



ARTICLE

# Pretending to Support? Duterte's Popularity and Democratic Backsliding in the Philippines

Yuko Kasuya<sup>1</sup>  and Hirofumi Miwa<sup>2</sup> 

<sup>1</sup>Faculty of Law, Keio University, Tokyo, Japan and <sup>2</sup>Faculty of Law, Gakushuin University, Tokyo, Japan  
**Corresponding author.** Yuko Kasuya Email: [ykasuya@keio.jp](mailto:ykasuya@keio.jp)

(Received 19 October 2022; revised 21 June 2023; accepted 26 June 2023)

## Abstract

The incumbent-led subversion of democracy represents the most prevalent form of democratic backsliding in recent decades. A central puzzle in this mode of backsliding is why these incumbents enjoy popular support despite their actions against democracy. We address this puzzle using the case of Philippine President Rodrigo Duterte. Although some Philippine analysts have speculated that his popularity was inflated due to social desirability bias (SDB) among survey respondents, there has been limited empirical examination. Our pre-registered list experiment surveys conducted in February/March 2021 detected SDB-induced overreporting at about 39 percentage points in face-to-face surveys and 28 percentage points in online surveys. We also found that the poor Mindanaoans, and those who believed their neighbors supported Duterte, were more likely to respond according to SDB. These possibly counter-intuitive results should be interpreted with caution because the survey was conducted during the height of the COVID-19 lockdown, and the findings cannot necessarily be extrapolated to the other period of his presidency. Nevertheless, this study suggests that preference falsification could be an alternative explanation for the puzzle of popular incumbents in democratic backsliding.

**Keywords:** democratic backsliding; social desirability bias; list experiment; Rodrigo Duterte; Philippines

## Introduction

Democratic backsliding<sup>1</sup> has attracted intensive scholarly attention in the past several decades.<sup>2</sup> We now know that the subversion of democracy by chief executives represents the predominant mode of backsliding (Bermeo 2016; Svobik 2020). A central puzzle about this phenomenon is that these executives obtain massive support. Accordingly, we must ask, in the words of Milan Svobik (2020, 5), “Why do voters who routinely profess pro-democratic values simultaneously support incumbents intent on subverting democracy?” Donald Trump of the United States, Viktor Orbán of Hungary, Narendra Modi of India, Jair Bolsonaro of Brazil, and Recep Erdoğan of Turkey are some manifestations of this puzzle.

Philippine's President Rodrigo Duterte joins the list of incumbents who assaulted democracy yet remain popular. Upon being elected in 2016, Duterte mounted attacks

© The Author(s), 2023. Published by Cambridge University Press on behalf of the East Asia Institute

on democratic norms and institutions and killed thousands of suspected drug users without due process (Rafael 2021). Despite local and international criticism, he remained very popular among Filipino voters, maintaining between about 60 percent and 90 percent approval according to most surveys. Scholars have provided a handful of ways to make sense of this, attributing his popularity to his personality (Arguelles 2019; Jennings 2019; Kenny and Holmes 2020), his moral appeal (Kusaka 2017a; Warburg and Jensen 2020a), his populist style (Curato 2016; Dulay, Hicken, and Holmes 2022), his policies (Kenny and Holmes 2020), and historical conjunctures (Arugay and Slater 2019; Garrido 2022; Teehankee 2017; Thompson 2018). These analyses provide significant insights into the ascent of Duterte and the broader dynamics of Philippine politics.

As a complementary contribution to the existing scholarship, we offer an alternative perspective that delves into Duterte's popularity by examining the phenomenon of preference falsification—specifically, the act of adjusting one's behavior to appear socially acceptable. We have focused on social desirability bias (SDB), a particular form of preference falsification. Nearly conceptually synonymous with preference falsification, SDB is a term commonly used in the survey research literature. It refers to the behavior of giving an untruthful but socially acceptable or desirable answer. Although the suspicion that Duterte's popularity rating may be inflated due to SDB has been discussed by some analysts (David 2020; Elemia 2020; Santos 2020; Teodoro 2020), we are unaware of any study that has empirically tested it.

To fill this gap, we employed the list experiment method, a survey method devised to explore socially sensitive questions. We conducted face-to-face and online surveys simultaneously from late February to early March 2021, which was a period of national lockdown and high unemployment rates induced by the COVID-19 pandemic. Our main finding was that Duterte's strikingly high approval rating was significantly inflated around that time, with more than one in three respondents who answered they support Duterte in both survey modes pretending to support the president. Specifically, we found that the magnitude of SDB was 39.5 percentage points for face-to-face survey and 28.3 percentage points for online survey. Further, three types of respondents were more likely to pretend to support the president: the socio-economically vulnerable, Mindanaoans, and those who perceived or believed that the president was popular among their neighbors.

Our results may appear counter-intuitive to some, due to the high magnitude of SDB we observed. It is important to underscore our commitment to transparency through the pre-registration of our research design and including a battery of sensitivity analyses, as detailed in the Online Appendix. Furthermore, it is crucial to note that our investigation pertains to a specific timeframe (February–March 2021), and therefore caution should be exercised when generalizing our findings to other periods of Duterte's presidency. Keeping these considerations in mind, we maintain that our results are reliable to the best of our knowledge.

This study contributes to multiple strands of scholarship. First, regarding the comparative study of democratic backsliding, our study suggests an alternative explanation for the puzzle of popular backsliding incumbents. Recent studies have argued that political polarization (Graham and Svulik 2020; Svulik 2020) and a majoritarian preference among voters (Grossman et al. 2021) are the major ways to understand

this puzzle. While further investigation is required to examine the applicability of theories within the Philippine context, our study introduces an additional factor for consideration: the presence of voters who engage in preference falsification regarding their opinions on regime support, who we call “pretenders.” Traditionally, preference falsification was a major concern in repressive authoritarian regimes (e.g., Kuran 1995; Wedeen 1999). Our study demonstrates that it is also a concern in the context of democratic backsliding. The existence of pretenders is non-trivial because a backsliding incumbent’s source of legitimacy is popular support. Accordingly, an inflated approval undermines public opinion to function as a check on the executive. This can send the incumbent an unwarranted signal that they can continue attacking democracy. As such, pretenders play a consequential role in backsliding democracies. Furthermore, our study expands the political context suitable for studying sensitivity bias to countries experiencing democratic backsliding, in addition to the authoritarian contexts suggested by Blair, Coppock and Moor (2020).

Second, this study provides insights into the democratic backsliding of the Philippines during the Duterte era. Our findings empirically confirm the suspicion held by some observers that his popularity was inflated, especially toward the latter part of his presidency (David 2020; Elemia 2020; Santos 2020; Teodoro 2020). Adding the element of preference falsification tempers our understanding of the Duterte administration. We discuss this point in this article’s conclusion.

Third and related to the second point, this study engages with the studies investigating Duterte’s popularity. As discussed earlier, numerous arguments have been put forth to comprehend why he is so popular, with a notable commonality that most such studies accepted his popularity as a given, or unquestioned the reliability of opinion poll results. Conversely, this study challenges the premise that he is genuinely popular. Our findings encourage scholars critically assess such underlying assumptions in analyzing the executive approval in the case of Duterte and possibly beyond.

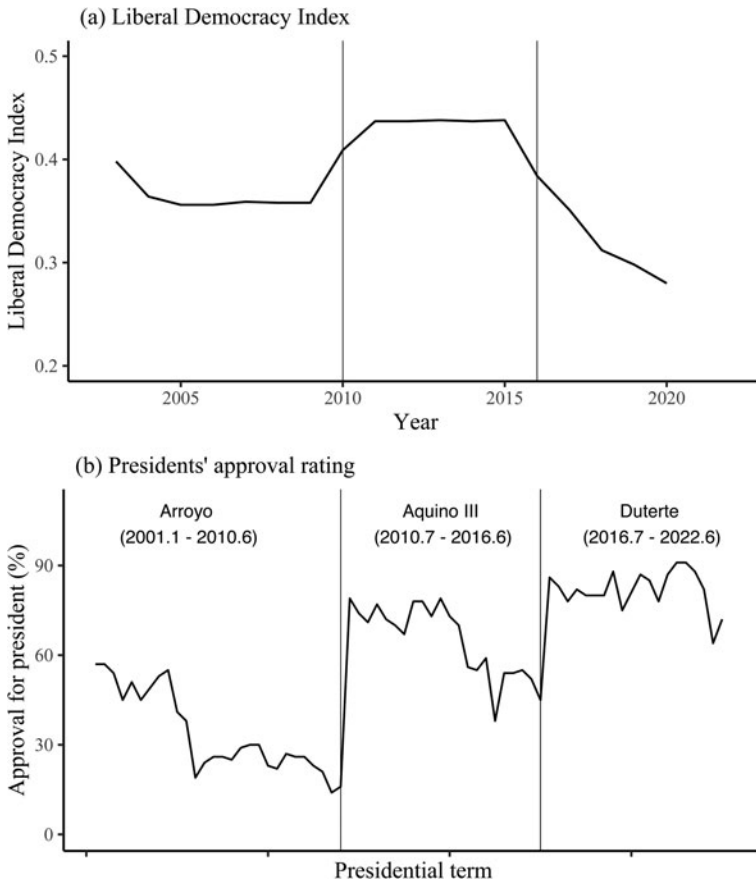
Fourth, this study adds to the nascent but rapidly expanding body of studies that utilize survey methods to study politics in developing countries (Lupu and Michelitch 2018). Our article joins the studies that use list experiments to understand the existence of SDB in opinion surveys about political leaders in developing countries. To the best of our knowledge, our paper is the first to implement a list experiment among Filipino respondents to study socio-politically motivated SDB.

## The puzzle of popular backsliding incumbents

### *Duterte’s popularity as an unresolved puzzle*

A puzzling aspect of recent incumbent-led backsliding is that these executives obtain significant support despite attacking democratic norms and institutions. Svobik (2020, 5) notes that it is often confusing that voters would elect such leaders and, “even more perplexingly, many undemocratic incumbents, including the examples of Chavez, Putin, and Erdoğan, enjoy significant and genuine popular support.” Luo and Przeworski (2019, 3) and Przeworski (2019, 187) raise similar concerns.

President Duterte is one case of a popular backsliding incumbent. Figure 1 compares the Liberal Democracy Index (LDI) of the V-Dem Project—a composite indicator of how a country achieves ideals of liberal democracy (Coppedge et al. 2021)



**Figure 1.** Duterte's Popularity amid Democratic Backsliding.  
 Source: 1(a) from Coppedge et al. 2021; 1(b) from Pulse Asia.

