

Preregistration:

Double standards in judging collective action

Background

Protest movements often spark contention about what actions are acceptable means of protest. Reactions to Black athletes kneeling during the national anthem to protest racist police violence in the United States are a case in point. Kneeling down is not violent, disruptive, or illegal and, therefore, does not seem like an extreme means of protest. And yet, only 29% of White Americans, compared to 66% of Black Americans, considered it appropriate for athletes to protest in this way (YouGov, 2017). At the time, commentators noted that public opinion seemed to condemn almost any action Black Americans took to protest police violence, leaving them with few acceptable avenues of protest (e.g., Noah, 2017). Protests following the killings of George Floyd, Breonna Taylor, and Ahmaud Arbery were again met with contention about what actions are proportionate responses to racist police violence. These examples suggest that *who* the protesters are and *what* they are protesting influences how acceptable observers judge their actions to be. In this study, we will examine double standards in judging collective action—that is, whether observers judge the same protest actions as more or less acceptable depending on their own and the protesters' group memberships and on the cause of the protest.

Hypotheses

In this study, we will test five hypotheses. We will test two predictions derived from social identity theory. First, we hypothesize that observers are more likely to judge an action as an acceptable means of protest when the protesters are ingroup members than when they are outgroup members (Hypothesis 1a). Second, we hypothesize that observers are more likely to judge an action as an acceptable means of protest when the goal of the protests would benefit the

observers' ingroup (Hypothesis 1b). We will test two predictions derived from system justification theory. First, we hypothesize that observers are more likely to judge an action as an acceptable means of protest when the goal of the protest is to support rather than challenge the system (Hypothesis 2a). Second, we hypothesize that observers who endorse system-justifying beliefs are more likely to consider system-supporting actions acceptable means of protest—and less likely to consider system-challenging actions acceptable means of protest—than observers who do not endorse system-justifying beliefs (Hypothesis 2b). In addition, we will test an alternative hypothesis. We hypothesize that observers are more likely to consider actions as acceptable means of protest when they support the cause of the protesters than when they oppose their cause (Hypothesis 3). We note that Hypotheses 1a, 1b, and 2a make competing predictions about differences between conditions (see Table A2) while Hypotheses 2b and 3 need not contradict each other or any of the other hypotheses.

Design

We will use a 2 (observers' group membership, quasi-experimental) x 2 (protesters' group membership, experimental) x 2 (protesters' cause, experimental) between-participants factorial design. Specifically, we will recruit Black and White Americans and ask them to judge the actions of either Black or White protesters who are either protesting for or against defunding the police. This design will allow us to assess the relative evidence for Hypotheses 1a, 1b, and 2a. We will measure system-justifying beliefs and support for the protesters' cause to gather correlational evidence for, respectively, Hypotheses 2b and 3.

Sample

We will recruit 1600 eligible participants for the Prolific subject pool who are 18 years old or older, live in the United States, and are American citizens. Of these, half will be Black Americans and half will be White Americans. We will continue collecting data until 1600 eligible participants (who do not fulfil our exclusion criteria) have completed the study. Simulations suggested that a sample size of 1600 participants is sufficient to estimate condition-wise standardized response tendencies with 95% uncertainty intervals of 0.18 to 0.32 in width.

Exclusions

We will exclude all participants who give a wrong answer to at least one of the three attention checks at the end of the study or who select a different answer to the ethnicity (simplified) question than they had selected when completing the Prolific pre-screening questionnaire.

Procedure

Participants will read the following paragraphs:

In 2020, police officers killed Ahmaud Arbery, Breonna Taylor, and George Floyd. Many were outraged that police officers had, once again, killed unarmed Black Americans. Across the United States, people called for changes to prevent future police violence.

Some argue that, to end police violence, we should take money away from police departments. Reducing police funding would mean fewer police officers on the street. Fewer police officers would mean fewer opportunities for them to turn violent. Reducing police funding would also leave more money for other services. Proponents argue for reallocating police funding to social services, housing, and education. Doing so would keep communities safer with fewer police officers. We refer to this position as “defunding the police”. This position differs from “reforming the police” which might mean increasing police funding and also differs from “abolishing the police” which means disbanding police departments altogether.

Participants will then be asked to answer whether the text had been about “reforming the police”, “defunding the police”, or “abolishing the police”. If they select the wrong answer, they will be instructed to re-read the text and select the right answer. On the next page, participants will state whether they supported or opposed the proposed solution.

Participants in the *system-challenging* protest condition will then read a text about protesters in support of defunding the police:

Earlier, you read about defunding the police as a possible solution to end police violence. Some local residents want to protest *for* defunding the police. They argue that reducing police funding would prevent police violence.

