Psychologically authentic versus inauthentic replication attempts

Christopher J. Bryan\textsuperscript{a,1}, Gregory M. Walton\textsuperscript{b}, and Carol S. Dweck\textsuperscript{b}

What is an authentic replication attempt and what is not? Gerber et al.’s paper \cite{1} gives us the opportunity to reflect on this issue of longstanding concern to us. Gerber et al. \cite{1} attempted to replicate our 2011 research \cite{2}, which showed that referring to voting with nouns rather than verbs in a pre-election survey (e.g., “How important is it to you to [be a voter/ vote] . . .”) could increase turnout. Although their study bears a superficial resemblance to ours, it is not an authentic replication. Gerber et al. \cite{1} do not create the psychological context in which the phenomenon could plausibly emerge.

As described in our paper \cite{2}, noun wording can offer one the opportunity to claim a valued identity by engaging in the relevant behavior. We tested our hypothesis in high-profile elections that received substantial public attention: the 2008 United States presidential election and the 2009 New Jersey gubernatorial election. In these cases, the opportunity to be “a voter” feels like a valued identity that can motivate behavior. Contrast this with the congressional primaries examined by Gerber et al. \cite{1}: The outcomes of almost none were in doubt—nearly half were uncontested—and few received any meaningful attention. Consider the 12 of 61 major party House primaries that Gerber et al. call “competitive” \cite{1}; one was uncontested and six others were decided by huge margins (20.5, 32.8, 33, 39, 47.8, and 58.6 percentage points). In reality, only 4 of those 61 primaries were meaningfully competitive by even a loose standard (details at https://osf.io/g96sc/).

To demonstrate the psychological significance of this difference, we asked 366 online participants to imagine either an election like the New Jersey gubernatorial race or a congressional primary. Participants reported how important and positive the identity “voter” would feel in each scenario. As expected, they judged voting in the high-profile election to have far more important and positive identity implications ($t = 8.31, P < 0.0005$, $d = 0.85$; details at https://osf.io/g96sc/).

Psychological experiments often seem simple. This is deceptive. They are predicated on a careful analysis of psychological processes and the contextual factors that influence them. An authentic replication begins with this psychological understanding \cite{3}. In the best cases it attempts to extend this understanding by, for example, directly comparing contexts in which a phenomenon is likely to emerge to ones in which it is not. In the present case, an electoral context that allows the identity “voter” to feel important is necessary to motivate behavior.

Misunderstandings like this may be inevitable as social sciences become more interdisciplinary. To prevent them, we can articulate our theoretical and methodological assumptions in our papers so they are clear to nonspecialists. And we can share expertise across disciplinary lines before attempting replications.

Understanding each other’s expertise and pooling that expertise through collaboration is essential. By such understanding and sharing we can pursue useful and rigorous science together.

\begin{thebibliography}{9}
\end{thebibliography}

\textsuperscript{a}Booth School of Business, University of Chicago, Chicago, IL 60637; and \textsuperscript{b}Stanford University, Stanford, CA 94305

Author contributions: C.J.B., G.M.W., and C.S.D. designed research; C.J.B. performed research; C.J.B. analyzed data; and C.J.B., G.M.W., and C.S.D. wrote the paper.

The authors declare no conflict of interest.

\textsuperscript{1}To whom correspondence should be addressed. Email: christopher.bryan@chicagobooth.edu.
Authentic Versus Inauthentic Replication Attempts

Online Supporting Information

Authors:
Christopher J. Bryan,*,† Gregory M. Walton,* and Carol S. Dweck*§

Affiliations:
* University of Chicago, Booth School of Business
† Stanford University
§ Correspondence to: christopher.bryan@chicagobooth.edu
Competitiveness of elections in Gerber and colleagues’ sample

Gerber and colleagues conducted their experiment during the August 2014 primary elections in Michigan, Tennessee, and Missouri. This included 61 major party U.S. House primaries, 4 major party U.S. Senate primaries, and 4 major party state gubernatorial primaries. In their online supporting information, Gerber and colleagues indicate that approximately 42% of their sample was drawn (oversampled) from the population of eligible voters in 12 U.S. House primary races that they describe as “competitive.” The criteria by which they made this judgment were not provided. Table 1 below lists the 12 races that Gerber and colleagues called competitive, along with the margin by which those races were decided.