—and the survey-based presidential approval rating provided by the local polling company Pulse Asia. As Figure 1(a) shows, on the one hand, democracy backslid under Duterte's presidency. Upon being elected in 2016, he attacked democratic institutions, including the legislature, courts, civil society groups, and the media. In the context of his "war on illegal drugs" campaign, thousands of people accused of selling drugs were killed without due legal process (Johnson and Fernquest 2018). On the other hand, Figure 1(b) shows that he remained very popular. Duterte's approval rating remained between 60 percent and 90 percent throughout his term. Moreover, unlike his predecessors Arroyo and Aquino III, whose popularity sharply declined during their final years, he maintained a high approval rating until the end of his term.

Duterte's high approval rating has been a subject of inquiry since the beginning of his term. Some analysts attributed his popularity to personality traits: he is charismatic (Kenny and Holmes 2020), easily identifies with the masses (Arguelles 2019;

Jennings 2019), and resonates with the morality of the poor (Kusaka 2017a, 2017b; Warburg and Jensen 2020a). Others emphasized the “us versus them” rhetoric of Duterte’s populist leadership style, characterizing his version as “penal populism” (Curato 2016; Kenny and Holmes 2020) and “ethno-populism” (Dulay, Hicken, and Holmes 2022). From a policy perspective, his signature policy, the war on illegal drugs, appealed to Filipinos who were tired of chronic drug-related problems (Kenny and Holmes 2020). Others placed his emergence as a political leader in history terms, claiming that voters were disillusioned with the liberal democracy initiated by his predecessors and considered Duterte’s emphasis on disciplined rule an attractive alternative (Curato 2017; Garrido 2022; Teehankee 2017; Thompson 2018) and part of cyclical careening between populism and elite democracy (Arugay and Slater 2019).

Duterte’s popularity was further scrutinized when Pulse Asia released the results of its regular executive approval survey in October 2020. The survey was conducted in September 2020, when the country was in the midst of a sharp surge in COVID-19 patients, very strict lockdown measures, and high unemployment rates. Despite these factors, Duterte obtained his highest recorded approval rating, 91 per cent (Calonzo 2020).<sup>3</sup>

That result prompted some to offer another interpretation of his popularity: Survey respondents were pretending to support Duterte out of social concern. For example, sociologist Randy David (2020) wrote in his newspaper column that most Filipinos “would readily give ‘safe’ answers [rather] than say something that could expose them to unwanted drug raids or to being denied ‘ayuda’ (government hand-outs).” Political analyst Richard Heydarian said at a press conference, “What’s the incentive for people to be completely honest about whether they like or dislike a person? The climate of fear cannot be taken out as a factor” (Elemia 2020). Veteran journalists Luis Teodoro (Teodoro 2020) and Vergel Santos (Santos 2020) wrote similar analyses.

Preference falsification is the social science term for the point made by David and Heydarian. This concept refers to the act of *not* publicly revealing one’s private (truthful) preferences due to feelings of fear or shame. Timur Kuran, who pioneered the study of preference falsification in political contexts, notes that this is a specific form of lying aimed at manipulating others’ perceptions of the speaker’s motivations or dispositions (Kuran 1995, 4–5).<sup>4</sup>

SDB is another term pertaining to our research context. While preference falsification can happen in any human interaction, SDB is usually restricted to the answers respondents give in surveys. This study uses the two terms as conceptual equivalents but distinguishes them according to context.<sup>5</sup>

Against these backdrops, we address political preference falsification among Filipino survey respondents. Was Duterte’s approval rating inflated due to SDB? If so, to what extent and under what conditions were respondents more vulnerable to such bias?

## Hypotheses

Our central hypothesis is that Duterte’s approval rating was significantly inflated by over-reporting due to SDB. The hypothesis empirically examines the concern

addressed by some Philippine analysts (David 2020; Elemia 2020; Santos 2020; Teodoro 2020).

*H1 Duterte's approval rating was inflated due to the presence of SDB in some respondents.*

Our second hypothesis examines whether the survey mode makes a difference in inducing SDB. This hypothesis also examines some suggestions of Filipino analysts. For example, after explaining how sensitive Filipino social interactions can be, Molo (2020) posed the following hypothetical:

[C]an you imagine a situation where a stranger arrives in your house and asks you, "How much do you approve of Duterte?"

Reflecting on this implies that the face-to-face survey method employed by Pulse Asia can intimidate some respondents. Furthermore, the survey method literature shows that when sensitive questions are asked, face-to-face respondents are more susceptible to SDB (due to the presence of an interviewer) than those who answer online surveys (Duffy et al. 2005; Heerwegh 2009; Richman et al. 1999). We expect that a similar mode effect exists in the Philippine context.

*H2: The face-to-face interview method produces a higher degree of SDB than the online survey method.*

The rest of our hypotheses examine conditions that might affect susceptibility to SDB. We focus on two types of conditions: "the climate of fear" as described by Philippine analysts, and social conformity pressure. Journalist Louis Teodoro explained the first condition: "The transformation of the country's political context from one that, prior to 2016, was relatively threat-free to the climate of fear that now defines it is most certainly a factor as well in the way the populace responds to surveys and public opinion polls" (Teodoro 2020). We examine the fear factor in Hypotheses 3, 4, and 5.

Hypothesis 3 operationalizes the fear argument by mentioning "extrajudicial killing" (EJK), that is, the killing of suspected illegal drug dealers and addicts without due legal process.<sup>6</sup> Although the estimated number of EJK victims under the Duterte administration varies,<sup>7</sup> EJKs began frequently making headlines soon after his inauguration. In a survey conducted by the local polling company Social Weather Stations in mid-2017, about 70 percent of respondents said they were concerned that they or somebody they knew would be a victim of EJK (ABS-CBN 2019). For these reasons, we expect respondents who personally knew EJK victims to feel more fear about giving unsafe answers and propose the following:

*H3: Respondents who personally know an EJK victim are more likely to falsify their preferences and express approval of Duterte.*

Our second operationalization of the fear thesis focuses on “red-tagging”: being labeled a communist or a member of a left-leaning political group. As one Philippines-based journalist noted, “[R]ed-tagging has become a dangerous tool amid the widespread culture of impunity under Duterte. While most, if not all, of these accusations are baseless, those targeted fear for their lives” (Gavilan 2021). Expecting those who personally knew victims of red-tagging to be more fearful and more likely to be influenced by SDB, we proposed the following.

*H4: Respondents who personally know a victim of “red-tagging” are more likely to falsify their preferences and express approval of Duterte.*

Analysts also mention “the regime’s keyboard army of trolls” as a reason for the fear (Teodoro 2020). Online political trolls intentionally leave offensive messages online. Since Duterte’s term, social media sites such as Facebook have been filled with repugnant messages against those who criticize Duterte (Cabanes and Cornelio 2017). Frequent internet users see more of such content, making them more likely to develop fear. This led to our third operationalization of the fear factor:

*H5: Respondents who spend more time using online social media (such as Facebook and Twitter<sup>8</sup>) are more likely to falsify their preferences and express approval of Duterte.*

Our sixth hypothesis examines the effects of what the social psychology literature calls “social conformity pressure” (e.g., Asch 1951).<sup>9</sup> This notion refers to the act of changing one’s originally correct answer to something incorrect but endorsed by the majority of people they encounter. Although social conformity pressure is similar to the notion of SDB, it differs in that it occurs in settings where people face direct pressure from the group they belong to. The source of social conformity pressure can be a vague notion of the people or the society. Using respondents’ perceptions of Duterte’s popularity in their neighborhood as a proxy for social conformity pressure, we hypothesized the following:

*H6: Respondents who believe President Duterte is highly popular in their neighborhood are more likely to falsify their preference and express approval of Duterte.*

Our hypothesis concerning social conformity pressure resembles what Canare et al. (2021) labeled “herd behavior.” Economists and sociologists use this term to refer to the phenomenon of people following a crowd. Canare et al.’s (2021) survey research, conducted among Metro Manila residents, found that respondents who believed the president was popular among people they knew were more likely to support him than those who did not. Although H6 is identical to their herd behavior hypothesis, we did not follow their theoretical framing because herd behavior may or may not involve preference falsification. Because our study focuses explicitly on preference falsification, we have used the more precise term social conformity pressure.

## Data and methods

### List experiment surveys

We conducted pre-registered list experiments to measure the proportion of Duterte's approval rating unaffected by SDB.<sup>10</sup> We commissioned Pulse Asia to include the list experiment and related questions in its periodic nationwide face-to-face survey, fielded from February 22 to March 3, 2021. The population was Filipino adults. Multi-stage random sampling with uniform quotas for gender  $\times$  region (four strata: National Capital Region [NCR], Luzon excluding NCR, Visayas, and Mindanao) was applied. There were 1,200 respondents. We simultaneously conducted an online survey with the same experiment and questions as the face-to-face survey. Respondents were recruited from Lucid Marketplace from February 26 to March 3. Given difficulties reaching older respondents in the Filipino online setting, we set the target of our online survey as citizens who were aged 44 or younger. Thus, we only use the data from respondents to the face-to-face survey who are in this age bracket when we compare the results of the two surveys. We set quotas for gender, age (three strata: 18–24, 25–34, and 35–44), and region to ensure that the sample's marginal distribution was as close as possible to that of the national population. We concurrently conducted an additional online survey containing a placebo list experiment and performed a novel sensitivity analysis to address the problem of non-strategic misreporting often associated with the list experiment method (Kuhn and Vivyan 2022; Rimbau and Ostwald 2021). We describe the design of the placebo experiment and provide the results of the sensitivity analysis in Section 'Sensitivity analysis for non-strategic misreporting'.

We randomly assigned respondents to treatment and control groups in all surveys. We conducted blocked randomization based on gender and region in the face-to-face survey and based on gender, age, education (completed high school or below, vocational or some college, and completed college or higher), and region in the online survey.<sup>11</sup> Our experiment adopted the following core question:

Of the politicians on this list, **how many** do you approve of in terms of performance of duties?

- MANUEL QUEZON
- FERDINAND MARCOS
- CORAZON "CORY" AQUINO
- (RODRIGO "DIGONG" DUTERTE)

The item "RODRIGO 'DIGONG' DUTERTE" was shown only to the treatment group. The order of the items was randomized in the online survey. Respondents chose one of the numbers from 0 to 3 (for the control group) or 0 to 4 (for the treatment group). Following Frye et al.'s (2017) design, we put Duterte alongside previous presidents of the Philippines.<sup>12</sup>

To test Hypothesis 1, we included a direct question about respondents' approval of Duterte. Because the presidential approval question in Pulse Asia's regular survey employs a five-point scale and is not comparable with our list question, we added an original question to the surveys: "Do you approve or disapprove of President



Duterte's performance of his duties?" The options were "Approve," "Disapprove," and "Don't know" (DK). Following the advice of Blair and Imai (2012), the direct question was asked after the list question. We considered the "direct questioning (DQ)-based" approval rating as the percentage of "Approve," with DK treated as a missing value.<sup>13</sup> Additionally, because the face-to-face survey contained Pulse Asia's question about support for Duterte based on a five-point scale, we added this question to the online survey for reference purposes.<sup>14</sup>

After the list question, we asked the following four questions to test Hypotheses 3–6. We first asked respondents whether they personally knew a victim of an EJK. Second, we asked whether they personally knew anyone labeled a communist or a member of a communist group (i.e., red-tagged). The options for these two questions were "Yes" and "None." Third, we asked respondents how many minutes they spent using their social networking services, such as Facebook and Twitter, on a typical weekday. Options were 0–60, 61–90, 91–120, and over 121 (minutes). We re-categorized respondents using a threshold of 60. The fourth question asked each respondent whether they perceive people in their *barangay* (neighborhood or the smallest administrative unit) to be satisfied with Duterte's performance of his duties. This question was measured using a four-point Likert scale with a "don't know" option. We divided respondents into two groups, excluding those who answered "don't know."