Participants in the *system-supporting* protest condition will instead read a text about protesters rallying against defunding the police:

Earlier, you read about defunding the police as a possible solution to end police violence. Some local residents want to protest *against* defunding the police. They argue that reducing police funding would mean fewer police officers serving their community.

Participants will then read either that “most of the protesters are Black” or that “most of the protesters are White”. Participants will again have to correctly answer multiple choice questions about the text (“Are the protesters for or against defunding the police?”, “Who are the protesters?”) before being able to proceed.

In the next pages, participants will complete the remaining measures (see below). On the final page, participants will respond to three attention checks: “In this study, you first read about a proposed solution to police violence. What was it?” (1 = *reforming the police*, 2 = *defunding the police*, 1 = *abolishing the police*); “In this study, you then read about protesters. Were these protesters for or against the proposed solution?” (1 = *for*, 2 = *against*); and “Were most of the protesters Black or White?” (1 = *Black*, 2 = *White*).

Measures

We will measure the outcome variable by asking participants to think about the protesters they have read about and to decide, for each of twenty-five protest actions, whether they think this action is “an acceptable means for them to protest [for/against] defunding the police” (1 = *yes*, 0 = *no*). We will encourage participants to ignore restrictions related to COVID-19 that might affect any of these actions. We will present items in a randomized order. We list all actions in Table A1.

We will measure support for defunding the police with one item: “Do you support or oppose defunding the police?” (1 = *strongly oppose*, 2 = *somewhat oppose*, 3 = *undecided*, 4 = *somewhat support*, 5 = *strongly support*).

We will measure system-justifying beliefs with eight items (adapted from Kay & Jost, 2003), for example, “in general, I find society to be fair” (1 = *strongly disagree*, 7 = *strongly agree*). We will report results from a confirmatory factor analysis to confirm whether all items load onto a single factor.

We will include additional measures to describe the sample, to report as descriptive statistics, or to use in not preregistered analyses. We will record participants’ gender and age with open-response items, import a range of other demographic variables from the Prolific database, and repeat the ethnicity (simplified) question from the Prolific pre-screening questionnaire. We will also

ask participants how outraged they are about recent incidents of police violence against Black Americans, to what extent they identify with the protesters described in this study, and to what extent they identify with their racial ingroup (1 = *not at all*, 7 = *very much*). We will also ask how often, if at all, participants have participated in protest actions such as the ones we had asked about earlier (1 = *never*, 5 = *very often*) and whether they have participated in protests *for* reforming, defunding or abolishing the police, *against* reforming, defunding or abolishing the police, or in neither. We will measure political orientation with a one-item semantic differential scale: “People often describe their political orientation as liberal or conservative. On a scale from liberal to conservative, where would you position yourself?”; 1 = *liberal*, 7 = *conservative*).

Transformations

We will transform the support for defunding the police item so that higher scores reflect support for the protesters’ cause. In the *system-supporting* protest condition, we will reverse code this item so that higher scores reflect opposition to defunding the police. In the *system-challenging* protest condition, we will leave this item as is so that higher scores reflect support for defunding the police. We will use factor loadings from a confirmatory factor analysis model in the model that includes system-justifying beliefs as a predictor variable.

Planned Analyses

To test our hypotheses, we will run a series of two-parameter logistic item response models with participants’ responses (1 = *yes*, 0 = *no*) to the question whether they thought each action was an acceptable means of protest as the outcome variable. This model estimates the probability that a participant considers an item to be an acceptable means of protest as a function of three parameters. For each item i , the model estimates how acceptable (β_i) and discriminating (α_i) the item is. For each participant j , the model estimates their unique propensity to consider protest actions acceptable (θ_j). For each of the $2 \times 2 \times 2 = 8$ conditions, the model estimates the average propensity, across participants in each condition k , to consider protest actions acceptable (δ_k).

We will estimate these models in CmdStanR (Gabry & Cešnovar, 2020) using Bayesian statistical methods. Bayesian inference involves choosing a likelihood function and prior distributions. The likelihood function links the observed data to the model parameters and states how likely the observed data

are given different values of said model parameters. Prior distributions state how plausible different values of said model parameters are before considering the observed data. Bayesian inference applies Bayes' theorem to update prior distributions in light of the observed data to produce posterior distributions. Posterior distributions state how plausible different values of the model parameters are given the observed data.

