is the factor that drives the results we find. Depending on whether we ran this study as a survey experiment—as in the News Articles group—or as a high-information static or dynamic processing tracing study, we would draw very different conclusions. Subjects who could view more information about our candidates exhibited lower treatment effects, producing far fewer significant results. We can safely conclude that the design of the study does influence the types of conclusions a researcher is likely to draw and that low-information studies seem much more likely to produce significant findings.

Substantively, most of our results from the low-information conditions are very much in line with current literature. Previous experimental evidence would lead us to expect female candidates to be rated as more compassionate and trustworthy, and better at handling “feminine” issues like dealing with the wage gap, and that is indeed what we find in our study. In both low-information conditions, these findings are statistically significant and in high-information conditions the pattern of results is nearly identical, but not significant. However, there are also instances in which we find no difference between male and female candidates when we expected one to exist (leadership, military issues, helping the poor), and there are also two dependent variables for which gender stereotypes seemed to work in the opposite way from what we expected (competence and economic issues). Interestingly, these less-expected results are consistent with some of the more recent work on gender stereotypes (Dolan 2010; Brooks 2013; Dolan 2014, e.g.). Dolan (2010), for example, finds no difference in stereotypic evaluations of men’s and women’s ability to handle the economy, or in their levels of ambition or assertiveness.

Women candidates are also rated more highly on feeling thermometer scores in both dynamic conditions, which suggests that gender may actually be a net benefit for women candidates in our study, regardless of information condition. This is also consistent with a number of previous studies (Sanbonmatsu 2002; Dolan 2004; Lawless 2004; Dolan 2010; Ditonto 2017), many of which find that women candidates can and do benefit from gender-based stereotypes in certain contexts.

## 6 Subject Behavior Results

We now seek to explain why these differences emerge. What is it about these different information presentation styles that lead subjects to behave so differently? We suggest that there are three main factors at play: the time subjects spend in the experiment, the level of information they encounter, and the importance of the information they view.

### 6.1 Time

One way in which differences can manifest in an experimental study is through the time subjects spend gathering, reading and considering the information they encounter. Particularly in a study such as this, where the treatment is viewed early on (though reinforced throughout), this greater amount of time provides an opportunity for the treatment effect to attenuate naturally. Table 3,

<sup>9</sup> Calculated using the `ttesti` command in Stata.

**Table 3.** One-way Anova’s of time in experiment, by Information Group.

	Information Group	N	Mean	F Stat	Scheffe Group
Pre-Q	News Articles	200	260.22 (11.97)	0.83	1
	Low-Info Dyn Brd	189	288.12 (16.43)		1
	Static Board	200	291.52 (17.97)		1
	High-Info Dyn Brd	187	279.17 (15.38)		1
	Total	776	279.44 (7.74)		
Practice	News Articles	200	148.47 (4.60)	15.138	1
	Low-Info Dyn Brd	189	176.27 (8.15)		1
	Static Board	200	228.02 (13.66)		2
	High-Info Dyn Brd	187	166.24 (6.52)		1
	Total	776	178.99 (4.48)		
Campaign	News Articles	200	115.02 (5.30)	210.36	1
	Low-Info Dyn Brd	189	113.83 (3.61)		1
	Static Board	200	306.35 (11.34)		3
	High-Info Dyn Brd	187	277.12 (6.06)		2
	Total	776	202.61 (4.77)		
Post-Q	News Articles	200	153.28 (4.99)	0.28	1
	Low-Info Dyn Brd	189	157.63 (6.68)		1
	Static Board	200	162.61 (6.69)		1
	High-Info Dyn Brd	187	156.46 (9.71)		1
	Total	776	157.40 (3.62)		
Total	News Articles	200	680.84 (18.14)	33.85	1
	Low-Info Dyn Brd	189	739.62 (22.88)		1
	Static Board	200	993.05 (31.51)		3
	High-Info Dyn Brd	187	882.96 (23.00)		2
	Total	776	822.48 (12.82)		

below, shows the average time subjects took to complete each substage within the experiment, and in total to complete the entire study. Clear differences in total time emerge, and reviewing the substage information makes it clear that these time differences come from where we would expect them to—the practice session and the actual campaigns where subjects are exposed to information.

