# Replication of Levy (2022)

Leo Y. Tien *

tien.y@wustl.edu

December 2023

## 1 Introduction

Despite abundant research on the diverse repertoire of violence (Dumaine et al., 2022; Fujii, 2013; Gutiérrez-Sanín & Wood, 2017; Stanton, 2016; Wood, 2006), few studies quantitatively compare the impact of various forms of violence. Furthermore, within the realm of conflict studies, the effects of wartime violence on elections have been extensively explored, yet the findings are inconclusive (Balcells, 2012; Daly, 2019; Gallego, 2018; Matanock, 2017; Weintraub et al., 2015).

Levy (2022) fills the gap by studying what kinds of alleged perpetrators voters are more likely to vote for. She argues that the voters appraise candidates' alleged involvement in violence by considering how their participation reflects competence and integrity. She also argues that sexual violence violates a stronger norm than lethal violence. Thus, sexual violence perpetrators should be perceived as having less integrity. In her empirical study, she conducts a conjoint survey experiment in Colombia within the context of an election featuring former combatants as candidates. Her findings reveal that respondents support candidates who adhere to less stringent norms, exhibit lower levels of agency, and possess less clearly defined causal connections to the victims. These results have significant implications for understanding people's perceptions of different violence and post-conflict elections.

*PhD student in the Department of Political Science, Washington University in St. Louis

## 2 Replication

Levy (2022) utilizes a conjoint experiment to study what kinds of alleged perpetrators in wartime voters are more likely to vote for in a post-conflict election. Levy argues that voters evaluate how the violence reflects a candidate's integrity and competence, particularly security competence. Thus, most attributes included in the conjoint design indicate the two aspects.

She implemented the experiment online in Colombia with the help of a local survey firm. There were 1589 respondents, and each should complete four forced-choice tasks in which they picked a profile they preferred out of a pair. They also rated "how likely they would be to vote for each candidate" from 1 to 5.

### 2.1 Replication Process

In this replication exercise, I replicate the Average Marginal Component Effect (AMCE) estimates in Figure 2 of Levy (2022). Data, original codes, and codebooks are provided by the author on her website. To analyze the conjoint experiment data, I follow her codes to:

1. establish a baseline value for each conjoint attribute/component (e.g., age, gender)
2. establish factor order for the analysis
3. trim the dataset by removing rows where the profile selection variable is NA (i.e., did not select a profile in the experiment).
4. run the analysis

## 2.2 Replication Results

Figure 1 below shows the results from my replication. My findings are identical to the ones in the original paper. Regarding the "dyadic integrity" indicator, involvement in sexual violence decreases the probability that the candidate is chosen by 4.2 percentage points compared with involvement in killings. Agency of the perpetrators is also important as an enthusiastic violent perpetrator is less likely to be chosen than a reluctant one by 17 percentage points. Forced recruits are more likely to be chosen than an ideologically committed person by 8.6 percentage points. Levy also argues that the clarity of causation between the perpetrator and victim is an important factor. This causation is captured by whether the perpetrator has physical contact with the victim and also whether the violence would have been unlikely to happen without the wrongdoings of the perpetrator. The results here show that candidates who ordered someone else to commit instead of committing on his/her own see an increase in the probability of being preferred by five percentage points. Candidates who are just following orders and who "did not prevent someone else from committing" are also more likely to be chosen than voluntary perpetrators by 11 and 10 percentage points, respectively.

Security competence-related hypotheses, on the other hand, receive little support as in the original paper. Winning all battles does not make the candidates more likely to be chosen than candidates who won half of the battles or only few battles. Additionally, targeting people living in an area that supported the enemy or those living in an area controlled by the enemy makes candidates less likely to be chosen than those who only target enemy informants.