The literature does not provide a single clear definition of a competitive election. Past studies have treated elections decided by 4 percentage points or less as competitive (1) and have referred to races decided by 0.25-, 0.50-, 1.0-, and 2.0-percentage point margins as “near wins” (2). Here, we use a winning margin of 5 percentage points or less as the definition of a competitive primary.

A primary, however, can be competitive without being meaningful. For example, a competitive Democratic primary in a heavily Republican district is neither important nor likely to receive significant attention. It is essentially a contest to decide who will lose the general election. Thus, we consider a primary to have been competitive and meaningful if (1) the primary was won by 5 percentage points or fewer and (2) the winner of the primary went on either to win the general election or to lose it by a small enough margin that this outcome cannot be considered to have been a foregone conclusion. Because the primaries took place roughly three months before the general election, when the general election outcome was less certain, we adopted a looser standard—15 percentage points—to consider a general election loss to have been potentially competitive.

Only 3 of the 12 primaries Gerber and colleagues treated as competitive were actually competitive and meaningful: the Democratic primaries in Michigan’s 8th and 14th congressional districts and the Republican primary in Tennessee’s 4th district. This information is presented in Table 1.

---

1 All 32 congressional districts in the three states included in Gerber and colleagues’ study held both a Republican and a Democratic primary, irrespective of whether there was more than one candidate. In the 1st congressional district in Tennessee, however, nobody ran for the Democratic nomination so there was no Democratic primary.

2 The most competitive of the races included Gerber and colleagues’ study that was decided by a margin larger than 5 percentage points was the Republican U.S. Senate election in Tennessee. In that primary, Senator Lamar Alexander beat a primary challenger by 9.04 percentage points. The race was sufficiently uncompetitive that, 6 days before Election Day, a leading newspaper (Roll Call) published, a “post mortem” analysis of how Senator Alexander had won (3).
Table 1: The 12 House Primary Races Call “Competitive” By Gerber and Colleagues, Along With the Actual Margin of Victory.

<table>
<thead>
<tr>
<th>Primary treated by Gerber and colleagues as “competitive”</th>
<th>Winning margin in the primary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Michigan 1(^{st}) District Democratic primary</td>
<td>Uncontested</td>
</tr>
<tr>
<td>Michigan 1(^{st}) District Republican primary</td>
<td>39 percentage points</td>
</tr>
<tr>
<td>Michigan 2(^{nd}) District Republican primary</td>
<td>14.88 percentage points</td>
</tr>
<tr>
<td>Michigan 4(^{th}) District Republican primary</td>
<td>16.08 percentage points</td>
</tr>
<tr>
<td>Michigan 7(^{th}) District Republican primary</td>
<td>58.60 percentage points</td>
</tr>
<tr>
<td>Michigan 8(^{th}) District Democratic primary*</td>
<td>5.10 percentage points</td>
</tr>
<tr>
<td>(Meaningful; primary winner lost the general election by only 12.53 percentage points)</td>
<td></td>
</tr>
<tr>
<td>Michigan 8(^{th}) District Republican primary</td>
<td>20.52 percentage points</td>
</tr>
<tr>
<td>Michigan 11(^{th}) District Republican primary</td>
<td>32.80 percentage points</td>
</tr>
<tr>
<td>Michigan 13(^{th}) District Democratic primary</td>
<td>47.76 percentage points</td>
</tr>
<tr>
<td>Michigan 14(^{th}) District Democratic primary*</td>
<td>3.23 percentage points</td>
</tr>
<tr>
<td>(Meaningful; primary winner won the general election)</td>
<td></td>
</tr>
<tr>
<td>Tennessee 4(^{th}) District Republican primary*</td>
<td>0.05 percentage points</td>
</tr>
<tr>
<td>(Meaningful; primary winner won the general election)</td>
<td></td>
</tr>
<tr>
<td>Tennessee 9(^{th}) District Democratic primary</td>
<td>33.64 percentage points</td>
</tr>
</tbody>
</table>