Unsurprisingly, subjects who were in the low-information conditions (News Articles and Low-Information Dynamic Board) spent far less time in the study overall, because they had less information to view, and thus less to actually do. Subjects in the News Articles group completed the study quickest, taking on average about 680 seconds, or 11 minutes. The Low-Information Dynamic Board was close to this, at about 740 seconds. While each of the four groups averaged a different average completion time, a Scheffe test (using a 0.05 significance level) demonstrates that the two low-information groups were statistically indistinguishable, but were both different than the two high-information groups. Interestingly, the Static Board group took significantly longer than the High-Information Dynamic Board group, requiring about 993 seconds on average compared with 883 seconds. Subjects in these two groups spent much more time learning about the candidates.

While this is due to the amount of information available to subjects, it is also a consequence of the design of the overall study. In order to proceed out of each section of the experiment, subjects must complete a certain task. In the pre- and postquestionnaire stages, subjects all answered the

same questions, so predictably took similar amounts of time. In the information-providing stages however, subjects necessarily faced different tasks. In the News Articles group, subjects were asked to read the two articles, and then were free to progress,<sup>10</sup> while in the Static Board subjects were forced to view at least 5 individual items of information of their choice before moving on.<sup>11</sup> Thus subjects in the Static Board group could choose to view 5 items about their in-party candidate and never learn anything about their opponent. Those in the News Articles group did not have that option (though they could still open an article and simply not read it). The requirements in these two scenarios were information-dependent, forcing subjects to encounter a certain number of information items.

While the two static groups were information-dependent, the dynamic information boards were time-dependent, and forced subjects to remain within the stage until all of the available information had scrolled by.<sup>12</sup> This took longer, but did not require subjects to actually view any information if they did not wish to. It was possible for subjects to view nothing at all and still advance through to the next substage of the study (though no subject actually did this—the minimum number of unique items opened was 2, and the minimum number of boxes was 3).

## 6.2 Information viewed

The differences in behavior between the information groups are strikingly apparent in the level of information about the candidates that those subjects viewed—and somewhat unexpected. There are two primary ways to examine the information subjects viewed—based upon the number of unique attributes viewed, and by the total number of information items opened. The count of unique items viewed records how many different attributes subjects chose to expose themselves to—that is, how many pieces of information about the candidates they chose to look at. This measure does not take into account if subjects view an item multiple times, but simply that they viewed an item at least once. However, subjects will oftentimes return to re-examine previously viewed information, meaning that the number of items opened will sometimes be far greater than the number of unique items viewed. Examining both measures provides a greater window into how subjects learned about the candidates.

Table 4 (below) shows the differences in the number of unique items viewed and the total items opened for each of the information presentation groups. Subjects in the Static Board group, despite taking the most amount of time during the Campaign stage, on average viewed the least amount of unique information, about seven items total. A Scheffe test demonstrates that this is statistically the same as the Low-Information Dynamic Board, where subjects tended to view about eight unique items. This is interesting in that, even though the Static Board group had four times the information available as the Low-Information Dynamic Board group, they both viewed statistically equally amounts of information. And the Static Information Board group took much longer to do so! As a contrast, the High-Information Dynamic Board produced by far the most information viewed, at almost 22 attributes viewed; three times the information in less time than the Static Board.

Interestingly, participants in the two high-information groups ended up evaluating the candidates very similarly, despite massive differences in how much they actually learned about them. Members of the Static Board group viewed far less information, on average,

<sup>10</sup> The program required that both news articles be read, but allowed subjects to read each one as many times as they wished. This is what most survey experiments require participants to do.

<sup>11</sup> After viewing 5 items subjects were provided with a special code that would permit them to proceed to the postquestionnaire. We decided to put the bar at 5 items because that is equivalent to the amount of information available about each candidate in the News Articles condition.

<sup>12</sup> There are numerous ways to allow subjects to proceed, including allowing them to choose when to advance on to the vote decision. We selected to keep them in the “campaign” for its full duration to ensure that subjects were in fact able to encounter all of the available information in all of the information groups. This replicates the real-world example of a political campaign, where most people still wait until Election Day to vote.