### Inferences

We obtain an unbiased estimate of the percentage of individuals who would select a sensitive item (in our case, "RODRIGO 'DIGONG' DUTERTE") if it were displayed in the list by computing the difference in outcome means between the treatment and control groups. We conducted linear regressions in which the dependent variable is each respondent's answer to the list experiment question, and the independent variable is an indicator for being assigned to the treatment group. The coefficient represents the estimate of Duterte's approval rate not affected by SDB.<sup>15</sup> Because we conducted blocked randomization in the online surveys, the intercept varied by block when we analyzed the online survey data (Imbens and Rubin 2015, chap. 9). We estimated the parameters of this model by ordinary least squares with HC2 robust standard errors.

To test Hypothesis 1, we estimated the magnitude of SDB in both the face-to-face and online surveys. Because we measured Duterte's approval ratings from estimates based on DQ and the list experiment, subtracting the latter value from the former yields an unbiased estimate of the magnitude of SDB. Following standard practice (e.g., Blair, Coppock, and Moor 2020), we estimated the variance of this estimate by summing the variances of the DQ- and list-experiment-based estimates. Additionally, we tested the null hypothesis using the *z*-test.

Hypotheses 2–6 concern the difference in the magnitudes of SDB between different types of surveys or survey subgroups. We obtained an unbiased estimate of this quantity by subtracting the point estimate of the magnitude of SDB in one group from that in the other group (i.e., taking the difference-in-differences in means). Assuming the independence of the estimated magnitude of SDB across surveys and

subgroups, the variances of these estimates were computed by summing the variances of the magnitudes of SDB in each group; the statistical tests were based on the standard normal distribution. Although we set a statistical significance threshold of 0.05, we also report exact *p*-values.

**Results**

**Social desirability bias and survey mode effects**

Figure 2 demonstrates the estimation results of Duterte’s approval ratings using both direct and list-type questions, indicating the differences observed between survey modes. The figure also displays the difference in SDB magnitudes between face-to-face and online surveys (in the bottom line). Cross marks and dots represent point estimates. Horizontal segments represent the 95 percent confidence intervals (CIs).

When we asked respondents directly whether they approved of Duterte’s performance of his duties, his approval rate was 0.897 [0.880, 0.915] in the face-to-face survey and 0.808 [0.795, 0.821] in the online survey (95 percent CIs appear in brackets; this applies consistently hereafter). These high approval ratings are consistent with previously reported surveys and also coincide with the aggregate results of the Pulse Asia-style direct question in our surveys.<sup>16</sup> In contrast, list-experiment-based estimates of Duterte’s approval rate were 0.502 [0.411, 0.593] for the face-to-face survey and 0.525 [0.464, 0.586] for the online survey. The differences in the DQ-based and list-experiment-based estimates were 0.391 [0.291, 0.491] for the face-to-face survey and 0.283 [0.220, 0.345] for the online survey. These values are significantly distinguishable from zero (*p*-value was almost zero for both survey modes). The difference can be interpreted as being due to SDB’s existence, providing strong support for Hypothesis 1.

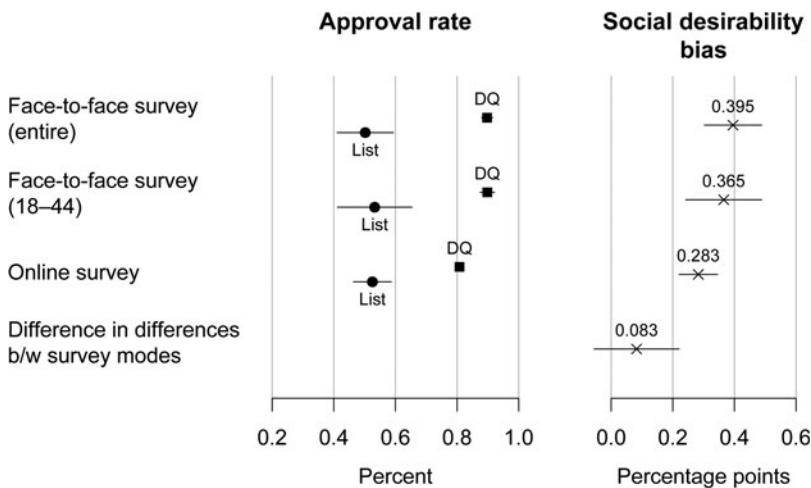


Figure 2. Estimates of Duterte’s approval ratings, SDB, and differences in SDB between survey modes. Note: Cross marks and dots represent point estimates. Horizontal segments represent the 95% CIs.

To test Hypothesis 2, we compared SDB magnitude between survey modes. We restricted the sample of the face-to-face survey to those who were 44 years old or younger to reconcile the distribution of respondent age across surveys. Figure 2 shows that this restriction did not substantially change the estimates of approval ratings or SDB in the face-to-face survey, with the bottom row showing the estimated difference in the magnitude of SDB to be 0.078 [-0.070, 0.226]. As predicted, SDB was greater in the online survey than the face-to-face survey. However, because this difference is not statistically significant ( $p = 0.301$ ), we cannot confirm Hypothesis 2.

We note that the difference in SDB magnitude between survey modes can be attributed to the difference in DQ-based approval ratings and that the list-experiment-based approval rates for Duterte are stable across surveys. These results indicate that it is likely that our list experiment detected the true level of Duterte's popularity.

### *Heterogeneity in susceptibility to social desirability bias*

Figure 3 shows the results of testing Hypotheses 3–6. Cross marks indicate the point estimate of SDB magnitude for each subgroup based on the comparison between the direct questions and the list experiment questions (note that the horizontal axis of Figure 3 differs from that of Figure 2). Dots indicate the point estimate of the difference in the magnitude of SDB between subgroups. Labels show the number of respondents in each subgroup.

In neither survey was personally knowing EJK or red-tagging victims significantly related to SDB magnitude. Because the number of such respondents was very small in the face-to-face survey, we must say that the tests for the face-to-face survey were underpowered. Nonetheless, we can confirm that knowing EJK or red-tagging victims was hardly related to SDB magnitude in the online survey.

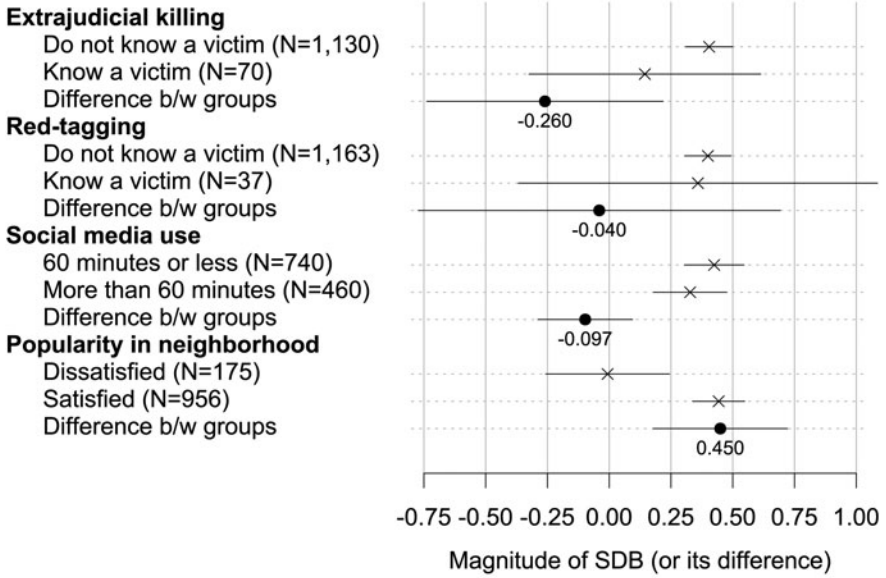
The results of both surveys demonstrate that the size of SDB did not significantly vary according to the time respondents spent on social media. Therefore, we can conclude that Hypothesis 5 does not hold.

In contrast, we found strong support for Hypothesis 6. The results indicate that individuals who believed that people in their *barangay* were dissatisfied with Duterte's performance would not alter their responses based on whether the question format was direct or list; they were completely unaffected by social desirability. Participants who perceived their neighbors as supporters of Duterte's performance were more likely to report that they approved of Duterte when directly asked than when the list question format was used. Such tendencies of social conformity pressure were observed in both surveys. The estimated difference in SDB size was 0.475 [0.193, 0.756] in the face-to-face survey and 0.397 [0.246, 0.549] in the online survey, a difference significantly distinguishable from zero ( $p < 0.001$  in both surveys).

### *Exploratory subgroup analyses*

In addition to testing SDB heterogeneity, we investigated how the degree of SDB differed according to certain demographic variables. We focused on gender, educational attainment, region of residence, and socioeconomic class.<sup>17</sup> Note that, as Figure 2

### Face-to-face survey



### Online survey

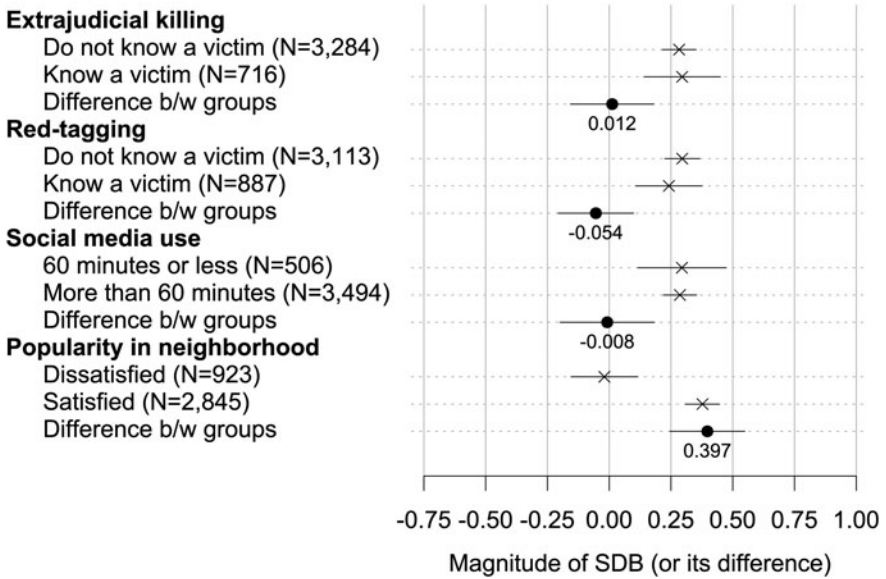


Figure 3. Subgroup comparison of the magnitudes of SDB for Hypotheses 3–6.  
 Note: Dots represent point estimates. Horizontal segments represent the 95% CIs.

shows, we mentioned that the difference in SDB caused by respondent age was negligible, with the SDB in the face-to-face sample of respondents aged 18–44 close to the entire sample's SDB.