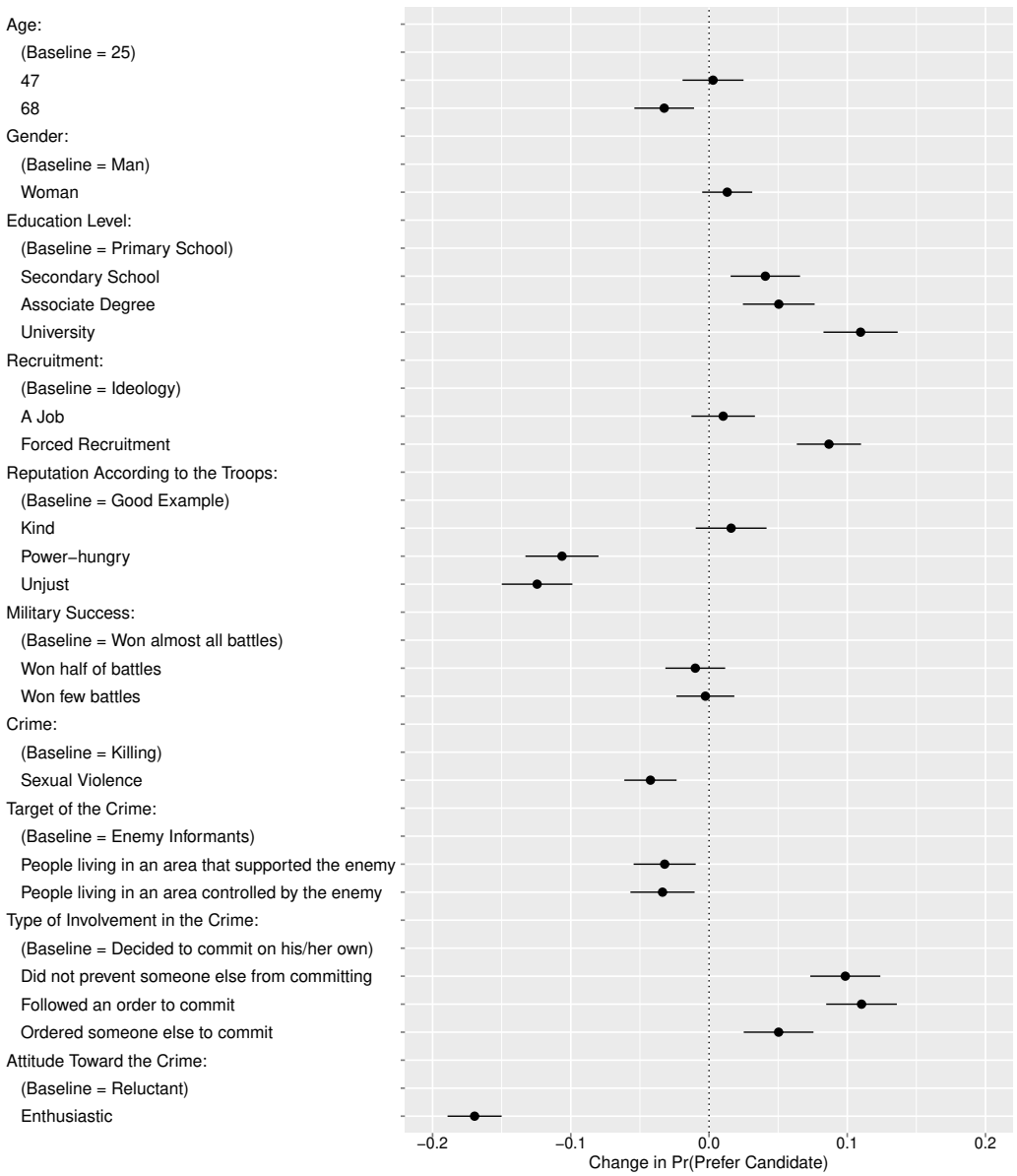


Figure 1: Replication of Figure 2 in Levy (2022)

## 3 The Limitation of AMCE

Despite the popularity of conjoint experiments in political science, recent scholars have raised valid concerns about the limitations of such research methods. In particular, Abramson et al. (2022) have advised scholars to exercise caution when interpreting the Average Marginal Component Effect (AMCE) introduced by Hainmueller et al. (2014) in the context of conjoint analysis. They highlight that political scientists sometimes misinterpret the results derived from AMCE as a representation of majority preferences. This misinterpretation can be misleading. For example, when the AMCE for males (baseline: female) is negative, the majoritarian interpretation could look like this: "A majority of voters prefer female to male candidates." The interpretation is wrong because AMCE considers not only the direction of preference (female preferred to male) but also the intensity of it. If there is a correlation between intensity and direction, AMCE cannot represent the majority preference, and thus, the statement mentioned above does not hold.

AMCE's link with majority preference is also sensitive to the number of attributes (i.e., gender) and the values within the attributes (i.e., male, female), (Abramson et al., 2022; Cuesta et al., 2022). This is because each additional attribute and value could change the preference ranking. As there are more and more attributes and values (subsequently, more total profiles), there is less and less certainty on the majority preference as each added value increases the number of possible profiles, changing and complicating the ranking preference.