Our models derive the likelihood of participants' responses from a Bernoulli likelihood function with a logistic regression equation linking the two item parameters (α_i, β_i), the participant parameter (θ_j), and the condition parameter (δ_k) to the observed data. To identify these models, we assign θ_j to follow a standard normal distribution and constrain α_i to be non-negative by estimating and transforming $\log \alpha_i$ instead. We use partial pooling to estimate α_i, β_i , and δ_k . Doing so adjusts estimates for multiple comparison (Gelman et al., 2012). Our models assign weakly informative prior distributions to all model parameters, that is, (Half-)Cauchy(0, 5) prior distributions.

We will estimate three models to test our hypotheses. First, we will estimate the model described above to compare the relative evidence for Hypotheses 1a, 1b, and 2a. We will use posterior samples (for δ_k) to estimate the mean (standardized) difference for three contrasts implied by the hypotheses (see Table A2). For example, Hypothesis 1a implies that, on average, participants are more accepting of protest actions by ingroup than by outgroup members. For each contrast, we will report, first, the point estimate and 95% uncertainty interval, based on the median and the quantiles of the posterior samples, for the standardized difference between the contrasted conditions (equivalent to Cohen's d) and, second, the posterior probability of the implied difference between the contrasted conditions, based on the proportion of posterior samples that satisfies the contrast implied by the condition. For example, we might report that, on average, participants rated actions performed by ingroup protesters as more acceptable than the same actions performed by outgroup protesters ($d = 0.20, [0.15, 0.35], \Pr(d > 0) > .99$), indicating strong evidence for Hypothesis 1a.

Second, we will estimate a model that extends the first model by including standardized factor loadings for the system-justifying beliefs measure as a predictor variable. We will estimate the effect of this predictor variable with two *fixed* effects, one estimating the effect on judgements about system-justifying protest actions and one estimating the difference to the effect on judgements about system-challenging protest actions, and one *varying* effect estimating variance across the eight conditions. We will use the two fixed effects

to test Hypothesis 2b which implies that system-justifying beliefs will be associated with participants being more accepting of system-supporting protest actions and less accepting of system-challenging protest actions. We will, first, report point estimates and uncertainty intervals for the fixed effects for system-supporting and system-challenging protest actions and, second, report the posterior probabilities of these effects being above or below zero.

Third, we will estimate a model that extends the first model by including support for the protesters' cause as a predictor variable. We will estimate the effect of this predictor variable as a monotonic effect (Bürkner & Charpentier, 2020) which estimates the average positive or negative change across predictor categories as well as how much of this change occurs between each of the four pairs of adjacent predictor categories. Our model assigns a Dirichlet prior, $\alpha = \{1, 1, 1, 1\}$, to model the proportion of the overall change that is expected to occur between each of the four pairs of predictor categories. We will estimate average change across predictor categories as varying across the eight conditions. We will report a point estimate and an uncertainty interval for the fixed effect of change across predictor categories as well as the posterior probability that the fixed effect is positive. If the effect is positive, we will consider this as evidence for Hypothesis 3. If any of the three models return results that do not correspond to any of our hypotheses, we will still report these results as preregistered since our analyses, by using partial pooling, adjust estimates for multiple comparison.

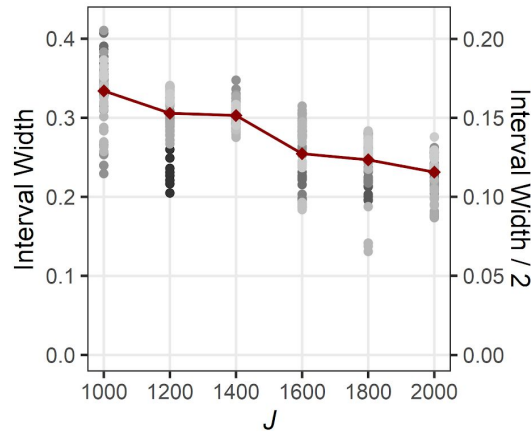
Simulations

For this preregistration, we ran simulations to determine the sample size for our study. We conducted this analysis in three steps. First, we applied the first model described under *Planned Analyses* to the data from a preceding study, which had used a similar outcome measure, with $I = 25$ items, $J = 443$ participants, and $K = 4$ conditions. Second, we randomly selected posterior samples from the model and used them to simulate 60 datasets, all with $I = 25$ items and $K = 8$ conditions. That is, we simulated 10 datasets for each sample size of $I = \{1000, 1200, 1400, 1600, 1800, 2000\}$ participants. We had to limit our simulations to 10 datasets per sample size as our analyses would otherwise have been too computationally demanding. Third, we applied the first model described under *Planned Analyses* to each of the 60 datasets. We calculated the width of the 95% uncertainty interval, based on the quantiles of posterior samples, for each condition in each dataset. This statistic reflects the precision of standardized condition-wise

estimates of the outcome variable; dividing this statistic by two reflects the size of the difference between two conditions that the model would be able to detect.