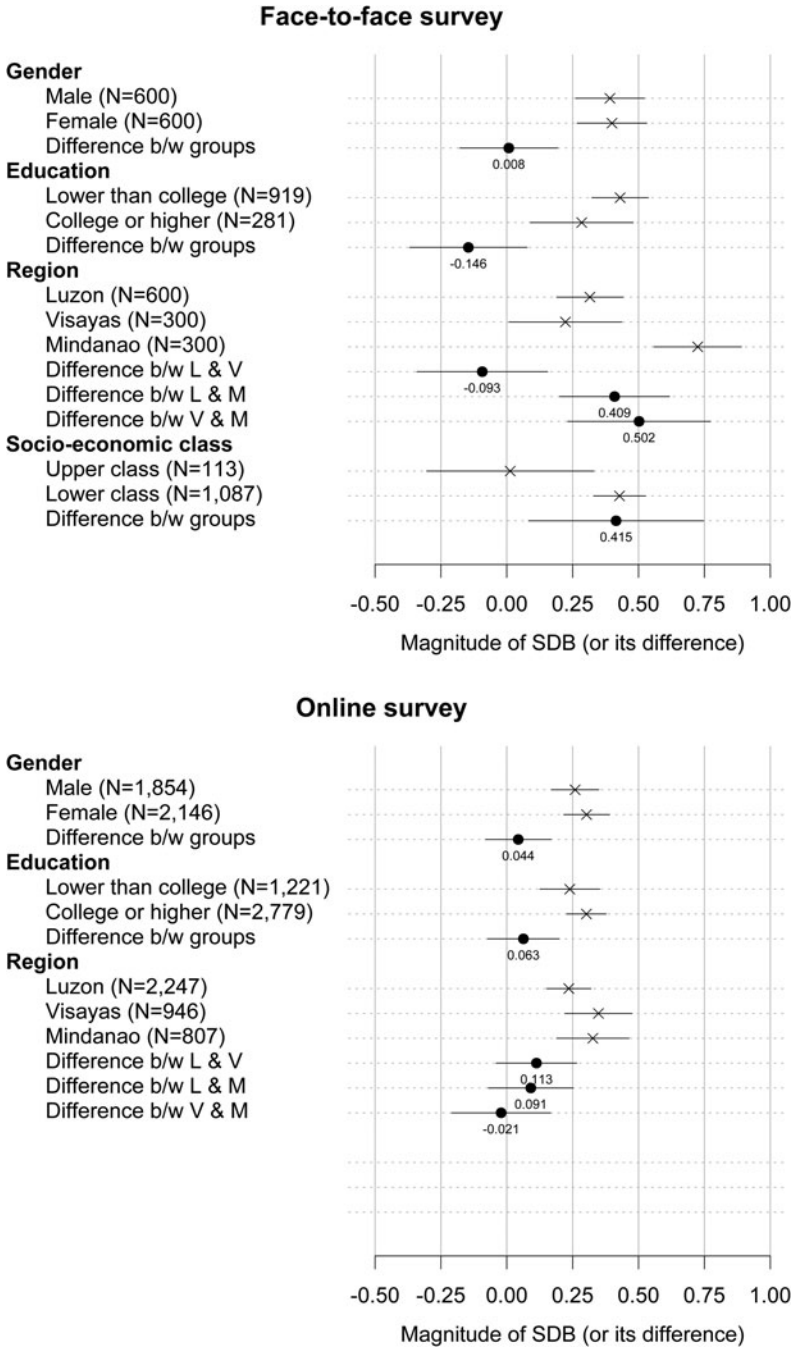
We split respondents into two groups based on education level: those with some college education or higher and those with less than some college education. Respondents' socioeconomic class was measured by the question on the type of housing, which is regularly used in Pulse Asia and other face-to-face polls in the Philippines. Following convention, we treated the respondents classified A, B, or C as the upper class and those in the D or E categories as the lower class. This information was only available for the face-to-face survey results.

We report the results in [Figure 4](#). As with [Figure 3](#), cross marks show the SDB point estimate, and dots show the point estimate of the difference in SDB between subgroups. The horizontal segments represent 95 percent CIs. Neither gender nor education level produced a significant SDB difference in either the face-to-face or the online survey.

However, the face-to-face survey results suggest a large difference across three major regions. While the degree of SDB was modest among respondents in Luzon and Visayas (about 27 and 24 percentage points, respectively), the magnitude reached 72.5 percentage points for Mindanaoans. As we report in the Online Appendix, the DQ-based approval estimate was greater in Mindanao (0.952 [0.928, 0.977]) than in other regions (0.883 [0.856, 0.909] in Luzon and 0.869 [0.829, 0.909] in Visayas); this is consistent with the conventional narrative that Mindanao represents Duterte's supporter base. However, the list-based estimate, which is assumed to reflect truthful answers, was remarkably low in Mindanao (0.228 [0.063, 0.392]) compared with other regions (0.567 [0.444, 0.691] in Luzon and 0.646 [0.436, 0.856] in Visayas).<sup>18</sup>

In the online mode, we did not find regional differences that resembled our face-to-face survey. This absence of regional differences in the online survey is due to the list-based estimate for Mindanao being at the same level as other regions (0.612 [0.475, 0.748]; see Online Appendix C), unlike in the face-to-face survey. The difference in Mindanao's SDB between surveys may be due to socioeconomic differences between online and face-to-face respondents. As the tables provided in Online Appendix C show, in Mindanao, the upper-class segment is smaller, and internet use is less extensive than in other regions. Due to these regional traits and the upper class being heavier internet users, it is likely that Mindanao's online survey respondents included a larger proportion of affluent Filipinos than other regions. As detailed in this paper's discussion section, upper-class respondents demonstrated low SDB. The approval rating among upper-class respondents remained very high (about 80 percent) when measured by the list question. We suspect that the online list-based estimate had an upward bias for Mindanao compared to the face-to-face results.

[Figure 4](#) reveals that poverty is a major source of preference falsification. A sizable amount of SDB—about 40 percentage points—was attributed to lower-class respondents (the DQ-based approval estimate was 0.897 [0.879, 0.916], and the list-based estimate was 0.470 [0.373, 0.566]). Among upper-class respondents, SDB was nearly zero (The DQ-based and list-based estimates were 0.896 [0.838, 0.954] and 0.883 [0.572, 1.195]).



**Figure 4.** Exploratory subgroup comparison of SDB magnitude according to demographic variables. *Note:* We did not compare socioeconomic class in the online survey because the respondent class was not measured in the online survey. The “college or higher” group includes respondents with some college education. Dots represent point estimates, and horizontal segments represent 95% CIs.

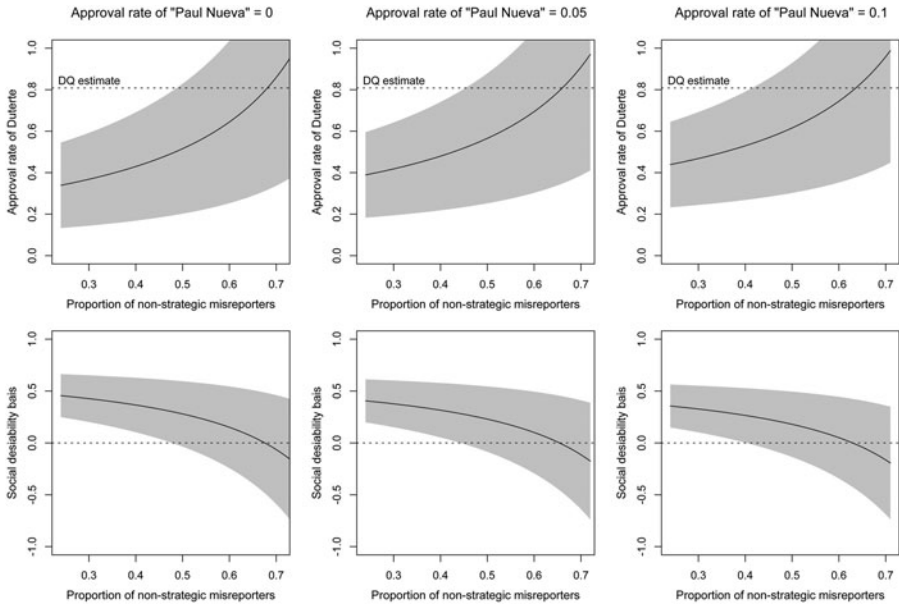
### *Sensitivity analysis for non-strategic misreporting*

Despite the scholarly belief that list experiments address the problem of strategic misreporting, some researchers have recently raised concerns that list experiments are vulnerable to non-strategic misreporting (Agerberg and Tannenber [2021](#); Ahlquist [2018](#); Kuhn and Vivyan [2022](#); Rimbau and Ostwald [2021](#)). To borrow Rimbau and Ostwald's ([2021](#)) explanation, while “[s]trategic errors arise when respondents lie to conceal their position on the sensitive issue” (172), “[n]on-strategic error includes such things as coding errors and poor quality responses that arise when respondents do not understand or rush through the list experiment” (173). If non-strategic misreporting exists, the estimated prevalence of affirmative response to a sensitive item using the difference-in-means estimator should be biased.

To address this issue, as indicated in our preregistration, we conducted an online survey including a placebo list experiment at the same time as the main survey and used the data obtained from this placebo survey for a sensitivity analysis.<sup>19</sup> One thousand respondents aged 44 or younger were recruited via Lucid Marketplace from February 26 to March 3, 2021. The recruitment procedure for this survey, including the quota settings and the method for omitting inattentive respondents, was identical to the main online survey, and we prevented respondents from giving answers to both surveys. We can thus consider the characteristics of respondents to be almost identical for the two surveys.