* Competitive primary: decided by 5 percentage points (rounded to the whole number) or fewer.
A fourth race not included in Gerber and colleagues’ list of those they treated as competitive also meets the five-percentage-point standard for competitiveness: the Republican primary in Tennessee’s 3rd district. That primary was decided by a 0.82-point margin and was meaningful because the nominee went on to win the general election.

Of course, even a competitive congressional primary is may be seen by most people as considerably less important and thus as offering lower identity stakes to a prospective voter than a competitive gubernatorial or presidential general election.

**Our Follow-up study method and results**

**Method**

We recruited 587 Amazon Mechanical Turk workers and randomly assigned them to read one of two election scenarios. In both conditions, participants received the following instruction: “Please imagine the following scenario and answer the questions that follow as though the scenario were real.”

Participants assigned to the Bryan et al. condition were asked to imagine a state-wide election for governor like the 2009 New Jersey gubernatorial election used in Experiment 3 of our paper (4):

> It’s November and your state is holding a **general election**. The election will be held this coming Tuesday. It will determine **who becomes the next governor of your state**. The race is tight and it is unclear who will win. Because of this, and because of the importance of the office at stake, the election has received extensive coverage in local, state-wide, and national news.

Participants assigned to the Gerber et al. condition were asked to imagine a primary like most of those studied by Gerber and colleagues (5):

> It’s early August and your state is holding a **primary election**. The primary will be held this coming Tuesday. It will select a **candidate from each party to run in November’s general election for your district’s representative in the U. S. House of Representatives** (that is, congressman or congresswoman). For both parties, the leading candidate is either unopposed (and, so, guaranteed to win) or leading by such a large margin that he is almost certain to win. Because the outcome is not in doubt, the local news has not covered the primary much.

After reading the scenario they were assigned to, all participants answered the following four questions:
1. “Keeping in mind all of the responsibilities you have on a typical Tuesday, how proud does it make you feel to imagine being a voter in this election?” (1 = Not at all proud, 2 = Slightly proud, 3 = Somewhat proud, 4 = Very proud, 5 = Extremely proud)

2. “Keeping in mind all of the responsibilities you have on a typical Tuesday, how worthwhile do you think it would be to make the time to be a voter in this election?” (1 = Not at all worthwhile, 2 = Slightly worthwhile, 3 = Somewhat worthwhile, 4 = Very worthwhile, 5 = Extremely worthwhile)

3. “Keeping in mind all of the responsibilities you have on a typical Tuesday, how important do you think it would be to make the time to be a voter in this election?” (1 = Not at all important, 2 = Slightly important, 3 = Somewhat important, 4 = Very important, 5 = Extremely important)

4. Keeping in mind all of the responsibilities most people have on a typical Tuesday, including work and family obligations, how true would you say it is that a good citizen has a responsibility to be a voter in this election? (1 = Not at all true, 2 = Slightly true, 3 = Somewhat true, 4 = Very true, 5 = Extremely true)

Next, participants were asked to indicate whether or not they were registered to vote, which political party they preferred, and their political orientation on a 7-point liberal to conservative spectrum. Finally, they were asked four quiz-style multiple choice questions to gauge their attention to the scenario they had read: (1) “Did you read a description of a primary election or a general election (Answer choices: Primary, General, Other); (2) “Was this a Presidential election? (Answer choices: Yes, No); (3) Was the election you read about close or was the outcome basically already known? (Answer choices: It was close, The outcome was basically already known); (4) “Was the election you read a description of receiving coverage or attention from state-wide or national news outlets? (Answer choices: Yes, No).