**Table 4.** One-way ANOVA's of information viewed, by Information Group.

Information Group	Unique Information Viewed				Total Information Items Viewed			
	N	Mean	F Stat	Scheffe Group	N	Mean	F Stat	Scheffe Group
News Articles	200	10.00 (0.00)		2	200	11.60 (0.34)		2
Low-Info Dyn Board	189	8.17 (0.12)		1	189	11.32 (0.31)		2
Static Board	187	7.14 (0.23)		1	187	9.34 (0.25)		1
High-Info Dyn Board	200	21.64 (0.47)		3	200	25.88 (0.78)		3
Total	776	11.87 (0.25)	617.42		776	14.67 (0.34)	258.69	

than High-Information Dynamic Board members and yet seemed to evaluate the candidates statistically identically. This is a strong indication that it is not necessarily viewing more information that is affecting subjects' evaluations of the candidates, but having *access* to certain kinds of information, and perhaps choosing to view information that is particularly influential.

It is worth noting that the News Articles Group is in its own Scheffe group, but this is due to the absence of variance in the number of items subjects viewed. Because each news article revealed five attributes about each candidate, and subjects were required to read both articles, we must assume that all of the subjects in this condition viewed the 10 available items with no variation between subjects. By lumping all of the available information into a single article, we have no choice but to assume that subjects fully read and paid attention to every portion of the text, even though we cannot verify this. While we can never truly be certain that subjects attend to any information they are exposed to (aside from perhaps using eye-tracking software combined with recall tests), presenting each piece of information in its own "box" (as the two dynamic groups and the Static Board do) lets us know for certain when subjects seek specific information, and thus conversely when they are *not* exposed to an item.

The pattern of viewing information changes slightly when we consider the total number of items opened. Using this metric, we can see that subjects in the Low-Information Dynamic Board group viewed more information, on average, than did subjects in the Static Board group, despite having much less information available to them. The High-Information Dynamic Board group again views much more information than the other conditions, at about 26 items. The News Articles group gets a slight boost here, with some subjects choosing to read the articles multiple times, raising the average items viewed to 11.60. Contrary to what we might have expected at the outset, subjects who had full control over a high-information environment (the Static Board) chose to view the fewest items out of all the groups, and were exposed to less information than the subjects who had only 25% of that information available to them.

### 6.3 Type of information

A final area in which differences in candidate evaluation can be generated is in the types of information subjects viewed. In each of our four conditions, subjects had access to the same five background pieces of information about the candidates, which are similar to information routinely used in vignette experiments, particularly those studying the effects of candidate attributes like gender. In the high-information conditions, we augmented this information with policy stances and general ideological information about the candidates. This allows us to compare what information subjects choose to view when they have no control over the information environment (News Articles), some control (Dynamic Boards), and total control (Static Information Board).

**Table 5.** Percentage of subjects viewing attribute, by Information Group.

Candidate Information	News Articles	Low-Info Dynamic Board	Static Board	High-Info Dynamic Board
Gun Control Policy	—	—	26.74%	86.50%
Taxation Policy	—	—	27.81%	84.50%
Health Care Policy	—	—	29.95%	84.00%
Abortion Policy	—	—	46.52%	83.50%
Immigration Policy	—	—	20.32%	82.50%
Defense Budget	—	—	18.18%	81.00%
Jobs Policy	—	—	29.95%	80.50%
Social Philosophy	—	—	34.76%	80.50%
Terrorism Policy	—	—	25.67%	79.50%
Crime Policy	—	—	16.58%	79.00%
Education Policy	—	—	18.18%	78.00%
Energy Policy	—	—	8.56%	77.50%
Economic Philosophy	—	—	46.52%	77.50%
Global Warming Stance	—	—	37.43%	77.00%
Iran Policy	—	—	6.95%	76.50%
Religion	100.00%	93.65%	18.72%	73.00%
Editorial About	100.00%	95.24%	10.70%	67.50%
Education	100.00%	92.59%	13.90%	65.50%
Family Background	100.00%	91.01%	5.35%	63.00%
Political Experience	100.00%	97.35%	13.90%	57.00%

Table 5 shows the percentage of subjects within each information group who selected to view each attribute (for either candidate). We rank the attributes by the percentage of subjects within the High-Information Dynamic Board who chose to view the attribute, because this is the group that tended to view the most information and we believe to be the most realistic scenario.