Except for demographic questions and a matrix-type item battery containing a directed question to detect satisficing participants, the placebo survey featured only a list experiment question. Respondents were block-randomized, as in the case of the main survey. The lists were the same as those used for the main survey, except that the list for the treatment group included the item “PAUL NUEVA” instead of “RODRIGO ‘DIGONG’ DUTERTE.” Although “Paul Nueva” is a completely plausible name in the Philippines, to the best of our knowledge, no politician has this name. As such, we can assume that most attentive respondents would not have considered “PAUL NUEVA” a politician they approved of; that is, “PAUL NUEVA” is assumed to have functioned as a placebo item. Under this assumption, we can check how sensitive our estimate of Duterte’s approval rate and the SDB to the proportion of non-strategic misreporters with simple computation.<sup>20</sup>

In this study, we found that the difference-in-means estimate for our placebo survey was 0.242, with a standard error of 0.060. Given the above assumption, if there were no non-strategic misreporters, this value should have been zero. Accordingly, it is certain that there were non-strategic misreporters. Furthermore, under an additional reasonable assumption, we can estimate that the proportion of non-strategic misreporters was no less than 24.2 percent (see Online Appendix B.4.3. for details). Using this value as the lower bound, we conducted a sensitivity analysis using a simulation. In that simulation, the proportion of non-strategic misreporters and the approval rate of “Paul Nueva.”<sup>21</sup> [Figure 5](#) presents the results of our sensitivity analysis. The top row shows how our estimate of Duterte’s approval rate (the vertical axis) changes depending on the assumed proportion of non-strategic misreporters (the horizontal axis). The bottom row shows the corresponding results for the SDB estimate. The first, second, and third columns show the results for the cases where the approval rate of “Paul Nueva” was set to 0, 0.05, and 0.1. The solid lines represent



**Figure 5.** Results of the sensitivity analysis for non-strategic misreporting using a placebo list experiment.

Note: The solid lines represent point estimates, the shaded areas represent 95% CIs, and the dotted lines in the top row panels represent the DQ-based estimates of Duterte’s approval rate in the main online survey.

point estimates, the shaded areas represent the 95 percent CIs, and the dotted lines in the top row panels indicate the DQ-based estimates of Duterte’s approval rate in the main online survey.

According to the top left panel, given that the approval rate of “Paul Nueva” is zero, only when the proportion of non-strategic misreporters is as high as 0.65 does the point estimate of Duterte’s approval rate reach its DQ-based estimate. Even when we focus on the upper limit of the 95 percent CI, the corresponding proportion does not exceed 0.49. Consequently, unless we suppose that more than half the respondents were non-strategic misreporters, we can be confident of SDB in respondents’ expression of their approval of Duterte.<sup>22</sup>

The bottom left panel shows that the SDB estimate is 0.436 [0.261, 0.610], 0.404 [0.214, 0.593], 0.336 [0.116, 0.557], and 0.242 [−0.023, 0.507] when we assume the proportion of non-strategic misreporters was 24 percent, 30 percent, 40 percent, and 50 percent. If we consider implausible the scenario in which more than half of the respondents were non-strategic misreporters, we can conclude that the SDB was 24 percentage points or higher. The second and third panels demonstrate that the setting of the approval rate of “Paul Nueva” did not substantially change the conclusion.

**Discussion**

Several caveats must be made before discussing our findings. First, our results only capture voter attitudes around February and March 2021, four years and eight



months into the Duterte presidency. This period coincided with a national lockdown and a high unemployment rate induced by the COVID-19 pandemic. Our survey also took place before the vaccination roll-out started.<sup>23</sup> This makes it possible to suggest that a significant portion of the Filipino population shared a prevailing sentiment of frustration and anxiety at the time of our survey. Accordingly, our results may not be relevant to other periods of Duterte's rule. For example, his popularity might have been authentic during the election campaign and earlier in his term. In an ethnographic analysis of the 2016 presidential campaign, Curato (2016) suggested that Duterte offered "hope" to the poor. However, as noted by Teodoro (2020), the country's relatively "threat-free" environment before 2016 was overwhelmed by a "climate of fear" due to the mass killings of drug suspects and severe punishments imposed on violators of COVID lockdown measures. It is crucial to note that our results should not be interpreted as suggesting a consistent inflation of Duterte's approval rating throughout his entire tenure due to SDB.

Second, Duterte was a relatively popular president even when SDB-induced answers are discounted, with the proportion of respondents who indicated genuine support at approximately 50 percent. Our findings should not be perceived as refuting the widely accepted notion that Duterte enjoyed significant popularity. Instead, this study can be regarded as an additional investigation that introduces a qualifying element, thereby tempering the certainty or extent of such understanding when most of the existing studies neglect to account for SDB in their interpretation of survey results.<sup>24</sup>

Third, Pulse Asia executive director Ana Tabunda, in an interview in June 2021, remarked that the non-response rate (those who refused to answer surveys) had not increased since the beginning of the Duterte administration, offering this as an indication of the absence of SDB (Rappler 2021). Regarding this point, Shen and Truex (2021, 1675, en.3) reported that surveys conducted in authoritarian regimes demonstrate only a slightly lower non-response rate than those in democracies, although it is usually thought that respondents in authoritarian regimes tend to exhibit higher levels of preference falsification. Furthermore, in a study of about 50 autocracies, Guriev and Treisman (2020) found that the level of state repression did not significantly affect the refusal rate among respondents. Although these findings may not squarely apply to this study's context, they nevertheless suggest that the rate of interview refusal may not relate to SDB concerning the regime support question, indicating that it remains possible to expect the presence of SDB even when the interview acceptance rate remains the same (as in the Duterte era).

With these caveats in mind, our study primarily confirms the suspicion that Duterte's popularity was inflated due to SDB. Roughly one in three respondents pretended to support the president in the face-to-face survey. Some may think that this is too high to be possible. Here, one can refer to studies concerning Russia's President Putin's popularity. In studies that employed the face-to-face list experiment, scholars have observed degrees of SDB concerning Putin's support ranging between 10 percentage points (Frye et al. 2017) and 44 percentage points (Kalinin 2015), depending on the time of the survey. Furthermore, in a preliminary study, Iglesias and Cheng (2022) conducted a face-to-face list experiment with residents of Metro Manila and its neighboring provinces in late 2021. They found that the magnitude of SDB

concerning support for Duterte was about 40 percentage points. These studies suggest that our results fall within the conceivable range.

Some may also think that our online survey's magnitude of SDB is too high at 28.3 percentage points. Preference falsification is generally associated with hesitation to reveal one's true preference in front of another person. In that case, one could argue that we should not observe a high degree of SDB online, where anonymity is high. Indeed, this was the logic behind our second hypothesis (face-to-face surveys invoke significantly higher SDB than online surveys), which was not supported empirically. It is conceivable that some respondents thought that responding online might not have guaranteed anonymity. Several studies conducted in other countries reported a considerable degree of SDB in online list experiments. On the support for the Chinese government, Li, Shi and Zhu (2018), Robinson and Tannenberg (2019), Nicholson and Huang (2022) reported approximately 28, 25, and 13 percentage points of SDB, respectively. On Russia's invasion of Ukraine, Chapkovski and Schaub (2022) reports 10 percentage points of overreporting on the support for Putin's decisions. These results suggest that some respondents do not regard online surveys as totally anonymous.

Furthermore, for both face-to-face and online surveys, we conducted a battery of tests to examine the validity of our assumptions and the robustness of our results, including placebo tests. These are reported in Section 'Sensitivity analysis for non-strategic misreporting' and Online Appendix. We found no issues regarding face-to-face survey. Although we detected a problem caused by non-strategic misreporters in online surveys, our sensitivity analysis showed that our overall conclusion holds under a reasonable assumption on the percentage of non-strategic misreporters. In sum, the available evidence suggests that our detection of SDB cannot be dismissed because it is too high.

It is critical to consider what drove respondents to pretend to support Duterte. One possible mechanism that emerges from our analyses is social conformity pressure. Testing our hypothesis (H6) concerning whether one's perception of Duterte's neighborhood popularity induces SDB produced strong and significant findings: In the face-to-face survey, the magnitude of SDB was 47.5 percentage points, while that of the online survey was 39.7 percentage points. Given this result, we posit that social pressure from "moral politics" (Kusaka 2017b) may have been an important mechanism behind conformity pressure.<sup>25</sup> Based on his ethnographic research in Metro Manila's urban slums, Kusaka (2017b) argued that the 1986 People Power was a catalyst event for the emergence of moral politics, which divided politics between "good versus evil" rather than between elites and the masses or the rich and poor. Subsequently, being on the side of the "good" became an important guiding principle for many Filipinos.

Furthermore, Kusaka (2017b, 2020) highlights that Duterte, as both a popular presidential candidate and president, fostered a prevailing moral discourse where supporters of his anti-drug campaign and strict COVID lockdown policies were regarded as "good," while those who opposed them were labeled "evil" or "*pasaway*" (naughty or trouble-maker in Filipino). This interpretation was corroborated by Warburg and Jensen (2020a, 2020b) through their long-term ethnographic research in Metro Manila's slum area. The amplification of Duterte's discourse by internet trolls further

served to silence his critics (Ong and Tapsell 2022). Under such circumstances, individuals likely experienced social pressure to conform to the state-imposed morality, which advocated for the killing of drug users and severe punishment for those who defied COVID-related restrictions. Notably, both Kusaka (2020) and Warburg and Jensen (2020a) observed the presence of discreet dissenters against Duterte's imposed morality. Considering these observations, we suspect that some survey respondents may have yielded to the influence of this dominant, state-imposed morality, feigning support for Duterte.

While this may come as a surprise to many, we found that SDB among Mindanaoans was extremely high. As reported in Section 'Exploratory subgroup analyses', our face-to-face study revealed that an overwhelming proportion of Mindanaoans (72.5 percentage points) pretended to support Duterte in our face-to-face survey. In Luzon and Visayas, SDB was below half (31.5 and 22.2 percentage points, respectively).

It is worth considering what might explain this regional variation. Again, the logic of morality-induced conformity pressure might be at play. Mindanaoans have been exposed to the moral discourse of Duterte for the longest because he served as mayor or vice mayor of Davao city for over 30 years before he emerged as a national-level politician. This makes it possible that the pressure to conform to Duterte-imposed morality was more strongly felt among Mindanaoans than Filipinos residing in other regions. The findings presented in Online Appendix C show that a significant proportion of Mindanaoans (91 percent) expressed the belief that individuals within their local community exhibited either a high or moderate level of satisfaction with Duterte. The corresponding figures for respondents from Luzon and Visayas are 82 percent and 80 percent. Although this discrepancy of 10 percent may not fully account for the regional disparity observed in SDB, we propose that variations in the intensity of social conformity pressure could potentially contribute to the differences.

Given the limited availability of data for comprehensive analysis, we are constrained to providing a speculative interpretation of our finding that individuals from lower socioeconomic backgrounds exhibit higher susceptibility to SDB than their wealthier counterparts. Specifically, the prevalence of SDB among respondents from lower-class strata was observed to be 42.8 percentage points, in contrast to a mere one percentage point among those from upper-class strata. However, unlike the previously discussed case of regional disparity, there was minimal divergence observed between different socioeconomic classes regarding respondents' perceptions of Duterte's popularity, as Online Appendix C shows.<sup>26</sup> This suggests that it is difficult to use higher social conformity pressure to explain high SDB among the poor (as operationalized by H6). Another plausible explanation—suggested by David (2020)—could be the fear of losing access to government pandemic subsidies, commonly referred to as "Ayuda." Considering the paucity of empirical investigations around this possibility, it is imperative for forthcoming studies to undertake an in-depth exploration of the mechanisms that establish a link between poverty and SDB.