Figure 1 shows the results from our simulations. We found that at $J = 1000$ participants, 125 per condition, we should expect standardized condition-wise estimates to have 95% uncertainty intervals with widths between $z = 0.23$ and $z = 0.41$. Based on a previous study, we consider these uncertainty intervals as too wide to reliably detect relevant differences between conditions. We also found that the precision of our estimates increased only marginally between 1600 and 2000 participants. At $J = 1600$ participants, 200 per condition, we should expect standardized condition-wise estimates to have 95% uncertainty intervals with widths between $z = 0.18$ and $z = 0.31$ —that is, narrow enough to detect even small differences between conditions from $z/2 = 0.09$ to $z/2 = 0.16$. We conclude that, based on our simulations, a sample size of $J = 1600$ offers the best compromise between precision and expense.

Figure 1
Results from our simulations



Note. Grey dots (grouped by dataset) show the width of the 95% uncertainty interval for each condition in a dataset while red diamonds show the mean interval width for each sample size.

Appendix

Table A1

List of actions to be used to measure judgements about protest actions

#	Action
1	<i>attend a protest march even though it might be unlawful</i>
2	<i>hand out flyers, leaflets, or pamphlets</i>
3	make a public speech
4	hold meetings to influence the public
5	attend or organise a protest march
6	<i>wear or display political symbols (e.g., patches, flags, bumper stickers)</i>
7	use social media (e.g., Facebook, Twitter, Instagram) to influence the public
8	refuse to buy goods or services from companies that advocate [against/for] defunding the police (boycott)
9	paste up posters with political messages in places where it is allowed and encouraged
10	join or form a group of activists
11	refuse to accept honors or awards in protest
12	donate to politicians who advocate [for/against] defunding the police
13	refuse to work (strike)
14	pay for adverts on social media (e.g., Facebook, Twitter, Instagram) to influence public opinion
15	donate to activist groups
16	visit people in their homes to convince them about the issue (canvassing, door knocking)
17	stand or sit in a building and refuse to leave (stand-in, sit-in)
18	<i>attend a protest march while carrying a firearm (where legal)</i>
19	enter and refuse to leave a building (occupation)
20	paste up posters with political messages in places where it is not allowed or encouraged
21	disrupt traffic (e.g., blocking roads)
22	<i>refuse to pay fees, fines, or taxes in protest</i>
23	mock or insult individuals who are [against/for] defunding the police
24	spray paint political messages in public places
25	<i>attend a protest march even though it might turn violent</i>

Note. Actions in *italics* replace actions used in a previous study that are either redundant or do not fit the context of the preregistered study.

Table A2

Contrasts implied by Hypotheses 1a, 1b, and 2a

Contrast	Group Membership		Protesters' Cause	
	Observers	Protesters	System-Challenging	System-Supporting
<i>Hypothesis 1a</i>	Black	Black	+	+
	Black	White	-	-
	White	Black	-	-
	White	White	+	+
<i>Hypothesis 1b</i>	Black	Black	+	-
	Black	White	+	-
	White	Black	-	+
	White	White	-	+
<i>Hypothesis 2a</i>	Black	Black	-	+
	Black	White	-	+
	White	Black	-	+
	White	White	-	+

Protesters' cause refers to whether protesters are protesting for defunding the police (challenging the system) or against defunding the police (supporting the system).

References

- Bürkner, P., & Charpentier, E. (2020). Modelling monotonic effects of ordinal predictors in Bayesian regression models. *British Journal of Mathematical and Statistical Psychology*, 73(3), 420–451. <https://doi.org/10.1111/bmsp.12195>
- Gabry, J., & Češnovar, R. (2020). *cmdstanr: R interface to CmdStan* [R package]. <https://mc-stan.org/cmdstanr>
- Gelman, A., Hill, J., & Yajima, M. (2012). Why we (usually) don't have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness*, 5(2), 189–211. <https://doi.org/10.1080/19345747.2011.618213>
- Noah, T. (2017, September 25). *When is the right time for Black people to protest?* [Video]. <http://www.cc.com/video-clips/nujtiv/the-daily-show-with-trevor-noah-when-is-the-right-time-for-black-people-to-protest>
- YouGov. (2017). *HuffPost: NFL, September 25–26, 2017* [Data set]. <https://perma.cc/3Q3V-4GC4>