To address the challenge of accurately inferring majority preference, Abramson et al. (2022) suggests establishing upper and lower bounds for the proportion of the population that favors one feature over another. The bounds are a

function of the AMCE, the total number of possible candidate profiles in the experiment, and the number of values in the attribute of interest. The bounds become wider when the number of profiles grows larger. The lower bounds of an AMCE estimate of 0.15 under five profiles would go below the 50% horizontal line, meaning the estimate may not indicate the majority preference. Since most conjoint experiments have profiles significantly more than 5, and also AMCE = 0.15 is reasonably large, the proposed bounds cast doubt upon the majority interpretation in many papers using AMCE.

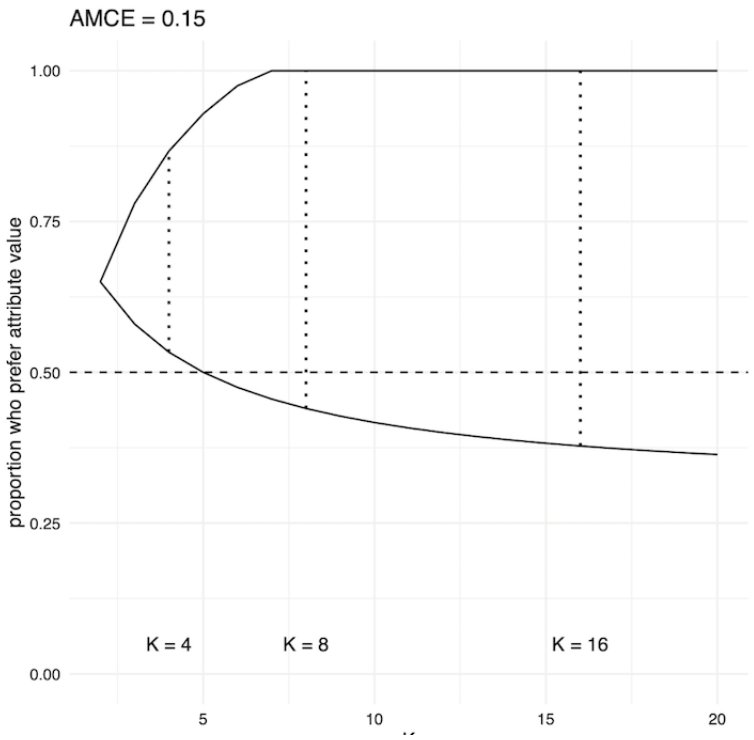


Figure 2: Source: Abramson et al. (2022)

Since Levy (2022) also implicitly interprets her results using a majoritarian approach, for example, by stating, "Respondents are more supportive of candidates who violate less strict norms," it becomes crucial to scrutinize the validity

of such assertions. I will discuss it more in the Extension section.

Not only does the AMCE fail to guarantee majority preference, but its validity in discovering preference is also in question. Ganter (2023) argues that scholars who use conjoint analysis often conflate between two goals: uncovering preference and understanding the probability of selection based on the specific experiment setup. The distinction matters because when we study the selection probability, we also consider the composition effects between attributes. However, when we are only interested in the preference, we should leave out the composition effect and only estimate the preference. For example, in Levy (2022) 's experiment, one profile contains ten attributes (e.g., gender, crime, military success...). A "selection question" related to this experiment could be: How likely is the candidate who committed sexual violence chosen over the other candidate who is a murder convict? Or, how likely the female candidate is to be chosen compared to a male candidate?

Note that this is different than asking: do people prefer sexual violence offenders candidates to murderers candidates? Or, do respondents prefer females to males? Do the crimes the candidates committed matter more than gender? The critical distinction is that the former question set considers the selection probability with consideration of the composite effect coming from the distribution of other attributes. Meanwhile, the preference questions, the later examples, concern the general preference between attributes and within attribute values. Ganter (2023) proves that AMCE is only suitable for answering the selection problem but not the preference problem, and he calls for an alternative estimand: Average Component Preference (ACP). Unlike the AMCE, the ACP teases out the preference effect from other composite effects that go into the

AMCE estimates. I will discuss the estimator for ACP in the next section.