On the question of "climate of fear" or respondents' SDB due to the fear of being a victim of Duterte's war on illegal drugs or red-tagging, our hypotheses (H3, H4, and H5) were not supported. This result can be interpreted in several ways. First, the fear

thesis was not relevant in reality despite the speculation of some analysts (Gavilan 2021; Teodoro 2020). Second, it is possible that the fear factor was relevant, but our operationalization was not valid. For example, asking whether participants personally knew victims of EJK (H3) or red-tagging (H4) may have provoked self-censorship.<sup>27</sup>

It is now worth considering how our empirical findings relate to existing studies concerning Duterte's popularity. One proposition that is testable using this study's dataset is the claim made by Dulay, Hicken and Holmes (2022) that Duterte's popularity was driven by ethnic affiliation: Filipino voters with non-Tagalog ethnicity tended to demonstrate greater support for Duterte. This argument holds even if the data used were contaminated by the SDB, provided the SDB did not correlate with respondent ethnicity. We used our data to examine whether the magnitude of SDB varied depending on respondent ethnicity and found that the difference was not statistically significant. Online Appendix C.4.3 reports that result. This analysis consequently substantiates that our findings corroborate and are compatible with the central argument presented by Dulay, Hicken and Holmes (2022).

## Conclusion

This article has empirically confirmed the suspicion of some Filipino analysts: Duterte's strikingly high popularity in the polls was highly inflated as of March 2021, the time of our survey. Our analyses also found that the poor, the Mindanaoans, and those who believed Duterte was popular in their neighborhoods were more likely to pretend to support him. Given that previous studies predominantly portray Duterte as a genuinely popular president, there may be concerns regarding the plausibility of our results. It is important to emphasize our commitment to transparency, as evidenced by pre-registering our research design and a battery of sensitivity analyses found in the Online Appendix. We also note that our study is constrained to capture a "snapshot" of voter attitudes at a given time and cannot easily be extrapolated to other time periods. Keeping these considerations in mind, we offer the following implications.

Our results elucidate the possible presence of another type of voter in backsliding democracies. Recent studies have shown that there are three types of voters who support illiberal incumbents: "autocrats" who dislike democracy, "militants" who support their party despite the party violating norms, and "majoritarians" who support the incumbents on the basis of being elected even when they violate norms (Grossman et al. 2021). Our study introduces a fourth type: "pretenders." These are voters who pretend to support the backsliding incumbent, especially in opinion polls. Our study suggests that such voters are of particular concern in developing countries, where most voters are socioeconomically vulnerable.<sup>28</sup> Thus, our findings imply that explanations for popular backsliding incumbents may differ between developing and developed countries.

Concerning democratic backsliding in the Philippines, our findings provide several new insights. The dominant narrative about Duterte features multiple aspects. First, voters were fascinated by the emergence of an illiberal leader, Duterte. They elected him and supported his illiberal policy (the war on drugs). Next, the implementation

of his signature policy was accompanied by attacks on democratic norms and institutions, yet voters overwhelmingly supported him. However, our study's results reveal that he was not as overwhelmingly popular as the polls suggested, at least toward the end of his term. His popularity was significantly inflated because many respondents pretended to support him. What would have happened if there had been no pretenders? In this counterfactual scenario, declining support might have sent Duterte an effective signal to refrain from further eroding democracy.<sup>29</sup> Thus, were survey results less inflated, poll results might have served as more effective checks on the executive, better pushing back against the executive aggrandizement.

This study also challenges the stereotype of Filipino voters. Existing studies tend to portray them as uncritical and naive, calling them “dutertards” (retarded Duterte supporters) or “bobotantes” (stupid voters). Our findings indicate such interpretations to be potentially superficial. In reality, many voters may be sophisticated strategists who play it “safe.” This might mean that Filipino voters are not “deluded”;<sup>30</sup> voters may be deluding the interviewers and observers. However, as we noted already, their individually optimal choice may have undesirable social and collective consequences: If a larger segment of respondents had answered that they sincerely did not support Duterte, the survey results might have functioned as a force against democratic backsliding.

What can our study tell us about post-Duterte politics? If Duterte's popularity was substantially inflated due to SDB, how can we make sense of the popularity of his successors, namely, President Ferdinand Marcos Jr. and Vice President Sara Duterte (Rodrigo Duterte's daughter)? Both won the 2022 election by broadly claiming to maintain Duterte's policies (Dulay et al. 2022). On this point, we maintain that it would be premature and inappropriate to extrapolate our findings to directly explain the electoral victories of Marcos Jr. and Sara Duterte for several reasons. First, voting outcomes involve not only individual psychology—our study's context—but also mobilization schemes encompassing patronage and vote buying (Aspinall et al. 2022). These broader factors significantly contribute to shaping electoral results. Second, as cautioned in the foregoing Discussion section, our study provides only a snapshot of respondent behavior in February/March 2021, and different dynamics may emerge during other time periods. Consequently, our study's main implication concerns not providing insights into post-Duterte politics but rather shedding light on the complexities associated with survey methodology and the interpretation of survey results.

Finally, we conclude this article by advocating for more studies addressing SDB in opinion surveys conducted in the Philippines and other backsliding democracies. An obvious first step is to conduct more surveys that employ indirect questioning techniques, such as the list experiments employed here. Furthermore, other methods, such as implicit association tests (Huang, Intawan, and Nicholson 2022) and the conjoint experiments (Horiuchi, Markovich, and Yamamoto 2022), may be useful. Second, scholars can triangulate various research approaches as suggested by Rosenfeld (2022). Examples include using combinations of indirect questioning, focus-group interviews, ethnographic studies, and direct questions from multiple survey sources. Considering what likely transpired in Duterte's Philippines, the inflated popularity of authoritarian leaders can justify their attacks on democracy. Properly estimating a nation's true “pulse” is critical to preventing democracy's slow death.

**Supplementary material.** The supplementary material for this article can be found at <https://doi.org/10.1017/jea.2023.18>. Replication materials are available at the Harvard Dataverse (<https://doi.org/10.7910/DVN/RQWIZL>).

**Acknowledgements.** For their helpful comments we thank the journal editor, two anonymous reviewers, Cleo Anne Calimbahin, Dean Dulay, Allen Hicken, Sol Iglesias, Marisa Kellam, Wataru Kusaka, Nico Ravanilla, Julio C. Teehankee, Mark Thompson, Marites Vitug and participants at the Philippine Political Studies Association annual meeting and at the Waseda University Contemporary Political Economy Institute's experimental political science study group meeting. We extend our special thanks to Ronald Holmes for his invaluable contributions during the design stage of this project.

**Sources of support.** This work was supported by the Japan Society for the Promotion of Science KAKENHI Grant Number 19H00584, the Sakuradakai Foundation grant, and Keio University's academic research promotion fund.

**Competing interests.** The authors declare none.

## Notes

1. Following Waldner and Lust (2018, 95), we use this term to describe backsliding that can happen in any type of regime, whether democratic or autocratic.
2. For example, see reviews provided in Waldner and Lust (2018) and Haggard and Kaufman (2021).
3. Pulse Asia adopted the following wording to measure Duterte's approval: "I have here names of some of our current government officials. Please tell us your opinion regarding their performance of their duties in the last three months. Using this board, Do you truly approve, somewhat approve, may approve or may disapprove, somewhat disapprove, or truly disapprove of the performance of Rodrigo 'Digong' Duterte of his duties as President or you have never heard, read, or watched anything about him at any time?" Pulse Asia calculated the sum of the percentage of "truly approve" and "somewhat approve" as the approval rating.
4. A related concept is "self-censorship," the act of suppressing one's potentially objectionable or embarrassing thoughts (Jiang and Yang 2016). Kuran (1995) notes that preference falsification is a broader concept than self-censorship because it can include both pretending and suppressing. Lying, hypocrisy, and insincerity should be distinguished from preference falsification, in that the latter "is a response to real or imagined social pressures [meant] to convey a particular preference" (Kuran 1995, 5).
5. A classic study of SDB defines it as "the tendency on behalf of the subjects to deny socially undesirable traits and to claim socially desirable ones, and the tendency to say things which place the speaker in a favourable light" (Nederhof 1985, 264). Blair, Coppock, and Moor (2020) prefers to use sensitivity bias instead of SDB.
6. The expression EJK has been used among Filipinos since the Marcos dictatorship; it refers to killing without due legal procedures. In the Duterte era, EJK attained a more specific meaning: killing for a drug-related incident.
7. According to the UN Office of the High Commissioner for Human Rights, the death toll associated with the government's anti-drug war operations from July 2016 to September 2020 is about 8,000; local civic organizations believe the number is much higher (Human Rights Watch 2021).
8. Twitter has been called X since July 2023.
9. When we preregistered our study in 2021, we referred to this hypothesis as herd behavior. Since then, we have come to realize that social conformity pressure is a more precise label that describes this hypothesis.
10. We registered our pre-analysis plan with AsPredicted: [https://aspredicted.org/TCX\\_VBL](https://aspredicted.org/TCX_VBL).
11. We conducted blocked randomization based on gender and region in the face-to-face survey and based on gender, age, education (completed high school or below, vocational or some college, and completed college or higher), and region in the online survey.
12. We conducted several diagnostic tests for assumptions required to make list experiments valid. Most diagnostics indicated no violations of the assumptions, but one test suggested a flaw in the online survey. This flaw may be due to non-strategic misreporters, which we discuss in detail and deal with in Section 'Sensitivity analysis for non-strategic misreporting' of this article. We report the results of the diagnostic tests in Online Appendix A.