What we find is again striking—the five attributes we included in the minimal conditions place in the bottom five slots of views in the High-Information Dynamic Board. That is—the types of information typically used in survey experiments alongside the treatment is the least desirable information for our subjects to want to view when given other options. If the intent of providing background information of this type in survey experiments is to avoid contaminating subjects’ decision-making processes with other considerations, we can now support this as a well-crafted design—subjects clearly have little interest in background information and do not seem to seek it out when making decisions.

However, this is also a strong indication of why we find such large differences in treatment effects between the high- and low-information groups. Background information in itself is simply not appealing to subjects in campaign style experiments, and presents little additional information for subjects to use when evaluating candidates. The result is that the treatment information—in this case the gender of the in-party candidate—is exaggerated in its importance because it is important relative to the other information available to draw from when evaluating a candidate. This is not to say that the treatment effect in low-information studies is wrong, only that it is exaggerated. By denying subjects the ability to access information that they might otherwise use to evaluate candidates, low-information studies force subjects to use treatment information alone. While the low-information conditions may accurately simulate very low-level elections, they certainly do not mimic higher-level national elections, which are those most commonly studied by political scientists.



## 7 Subject Behavior Conclusions

From this examination of how subjects spent their time within the experiment in the four information presentation groups, we get a sense of why there are such great differences between the findings in the gender cue analysis. We see significant differences by group in the amount of time spent in the study, the amount of information accessed, and the type of information that subjects cared about. These findings complicate our understanding of experimental design, however. Contrary to our expectations, giving subjects access to more information did not guarantee that they looked at more information. The group that took the longest to complete the study was also the group that viewed the least amount of information. And yet they acted remarkably similar to the High-Information Dynamic Group that viewed much greater amounts of information in less time. Given this design, we have little ability to tease out why this is the case, but we do now know that this is an important area for follow-up research.

The larger question we are left with, as researchers, is: which method is best for accomplishing our research goals? What we find here is that perhaps there is no single best answer. Low-information experiments seem better at determining if treatments *can* have effects, and whether they do in very low-information elections. High-information experiments appear better at determining if treatments have effects when subjects have other information at their disposal. In a real election, either of these types of studies may best mimic reality, depending upon the level of office and the amount of media attention for a particular race. We know that candidates running for office do not all have the same ability to inform the electorate about their campaigns, creating unique information environments around each office. For presidential candidates, information floods the media environment, almost guaranteeing that citizens learn at least some attributes about the candidates. In such races, experiments should likely mimic this and design high-information studies and we may want to approach low-information designs with greater skepticism.

But not all offices are like this. Lower-ballot elections, such as state legislative races and local contests suffer from much lower campaign spending and media attention. In these situations, low-information experiments such as survey experiments may be more accurate, because they better mimic the information environments that typically exist. Still, we do wonder whether restricting information from subjects mimics this situation better than does providing information and giving subjects the freedom to choose whether or not they wish to view it. In theory, all citizens can simply Google their local candidates and find out a great deal about them, even if few people actually do so.

## 8 General Discussion and Conclusions

Low-information survey experiments can clearly demonstrate that various treatments can produce behavioral effects, and field experiments can clearly demonstrate when effects do occur in the real world. The downsides to these two types of experiments are also apparent. Survey experiments can lack external validity, and unrealistically bar access to information that might diminish treatment effects. Field experiments are, at least in part, dependent on the events of the real world, forcing researchers to tailor their research questions to the available political environment (it is difficult to imagine how we could have run a field experiment in the scenario used here). We believe that high-information laboratory experiments are a possible middle ground, where researchers have the freedom to create scenarios they are interested in studying and a more realistic environment that allows treatment effects to dissipate.

The case for high-information process-tracing experiments is not perfect, however. Among the other findings, we do show that high-information experiments take longer, and thus will require larger payments for subject participation (Andersen and Lau 2018; Zechmeister 2015). Given confined research budgets, this means that such studies will likely draw smaller subject pools,

making examining subgroups within the sample more difficult. It is within these subgroups that the most interesting developments are likely to be found in future studies. We suspect (though do not present evidence here) that the drop in significant treatment effects is created by some subjects reacting to acquired information that allows the treatment to have a weaker effect. We doubt that this happens equally to all participants, and believe that this is likely localized to certain subgroups, possibly the most politically sophisticated participants who are most likely to process new information and update their evaluations in accordance.