Results

Results are reported for the 381 participants who correctly answered all four attention check questions described above, however none of the results change if all 587 MTurk workers recruited are retained.

Compared to participants in the Gerber condition, those in the Bryan condition reported that they would feel significantly prouder to be a voter, that being a voter would feel more important and more worthwhile, and that a good citizen had more of a responsibility to be a voter in the election scenario they read. An “identity stakes” composite of all four items (α=0.92) revealed that participants judged the identity stakes of the high-profile general election to be far greater than those of the low-profile congressional primary. See Table 2. The effect was not moderated by registration status, party preference, or ideological orientation (all interaction ps > 0.19).
Table 2: Means, Standard deviations, t-test Results, and Effect Sizes for All Dependent Variables in Our Empirical test of the identity stakes of the type of election studied in the Bryan et al. and Gerber et al. experiments.

<table>
<thead>
<tr>
<th>Item</th>
<th>Gerber et al. Condition: mean (SD)</th>
<th>Bryan et al. Condition: mean (SD)</th>
<th>t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proud</td>
<td>2.81 (1.28)</td>
<td>3.48 (1.20)</td>
<td>$t = 5.15, p &lt; 0.0005, d = 0.54$</td>
</tr>
<tr>
<td>Worthwhile</td>
<td>2.71 (1.28)</td>
<td>3.92 (1.12)</td>
<td>$t = 9.71, p &lt; 0.0005, d = 1.01$</td>
</tr>
<tr>
<td>Important</td>
<td>2.78 (1.32)</td>
<td>3.99 (1.08)</td>
<td>$t = 9.61, p &lt; 0.0005, d = 1.01$</td>
</tr>
<tr>
<td>Good citizen responsibility</td>
<td>3.33 (1.27)</td>
<td>3.90 (1.07)</td>
<td>$t = 4.64, p &lt; 0.0005, d = 0.49$</td>
</tr>
<tr>
<td>Identity stakes composite</td>
<td>2.91 (1.15)</td>
<td>3.82 (0.99)</td>
<td>$t = 8.19, p &lt; 0.0005, d = 0.85$</td>
</tr>
</tbody>
</table>

In sum, participants saw the gubernatorial general election scenario, which represented the context of the 2009 NJ gubernatorial election used in our paper, as having far greater identity stakes than the congressional primary election scenario, which represented the context in which almost all of Gerber and colleagues’ participants were tested.

It is notable that these differences emerged in response to a brief hypothetical scenario. The real life experience of a hotly contested election receiving heavy coverage, where analysts and pollsters are continually reporting on and speculating about what “voters” think and what “voters” will decide, may only strengthen the effect. This is especially the case when compared with the real life experience of a low-salience, low-importance primary one is only vaguely aware of amid all the other things going on in one’s life.

Conclusion

The most important problem with Gerber and colleagues’ replication attempt is that it was not conducted in a context in which the original theory would have predicted an effect. This lesson—that psychological context plays an essential role in shaping how an intervention is construed by participants and therefore how they respond to it—is important not only for this research but for any attempt to replicate social psychological findings (6).

In the present case, however, there are also other potentially important differences between the studies, ones that could change participants’ experiences, altering key
psychological processes and their downstream behavioral consequences. For example, most participants in the Gerber study were not contacted the day before the election as our participants were. Instead, they were contacted in the 4 days leading up to the election. We deliberately chose to contact participants as close in time as we could to the opening of the polls because we assumed that any reframing of the way participants construed the act of voting could otherwise be “drowned out” by subsequent experiences. In addition, Gerber and colleagues’ test was conducted by telephone interviewers rather than via a web-based survey, most likely taken in private, as ours was. We do not know how this more overtly social version of the treatment might alter participants’ experience, but it is something that needs to be considered. Both the timing and method of the survey administration are variations on our method that could be tested systematically to further illuminate our effect. Such tests would be of great interest to us and we stand ready to collaborate with researchers who seek a greater understanding of psychological phenomena.
References