13. Although we preregistered to compute DQ-based estimates using the entire sample, the results did not change when we utilized only the control group's responses to obtain the estimates. We considered this against Blair and Imai's (2012) concern that "indirect questioning may prime respondents [in the treatment group], invalidating comparison" (54). Furthermore, we confirmed that our substantive conclusions do not change when treating DK as "Disapprove." Details of these robustness verification methods are reported in Online Appendices B.1 and B.2.
14. We asked the following question that corresponds to Pulse Asia's wording shown in note 3: "Please tell us your opinion regarding President Duterte's performance of his duties during the last three months." We provided five options included in Pulse Asia's wording and DK.
15. We repeated our analysis using an alternative estimator that combines responses to the list experiment and the direct question proposed by Aronow et al. (2015), and we obtained nearly the same results. See Online Appendix B.3 for details.
16. The percentage of respondents who truly or somewhat approve of Duterte's performance of his duties during the last three months was 0.887 and 0.686 in the face-to-face and online surveys, respectively, excluding DK respondents. There is a noteworthy non-negligible difference in approval rating between the dichotomous questions and the Pulse Asia-style questions that can be attributed to the high prevalence of intermediate response categories ("may approve or may disapprove") in the online survey (0.156; the corresponding value was 0.065 in the face-to-face survey). We conjecture that online respondents' satisficing tendency (inattentiveness) caused the inflation of the middle-category responses.
17. Unlike the previous analyses, these analyses are exploratory and were not preregistered.
18. Online Appendix C illustrates the DQ- and list-based estimates of approval rating across subgroups.
19. Note that our placebo survey design is different from Riambau and Ostwald's (2021), though they also use the term "placebo." We discuss this issue in more detail in Online Appendix B.4.1.
20. We explain how our placebo design works in detail in Online Appendix B.4.2.
21. A recent article by Dulay, Hicken, and Holmes (2022) cited the working paper version of this article and claimed that our SDB "falls to 4 to 15 percentage points once the likelihood of non-strategic misreporting is taken into account via a 'placebo' list experiment" (544). Seemingly, they have interpreted the difference in means between the treatment and control groups, 0.242, to represent the magnitude of the upward bias in our SDB estimates due to non-strategic misreporting. That is, 0.395 (the SDB estimate in the face-to-face survey) minus 0.242 is 0.153, and 0.283 (the SDB estimate in the online survey) minus 0.242 is 0.041, which apparently informs their understanding of 4–15 percentage points. However, the value 0.242 cannot be simply interpreted as a representation of the magnitude of the bias. Our placebo list experiment can be interpreted only when we perform sensitivity analysis using the calculations described in conjunction with Figure 5 (see also Online Appendix B.4.2).
22. Because Riambau and Ostwald (2021) recognized that the problem of non-strategic misreporting is particularly severe among lower-educated people, we applied the sensitivity analysis to split samples based on educational level. As Online Appendix B.4.4 demonstrates, our conclusion—that it is safe to say that the SDB existed unless we suppose that no less than half the respondents were non-strategic misreporters—can be maintained.
23. The vaccination in the Philippines officially started on March 1, 2021 (Venzon 2021), while the major part of our survey was conducted in late February of 2021.
24. A notable exception is Dulay, Hicken, and Holmes (2022).
25. The social psychology literature informs various types of conformity mechanisms, and morality is among them. For a general review, see Schnuerch and Gibbons (2014); for a study focused on morality and SDB, see Kundu and Cummins (2013).
26. The percentage of respondents who answered that they believed people in their neighborhood were either very or somewhat satisfied with Duterte among ABC class was 80.4% while that percentage among DE class was 84.9%.
27. Figure 3 shows that 5.8% and 3.0% of face-to-face survey respondents confirmed that they knew victims of EJK and red-tagging, with 17.9% and 22.1% of online respondents confirming that they knew someone. Given that the online mode is less prone to SDB, we suspect that the face-to-face survey respondents may have been more likely to self-censor their response to this question item. Another possible reason of this disparity is that online respondents could have interpreted the question ("do you personally know a victim of EJK/red-tagged?") to include the victims they learned about online.

28. For example, democratic backsliding in Modi's India and Rajapaksa's Sri Lanka may involve "pretenders." Reports by Regan (2019) on Modi and by Human Rights Watch (2020) on Rajapaksa suggest that there is a context wherein survey respondents may be influenced by SDB.
29. We do not believe this scenario is unrealistic. For example, in October 2021, Duterte withdrew his plan to run for vice president when the majority of survey respondents answered that doing so would violate the constitution (Ranada 2021).
30. Consider, for example, the discussion of this view on this podcast: [www.newmandala.org/philippines-beyond-cliches-populist-voters](http://www.newmandala.org/philippines-beyond-cliches-populist-voters), accessed on January 29, 2022.

## References

- ABS-CBN. 2019. "Most Filipinos Worried Becoming EJK victims: SWS." <https://news.abs-cbn.com/news/03/01/19/most-filipinos-worried-becoming-ejk-victims-sws>. Accessed on January 2, 2023.
- Agerberg, Mattias, and Marcus Tannenbergh. 2021. "Dealing with Measurement Error in List Experiments: Choosing the Right Control List Design." *Research & Politics* 8 (2): 1–8.
- Ahlquist, John S. 2018. "List Experiment Design, Non-Strategic Respondent Error, and Item Count Technique Estimators." *Political Analysis* 26 (1): 34–53.
- Arguelles, Cleve V. 2019. "We are Rodrigo Duterte: Dimensions of the Philippine Populist Publics' Vote." *Asian Politics & Policy* 11 (3): 417–437.
- Aronow, Peter M., Alexander Coppock, Forrest W. Crawford, and Donald P. Green. 2015. "Combining List Experiment and Direct Question Estimates of Sensitive Behavior Prevalence." *Journal of Survey Statistics and Methodology* 3 (1): 43–66.
- Arugay, Aries A., and Dan Slater. 2019. "Polarization Without Poles: Machiavellian Conflicts and the Philippines' Lost Decade of Democracy, 2000–2010." *The Annals of the American Academy of Political and Social Science* 681 (1): 122–136.
- Asch, Solomon E. 1951. "Effects of Group Pressure upon the Modification and Distortion of Judgments." In *Groups, Leadership and Men: Research in Human Relations*, ed. Harold Guetzkow. Pittsburgh: Carnegie Press, 177–190.
- Aspinall, Edward, Meredith L. Weiss, Allen Hicken, and Paul D. Hutchcroft. 2022. *Mobilizing for Elections: Patronage and Political Machines in Southeast Asia*. Cambridge: Cambridge University Press.
- Bermeo, Nancy. 2016. "On Democratic Backsliding." *Journal of Democracy* 27 (1): 5–19.
- Blair, Graeme, Alexander Coppock, and Margaret Moor. 2020. "When to Worry about Sensitivity Bias: A Social Reference Theory and Evidence from 30 Years of List Experiments." *American Political Science Review* 114 (4): 1297–1315.
- Blair, Graeme, and Kosuke Imai. 2012. "Statistical Analysis of List Experiments." *Political Analysis* 20 (1): 47–77.
- Cabanes, J. V. A., and J. S. T. Cornelio. 2017. "Was Duterte's Rise Inevitable?" In *The Rise of Trolls in the Philippines (And What We Can Do About It)*, ed. Nicole Curato. Ithaca, NY: Cornell University Press, 233–252.
- Calonzo, Andrea. 2020. "Duterte's Trust Rating Rises to 91% Amid Region's Worst Outbreak." [www.bloomberg.com/news/articles/2020-10-05/duterte-s-trust-rating-rises-to-91-amid-region-s-worst-outbreak?leadSource=verify%20wall](http://www.bloomberg.com/news/articles/2020-10-05/duterte-s-trust-rating-rises-to-91-amid-region-s-worst-outbreak?leadSource=verify%20wall). Accessed on January 29, 2022.
- Canare, Tristan A., Ronald U. Mendoza, Jurel K. Yap, Leonardo M. Jaminola, and Gabrielle Ann S. Mendoza. 2021. "Unpacking Presidential Satisfaction: Preliminary Insights from Survey Data on the Bottom Poor in Metro Manila." *Philippine Political Science Journal* 42 (1): 1–29.
- Chapkovski, Philipp, and Max Schaub. 2022. "Do Russians Tell the Truth When They Say They Support the War in Ukraine? Evidence from a List Experiment." *LSE European Politics and Policy (EUROPP) blog*.
- Coppedge, Michael, John Gerring, Carl Henrik Knutsen, Staffan I. Lindberg, Jan Teorell, David Altman, Michael Bernhard, Agnes Cornell, M. Steven Fish, Lisa Gastaldi, et al. 2021. "V-Dem codebook v11."
- Curato, Nicole. 2016. "Politics of Anxiety, Politics of Hope: Penal Populism and Duterte's Rise to Power." *Journal of Current Southeast Asian Affairs* 35 (3): 91–109.
- . 2017. "Flirting with Authoritarian Fantasies? Rodrigo Duterte and the New Terms of Philippine Populism." *Journal of Contemporary Asia* 47 (1): 142–153.



- David, Randy. 2020. "The Sociology of Opinion Surveys." <https://opinion.inquirer.net/134356/the-sociology-of-opinion-surveys#ixzz6aYIKX6Lg>. Accessed on January 29, 2022.
- Duffy, Bobby, Kate Smith, George Terhanian, and John Bremer. 2005. "Comparing Data from Online and Face-to-Face Surveys." *International Journal of Market Research* 47 (6): 615–639.
- Dulay, Dean, Allen Hicken, and Ronald Holmes. 2022. "The Persistence of Ethnopolitist Support: The Case of Rodrigo Duterte's Philippines." *Journal of East Asian Studies* 22 (3): 525–553.
- Dulay, Dean, Allen Hicken, Ronald Holmes, and Anil Menon. 2022. "Who's Voting for 'Bongbong' Marcos to Be the Next Filipino President?" *Washington Post*, May 6, [www.washingtonpost.com/politics/2022/05/06/bongbong-marcos-duterte-philippines-election-may9/](http://www.washingtonpost.com/politics/2022/05/06/bongbong-marcos-duterte-philippines-election-may9/). Accessed on January 2, 2023.
- Elemia, Camille. 2020. "EXPLAINER: Duterte's High Ratings Despite Poor COVID-19 Response." [www.rappler.com/newsbreak/explainers/explainers-reasons-duterte-high-ratings-poor-covid-19-response/](http://www.rappler.com/newsbreak/explainers/explainers-reasons-duterte-high-ratings-poor-covid-19-response/). Accessed on January 29, 2022.
- Frye, Timothy, Scott Gehlbach, Kyle L. Marquardt, and Ora John Reuter. 2017. "Is Putin's Popularity Real?" *Post-Soviet Affairs* 33 (1): 1–15.
- Garrido, Marco. 2022. "The Ground for the Illiberal Turn in the Philippines." *Democratization* 29 (4): 673–691.
- Gavilan, Jodesz. 2021. "From UN Reports to Congress: The Many Times 'Red-Tagging' Was Used." [www.rappler.com/newsbreak/iq/list-times-term-red-tagging-use-united-nations-legislators-philippines/](http://www.rappler.com/newsbreak/iq/list-times-term-red-tagging-use-united-nations-legislators-philippines/). Accessed on January 29, 2022.
- Graham, Matthew H., and Milan W. Svobik. 2020. "Democracy in America? Partisanship, Polarization, and the Robustness of Support for Democracy in the United States." *American Political Science Review* 114 (2): 392–409.
- Grossman, Guy, Dorothy Kronick, Matthew Levendusky, and Marc Meredith. 2021. "The Majoritarian Threat to Liberal Democracy." *Journal of Experimental Political Science* 9 (1): 36–45.
- Guriey, Sergei, and Daniel Treisman. 2020. "The Popularity of Authoritarian Leaders: A Cross-National Investigation." *World Politics* 72 (4): 601–638.
- Haggard, Stephan, and Robert Kaufman. 2021. *Backsliding: Democratic Regress in the Contemporary World*. Cambridge: Cambridge University Press.
- Heerwegh, Dirk. 2009. "Mode Differences between Face-to-Face and Web Surveys: An Experimental Investigation of Data Quality and Social Desirability Effects." *International Journal of Public Opinion Research* 21 (1): 111–121.
- Horiuchi, Yusaku, Zachary Markovich, and Teppei Yamamoto. 2022. "Does Conjoint Analysis Mitigate Social Desirability Bias?" *Political Analysis* 30 (4): 535–549.
- Huang, Haifeng, Chanita Intawan, and Stephen P. Nicholson. 2022. "In Government We Trust: Implicit Political Trust and Regime Support in China." *Perspectives on Politics*: 1–19, doi:10.1017/S1537592722001037.
- Human Rights Watch. 2020. "Sri Lanka: Increasing Suppression of Dissent." Human Rights Watch, [www.hrw.org/news/2020/08/08/sri-lanka-increasing-suppression-dissent](http://www.hrw.org/news/2020/08/08/sri-lanka-increasing-suppression-dissent). Accessed on January 29, 2022.
- . 2021. "Philippines: Events in 2020." Human Rights Watch, [www.hrw.org/world-report/2021/country-chapters/philippines](http://www.hrw.org/world-report/2021/country-chapters/philippines). Accessed on January 29, 2022.
- Iglesias, Sol, and Chao-Yo Cheng. 2022. "Does State-Sponsored Violence Lead to Democratic Erosion? Evidence from a List Experiment in the Philippines." Presentation made at the Saw Swee Hock Southeast Asia Centre seminar on October 22, 2022.
- Imbens, Guido W., and Donald B. Rubin. 2015. *Causal Inference in Statistics, Social, and Biomedical Sciences*. New York: Cambridge University Press.
- Jennings, Ralph. 2019. "Why Philippines President, Criticized Abroad, Has Record High Approval." [www.voanews.com/a/east-asia-pacific\\_why-philippines-president-criticized-abroad-has-record-high-approval/6172458.html](http://www.voanews.com/a/east-asia-pacific_why-philippines-president-criticized-abroad-has-record-high-approval/6172458.html). Accessed on January 29, 2022.
- Jiang, Junyan, and Dali L. Yang. 2016. "Lying or Believing? Measuring Preference Falsification from a Political Purge in China." *Comparative Political Studies* 49 (5): 600–634.
- Johnson, David T., and Jon Fernquest. 2018. "Governing Through Killing: The War on Drugs in the Philippines." *Asian Journal of Law and Society* 5 (2): 359–390.
- Kalinin, Kirill. 2015. "Exploring Putin's Post-Crimean Supermajority." Available at SSRN 2658829, <https://ssrn.com/abstract=2658829>.