We see high-information experiments as a useful tool for political scientists, adding an additional layer of realism and complexity over traditional vignette-style experiments. Future developments can continue this progress, particularly by lengthening the duration of studies (over multiple days or weeks, for instance) and testing the effects of a wider variety of types of information (topics that subjects gravitate toward vs. those they ignore). This study has demonstrated that our variable of interest—candidate gender—did not produce significant treatment effects across a variety of methods. However, we believe that is because the effect of gender can be moderated by other information. It is less clear if things like partisanship or declared ideology would be similarly affected by additional information, unless it was directly contradictory. There is still a great deal of room for studying what information is influential to voters, and how the overall information environment influences the effect of any single item of information. In summary, we believe that by complicating the information environment we can create more externally valid studies that will better capture how people learn about and evaluate the political world.

## Supplementary material

For supplementary material accompanying this paper, please visit

<https://doi.org/10.1017/pan.2018.21>.

## References

- Alexander, Deborah, and Kristi Andersen. 1993. Gender as a factor in the attribution of leadership traits. *Political Research Quarterly* 46(3):527–545.
- Andersen, David. 2018. Replication data for: Information and its presentation: Treatment effects in low-information vs. high-information experiments, <https://doi.org/10.7910/DVN/TGFAOH>, Harvard Dataverse, V1, UNF:6:Nxzb+4xRVIDBTMLctotUVQ==.
- Andersen, David J., and Richard R. Lau. 2018. Pay rates and subject performance in social science experiments using crowdsourced online samples. *Journal of Experimental Political Science*, doi:10.1017/XPS.2018.7.
- Anderson, John. 1983. *The architecture of cognition*. Cambridge, MA: Harvard University Press.
- Banducci, Susan, Jeffrey Karp, Michael Thrasher, and Colin Rallings. 2008. Ballot photographs as cues in low-information elections. *Political Psychology* 29(6):903–917.
- Barabas, Jason, and Jennifer Jerit. 2010. Are survey experiments externally valid? *American Political Science Review* 104(2):226–242.
- Berinsky, Adam, Gregory Huber, and Gabriel Lenz. 2012. Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Political Analysis* 20:351–368.
- Brooks, Deborah Jordan. 2013. *He runs, she runs: Why gender stereotypes do not harm women candidates*. Princeton, NJ: Princeton University Press.
- Brooks, Deborah Jordan, and John Geer. 2007. Beyond negativity: The effects of incivility on the electorate. *American Journal of Political Science* 51:1–16.
- Burrell, Barbara. 1994. *A woman's place is in the House: Campaigning for congress in the feminist era*. Ann Arbor: University of Michigan Press.
- Carroll, Susan J., and Kelly Dittmar. 2010. The 2008 candidacies of Hillary Clinton and Sarah Palin: Cracking the “highest, hardest glass ceiling”. In *Gender and elections: Shaping the future of American politics*, ed. Susan J. Carroll and Richard L. Fox. New York: Cambridge University Press.
- Carroll, John S., and Eric Johnson. 1990. *Decision research: a field guide*. Newbury Park: Sage Publications.
- Carsey, Thomas, and Gerald Wright. 1998. State and national factors in gubernatorial and senatorial elections. *American Journal of Political Science* 42(3):994–1002.