# 4 Extention

## 4.1 Majority Preference

Based on the limitation of AMCE, we should be cautious when interpreting the results in Levy (2022). First, there should be some correlation between the direction of preference and the intensity of it. Levy's theory suggests that sexual violence violates stricter norms than lethal violence, and thus, respondents should support sexual violence perpetrators less. Consequently, knowing the respondents prefer candidates who perpetrate lethal violence instead of sexual violence, we should expect the respondents to prefer the lethal violence perpetrator strongly over the sexual violence offender. Second, there are several attributes, and the values for each attribute result in a large number of profiles (K = 3*4*4*2*4*2*3*3*2 = 41472). Hence, the lower bound of the fraction of people who prefer a feature should be substantial, making the majoritarian interpretation in the paper spurious. Overall, I expect that:

H1: *All the lower bound of the AMCE estimates will be below 50%*

To construct the lower and higher bounds, I use the formula provided by Abramson et al. (2022):

$$y \in [max\{\frac{\pi(t_1,t_0)\tau K+\tau}{K(\tau-1)+\tau}, 0\},\ min\{\frac{\pi(t_1,t_0)\tau K+K(\tau-1)}{K(\tau-1)+\tau}, 1\}]$$

where $\pi$ is the AMCE, $K$ is the number of potential profiles, $\tau$ is the number of distinct values in the attribute of interest that can be taken.

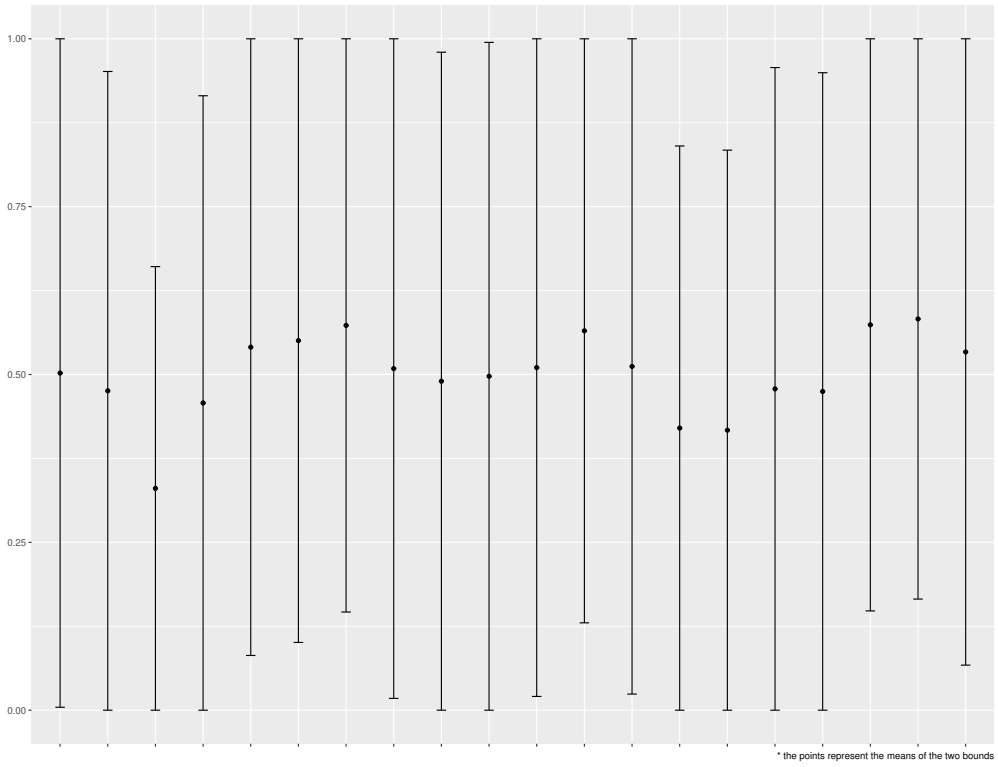


Figure 3: Do the AMCE Estimates Infer Majority Preference?

Recall we can only be certain that the estimates represent majority preference when the lower bounds are above 50% (i.e., over half of the population favors one feature over another). In Figure 3, I plot the percentage on the y-axis, and each vertical line represents an AMCE estimate. It should be noted that the points between the bounds are the average between the two bounds (i.e., $\frac{upper + lower\ bound}{2}$). Not surprisingly, none of the 20 estimates reported in Levy (2022) has a lower bound over 50%, meaning none of her AMCE estimates can be understood as the majority preference.