- Kenny, Paul D., and Ronald Holmes. 2020. "A New Penal Populism? Rodrigo Duterte, Public Opinion, and the War on Drugs in the Philippines." *Journal of East Asian Studies* 20 (2): 187–205.
- Kuhn, Patrick M., and Nick Vivyan. 2022. "The Misreporting Trade-Off between List Experiments and Direct Questions in Practice: Partition Validation Evidence from Two Countries." *Political Analysis* 30 (3): 381–402.
- Kundu, Payel, and Denise Dellarosa Cummins. 2013. "Morality and Conformity: The Asch Paradigm Applied to Moral Decisions." *Social Influence* 8 (4): 268–279.
- Kuran, Timur. 1995. *Private Truths, Public Lies*. Cambridge, MA: Harvard University Press.
- Kusaka, Wataru. 2017a. "Bandit Grabbed the State: Duterte's Moral Politics." *Philippine Sociological Review* 65: 49–75.
- . 2017b. *Moral Politics in the Philippines: Inequality, Democracy and the Urban Poor*. Singapore: NUS Press.
- . 2020. "Duterte's Disciplinary Quarantine: How a Moral Dichotomy was Constructed and Undermined." *Philippine Studies: Historical and Ethnographic Viewpoints* 68 (3/4): 423–442.
- Li, Xiaojun, Weiyi Shi, and Boliang Zhu. 2018. "The Face of Internet Recruitment: Evaluating the Labor Markets of Online Crowdsourcing Platforms in China." *Research & Politics* 5 (1): 1–8.
- Luo, Zhaotian, and Adam Przeworski. 2019. "Democracy and its Vulnerabilities: Dynamics of Democratic Backsliding." Available at SSRN 3469373, <https://ssrn.com/abstract=3469373>.
- Lupu, Noam, and Kristin Michelitch. 2018. "Advances in Survey Methods for the Developing World." *Annual Review of Political Science* 21: 195–214.
- Molo, John. 2020. "When Getting '91% Approval' Means Things Are Really Bad." <https://www.rappler.com/voices/thought-leaders/opinion-when-getting-91-percent-approval-means-things-really-bad>. Accessed on January 29, 2022.
- Nederhof, Anton J. 1985. "Methods of Coping with Social Desirability Bias: A Review." *European Journal of Social Psychology* 15 (3): 263–280.
- Nicholson, Stephen P., and Haifeng Huang. 2022. "Making the List: Reevaluating Political Trust and Social Desirability in China." *American Political Science Review* 117 (3): 1158–1169.
- Ong, Jonathan Corpus, and Ross Tapsell. 2022. "Demystifying Disinformation Shadow Economies: Fake News Work Models in Indonesia and the Philippines." *Asian Journal of Communication* 32 (3): 251–267.
- Przeworski, Adam. 2019. *Crises of Democracy*. Cambridge: Cambridge University Press.
- Rafael, Vicente L. 2021. *The Sovereign Trickster: Death and Laughter in the Age of Duterte*. Durham, NC: Duke University Press.
- Ranada, Pia. 2021. "Big Majority of Filipinos Think Duterte Bid for Vice President Violates Constitution – SWS." [www.rappler.com/nation/elections/most-filipinos-think-duterte-vice-presidential-bid-violates-constitution-sws-survey-june-2021/](http://www.rappler.com/nation/elections/most-filipinos-think-duterte-vice-presidential-bid-violates-constitution-sws-survey-june-2021/). Accessed on January 29, 2022.
- Rappler. 2021. "Rappler Talk: Duterte's Popularity as He Nears Last Year of Presidency." [www.rappler.com/nation/interview-ana-tabunda-duterte-popularity-last-year-presidency/](http://www.rappler.com/nation/interview-ana-tabunda-duterte-popularity-last-year-presidency/). Accessed on January 29, 2022.
- Regan, Helen. 2019. "Modi Says India's Minorities Are Living in World of Imaginary Fear. Muslims Disagree." CNN, <https://edition.cnn.com/2019/05/29/asia/india-modi-muslim-fear-intl/index.html>. Accessed on January 29, 2022.
- Riambau, Guillem, and Kai Ostwald. 2021. "Placebo Statements in List Experiments: Evidence from a Face-to-Face Survey in Singapore." *Political Science Research and Methods* 9 (1): 172–179.
- Richman, Wendy L., Sara Kiesler, Suzanne Weisband, and Fritz Drasgow. 1999. "A Meta-Analytic Study of Social Desirability Distortion in Computer-Administered Questionnaires, Traditional Questionnaires, and Interviews." *Journal of Applied Psychology* 84 (5): 754.
- Robinson, Darrel, and Marcus Tannenber. 2019. "Self-Censorship of Regime Support in Authoritarian States: Evidence from List Experiments in China." *Research & Politics* 6 (3): 1–9.
- Rosenfeld, Bryn. 2022. "Survey Research in Russia: In the Shadow of War." *Post-Soviet Affairs* 45 (6): 38–48.
- Santos, Vergel O. 2020. "False Trust." [www.rappler.com/voices/thought-leaders/opinion-newspoint-false-trust/](http://www.rappler.com/voices/thought-leaders/opinion-newspoint-false-trust/). Accessed on January 29, 2022.
- Schnuerch, Robert, and Henning Gibbons. 2014. "A Review of Neurocognitive Mechanisms of Social Conformity." *Social Psychology* 45 (6): 466–478.
- Shen, Xiaoxiao, and Rory Truex. 2021. "In Search of Self-Censorship." *British Journal of Political Science* 51 (4): 1672–1684.

- Svolik, Milan. 2020. "When Polarization Trumps Civic Virtue: Partisan Conflict and the Subversion of Democracy by Incumbents." *Quarterly Journal of Political Science* 15 (1): 3–31.
- Teehankee, Julio C. 2017. "Was Duterte's Rise Inevitable?" In *A Duterte Reader: Critical Essays on Rodrigo Duterte's Early Presidency*, ed. Nicole Curato. Ithaca, NY: Cornell University Press, 37–56.
- Teodoro, Luis V. 2020. "Surveys and the Fear Factor." [www.bulatlat.com/2020/10/17/surveys-and-the-fear-factor/](http://www.bulatlat.com/2020/10/17/surveys-and-the-fear-factor/). Accessed on January 29, 2022.
- Thompson, Mark R. 2018. "Why Duterte Remains So Popular: The Failure of the Philippines' Liberal Reformism." *Foreign Affairs*, [www.foreignaffairs.com/articles/philippines/2018-10-09/why-duterte-remains-so-popular](http://www.foreignaffairs.com/articles/philippines/2018-10-09/why-duterte-remains-so-popular). Accessed on January 29, 2022.
- Venzon, Cliff. 2021. "Philippines Starts COVID Vaccinations Courtesy of China." <https://asia.nikkei.com/Spotlight/Coronavirus/COVID-vaccines/Philippines-starts-COVID-vaccinations-courtesy-of-China>. Accessed on January 2, 2023.
- Waldner, David, and Ellen Lust. 2018. "Unwelcome Change: Coming to Terms with Democratic Backsliding." *Annual Review of Political Science* 21: 93–113.
- Warburg, Anna Bræmer, and Steffen Jensen. 2020a. "Ambiguous Fear in the War on Drugs: A Reconfiguration of Social and Moral Orders in the Philippines." *Journal of Southeast Asian Studies* 51 (1–2): 5–24.
- Warburg, Anna Braemer, and Steffen Jensen. 2020b. "Policing the War on Drugs and the Transformation of Urban Space in Manila." *Environment and Planning D: Society and Space* 38 (3): 399–416.
- Wedeen, Lisa. 1999. *Ambiguities of Domination*. Chicago: University of Chicago Press.

**Yuko Kasuya** is a professor of Comparative Politics at Keio University. Her research interests include regime transition, comparative political institutions, electoral systems, party politics, and East and Southeast Asian politics.

**Hirofumi Miwa** is a professor of Political Science at Gakushuin University. His research interests lie in the areas of political behavior, including voter ideology and partisanship, and Japanese politics.

---

**Cite this article:** Kasuya Y, Miwa H (2023). Pretending to Support? Duterte's Popularity and Democratic Backsliding in the Philippines. *Journal of East Asian Studies* 23, 411–437. <https://doi.org/10.1017/jea.2023.18>