- Cook, Elizabeth Adell, Sue Thomas, and Clyde Wilcox, eds. 1994. *The year of the woman: Myths and realities*. Boulder: Westview Press.
- Darcy, R., S. Welch, and J. Clark. 1997. *Women, elections, and representation*. second ed. New York: Longman.
- Ditonto, Tessa. 2017. A high bar or a double standard? Gender, competence, and information in political campaigns. *Political Behavior* 39(2):301–325.
- Ditonto, Tessa, Allison Hamilton, and David Redlawsk. 2014. Gender stereotypes, information search, and voting behavior in political campaigns. *Political Behavior* 36(2):335–358.
- Dolan, Kathleen. 2004. *Voting for women how the public evaluates women candidates*. Boulder, CO: Westview Press.
- Dolan, Kathleen. 2010. The impact of gender stereotyped evaluations on support for women candidates. *Political Behavior* 32(1):69–88.
- Dolan, Kathleen. 2014. *When does gender matter? Women candidates and gender stereotypes in American elections*. New York: Oxford University Press.
- Druckman, James, Donald Green, James Kuklinski, and Arthur Lupia. 2006. The growth and development of experimental research in political science. *American Political Science Review* 100(4):627–635.
- Ericsson, Anders, and Herbert Simon. 1980. Verbal reports as data. *Psychological Review* 87(3):215–251.
- Fiske, Susan T., and Steven L. Neuberg. 1990. A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. *Advances in Experimental Social Psychology* 23:1–74.
- Gaetano, Justin. 2013. Holm-Bonferroni sequential correction: An EXCEL calculator (corrected by Pawel Kleka (2015)) Retrieved from [http://www.staff.amu.edu.pl/~kleka/\\_uploads/Holms-correction-calculator.xlsx](http://www.staff.amu.edu.pl/~kleka/_uploads/Holms-correction-calculator.xlsx).
- Gaines, Brian J., James Kuklinski, and Paul Quirk. 2007. The logic of the survey experiment reexamined. *Political Analysis* 15(winter):1–20.
- Gerber, Alan, and Donald Green. 2012. *Field experiments: design, analysis and interpretation*. New York: W.W. Norton Publishing.
- Gilens, Martin. 2001. Political ignorance and collective policy preferences. *American Political Science Review* 95:379–396.
- Hastie, Reid. 1986. A primer of information-processing theory for the political scientist. In *Political cognition: The nineteenth annual Carnegie Symposium on political science*, ed. Richard Lau and David Sears. Hillsdale, NJ: Erlbaum, pp. 11–39.
- Hayes, Danny. 2011. When gender and party collide: Stereotyping in candidate trait attribution. *Politics & Gender* 7(2):133–165.
- Higgle, Ellen, Penny M. Miller, Todd G. Shields, and Mitzi M. S. Johnson. 1997. Gender stereotypes and decision context in the evaluation of political candidates. *Women and Politics* 17(3):69–88.
- Highton, Benjamin. 2004. Policy voting in senate elections: The case of abortion. *Political Behavior* 26(2):181–200.
- Holm, Sture. 1979. A simple sequential rejective method procedure. *Scandinavian Journal of Statistics* 6:65–70.
- Huddy, Leonie, and Nayda Terkildsen. 1993a. Gender stereotypes and the perception of male and female candidates. *American Journal of Political Science* 37(1):119–147.
- Huddy, Leonie, and Nayda Terkildsen. 1993b. The consequences of gender stereotypes for women candidates at different levels and types of office. *Political Research Quarterly* 46(3):503–525.
- Iyengar, Shanto. 2011. Laboratory experiments in political science. In *Handbook of experimental political science*, ed. Druckman James, Donald Green, James Kuklinski, and Arthur Lupia. New York: Cambridge University Press.
- Jacoby, Jacob, Donald Speller, and Carol Kohn. 1974. Brand choice behavior as a function of information load. *J. Mar. Res.* 11(1):63–69.
- Jerit, Jennifer, Jason Barabas, and Scott Clifford. 2013. Comparing contemporaneous laboratory and field experiments on media effects. *Public Opinion Quarterly* 77(1):256–282.
- Kahn, Kim Fridkin. 1996. *The political consequences of being a woman: How stereotypes influence the conduct and consequences of political campaigns*. Columbia.
- Kinder, Donald. 2007. Curmudgeonly advice. *Journal of Communication* 57(March):155–162.
- Koch, Jeffrey. 2000. Do citizens apply gender stereotypes to infer candidates' ideological orientations? *The Journal of Politics* 62:414–429.
- Koch, Jeffrey. 2002. Gender stereotypes and citizens' impressions of house candidates' ideological orientations. *American Journal of Political Science* 46(2):453–462.
- Lau, Richard R. 1995. Information search during an election campaign: introducing a process tracing methodology for political scientists. In *Political judgment: Structure and Process*, ed. M. Lodge and K. McGraw. Ann Arbor, MI: University of Michigan Press, pp. 179–206.
- Lau, Richard R., David J. Andersen, and David P. Redlawsk. 2008. An exploration of correct voting in recent presidential elections. *American Journal of Political Science* 52(2):395–411.

- Lau, Richard R., and David P. Redlawsk. 1997. Voting correctly. *American Political Science Review* 91(September):585–599.
- Lau, Richard R., and David P. Redlawsk. 2006. *How voters decide: Information processing during election campaigns*. New York: Cambridge University Press.
- Lawless, Jennifer. 2004. Women, war and winning elections: Gender stereotyping in the post-September 11th era. *Political Research Quarterly* 57(3):479–490.
- Leeper, M. S. 1991. The impact of prejudice on female candidates: An experimental look at voter inference. *American Politics Quarterly* 19(2):248–261.
- Lodge, Milton, and Ruth Hamill. 1986. A partisan schema for political information processing. *American Political Science Review* 80(2):505–519.
- Lodge, M., P. Strohm, and J. Wahlke. 1990. Black-box models of evaluation. *Political Behavior* 12(1):5–18.
- Lodge, M., M. Steenbergen, and Shawn Brau. 1995. The responsive voter: Campaign information and the dynamics of candidate evaluation. *American Political Science Review* 89(2):309–326.
- Matson, Marsha, and Terri Susan Fine. 2006. Gender, ethnicity, and ballot information: Ballot cues in low-information elections. *State Politics and Policy Quarterly* 6(1):49–72.
- McDermott, Monika, and David Jones. 2005. Congressional performance, incumbent behavior and voting in Senate Elections. *Legislative Studies Quarterly* 30:235–247.
- McDermott, Monika L. 1997. Voting cues in low-information elections: Candidate gender as a social information variable in contemporary United States elections. *American Journal of Political Science* 41(January):270–283.
- McDermott, Monika. 1998. Race and gender cues in low-information elections. *Political Research Quarterly* 51(4):895–918.
- McDermott, Rose. 2002. Experimental methods in political science. *Annual Review of Political Science* 5:31–61.
- McGraw, Kathleen, Milton Lodge, and Patrick Strohm. 1990. On-line processing in candidate evaluation: The effects of issue order, issue importance, and sophistication. *Political Behavior* 12(1):41–58.
- Morton, Rebecca, and Kenneth Williams. 2010. *Experimental political science and the study of causality: From nature to the lab*. New York: Cambridge University Press.
- Mutz, Dianna. 2011. *Population-based survey experiments*. Princeton, NJ: Princeton University Press.
- Payne, J. W. 1976. Task complexity and contingent processing in decision making: An information search and protocol analysis. *Organizational Behavior and Human Performance* 16:366–387.
- Redlawsk, David P. 2004. What voters do: Information search during election campaigns. *Political Psychology* 25(August):595–610.
- Sanbonmatsu, Kira. 2002. Gender stereotypes and vote choice. *American Journal of Political Science* 46:20–34.
- Sanbonmatsu, Kira, and Kathleen Dolan. 2009. Do gender stereotypes transcend Party? *Political Research Quarterly* 62(3):485–494.
- Sapiro, Virginia. 1981. If US Senator Baker were a woman: An experimental study of candidate images. *Political Psychology* 3(1/2):61–83.
- Seltzer, Richard, Jody Newman, and Melissa Voorhees Leighton. 1997. *Sex as a political variable: Women as candidates and voters in US elections*. Boulder, CO: Lynn Reinner Publications.
- Sigelman, Lee, and Carol K. Sigelman. 1982. Sexism, racism, and ageism in voting behavior: An experimental analysis. *Social Psychology Quarterly* 45(4):263–269.
- Simon, Herbert A. 1979. Information processing models of cognition. *Annual Review of Psychology* 30:363–396.
- Woods, Harriet. 2000. *Stepping up to power*. Boulder, CO: Westview Press.
- Zaller, John. 1992. *The nature and origins of mass opinion*. New York: Cambridge University Press.
- Zaller, John, and Stanley Feldman. 1992. A simple theory of the survey response: answering questions versus revealing preferences. *American Journal of Political Science* 36:579–616.
- Zechmeister, Elizabeth. 2015. Ethics and research in political science: The responsibilities of the researcher and the profession. In *Ethical Challenges in Political Science Experiments*, ed. Scott Desposato. New York: Routledge, pp. 255–261.