## 4.2 Average Component Preference for Measuring Preferences

Another approach to extend from Levy (2022) is rigorously examining preference patterns, both cross-attribute and within-attribute. Ganter (2023) argues that AMCE cannot identify both the preference and causal effect of an attribute on the profile's selection at the same time. He introduces the Average Component Preference (ACP), which can better explain preference patterns. In the case of Levy (2022), she claims to find support for the integrity hypothesis but no support for the security competence hypothesis. However, the ACP is needed to ascertain which attribute is more important in the respondents' minds.

Luckily, the ACP estimator is consistent and requires the same data as AMCE for a forced-choice conjoint analysis. Formally, it can be estimated with:

$$\hat{\pi}_l(t_l; w_{t_l}) = \sum_{\substack{t'_l \in \mathrm{T}_l \\ t'_l \neq t_l}} w_{t_l}(t'_l)\{\mathbb{E}[Y_{ij}|T_{ijl} = t_l, T_{i[-j]l} = t'_l] - 0.5\}$$

where $t_l$ is a profile (the value $t$ for the attribute $l$), $w_l$ is a vector of level weights, $\mathrm{T}_l$ is the set of levels for attribute $l$, $i$ and $j$ represent the respondent and their choice within a pair, respectively. Despite the usefulness of ACP in this scenario, two issues should be addressed.

First, the concepts of security competence and dyadic integrity are composite in nature and in operationalization, and their components are not independent. Hence, we should be cautious about making a general claim on their relative importance based on the original analysis or mine here. For example, a component categorized as a security competence component by Levy, the target of crimes, could also be counted as a component of integrity (and its mid-level component:

norm violation) because indiscriminate violence harms ordinary civilians who pose no threats to the perpetrators. The action even violates humanitarian law (i.e., Geneva Conventions Additional Protocol I (1977)). Second, because the two concepts are captured by multiple different indicators, making claims about the relative importance of the big concepts is complex. However, suppose all the indicators of the other competence turn out to be insignificant, and some of the indicators of the focal competence are significant. In that case, it is still straightforward to interpret that one kind of competence matters more. Thus, I hypothesize that:

H2: *Integrity is a more important factor than security competence*

And my hypothesis is supported only when all the indicators of security competence turn out to be insignificant and some of the indicators of integrity competence are significant. Figure 4 which is shown in Levy (2022) provides an overview of the concept-indicator operationalization:

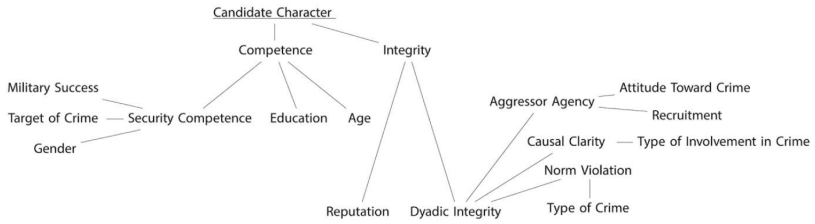


Figure 4: Overview of operationalization - original Figure 1 (Levy, 2022)

Figure 5 shows the ACP estimates of all the attributes and their values (levels) analyzed in Figure 1. The error bars represent the 95% confidence interval. First, the only attributes with no statistically significant levels are *Gender, Military Success* and *Target of the Crime*. The three attributes are the exact three

components of security competence. The evidence thus provides support for my hypothesis 2.

Regarding how to interpret the ACP estimates, Ganter (2023) recommends a more holistic approach to understanding preference direction and intensity instead of focusing on the point estimates. In this manner, the results demonstrate that the respondents prefer younger (25 or 47) candidates to older ones (65). Respondents also like one with only primary school education less than those who receive higher degrees. Forced recruits are more preferred than ideologically committed personnel and economic recruits. Respondents also prefer those with kind reputations and they prefer those with a name of unjust less. Killer candidates are viewed more favorably than sexual offender candidates. Those who follow orders to commit violence are preferred over other conditions. Finally, reluctant violent perpetrators are more preferred to enthusiastic perpetrators.

Overall, results from the ACP estimates largely resemble the AMCE results. It is normal to see high correlations between the two sets of results, and the similarity does not mean ACP is not needed, especially if the goal is to uncover preference patterns. That said, there is one difference that is worth mentioning. The results in Figure 1 might suggest that "Target of the Crime" (or, targeted versus indiscriminate violence) matters in respondent's preference. However, the ACP estimates for the levels in the same attribute are not statistically significant. Therefore, we should be less confident about whether the target of crime contributes to respondent's preference formation.

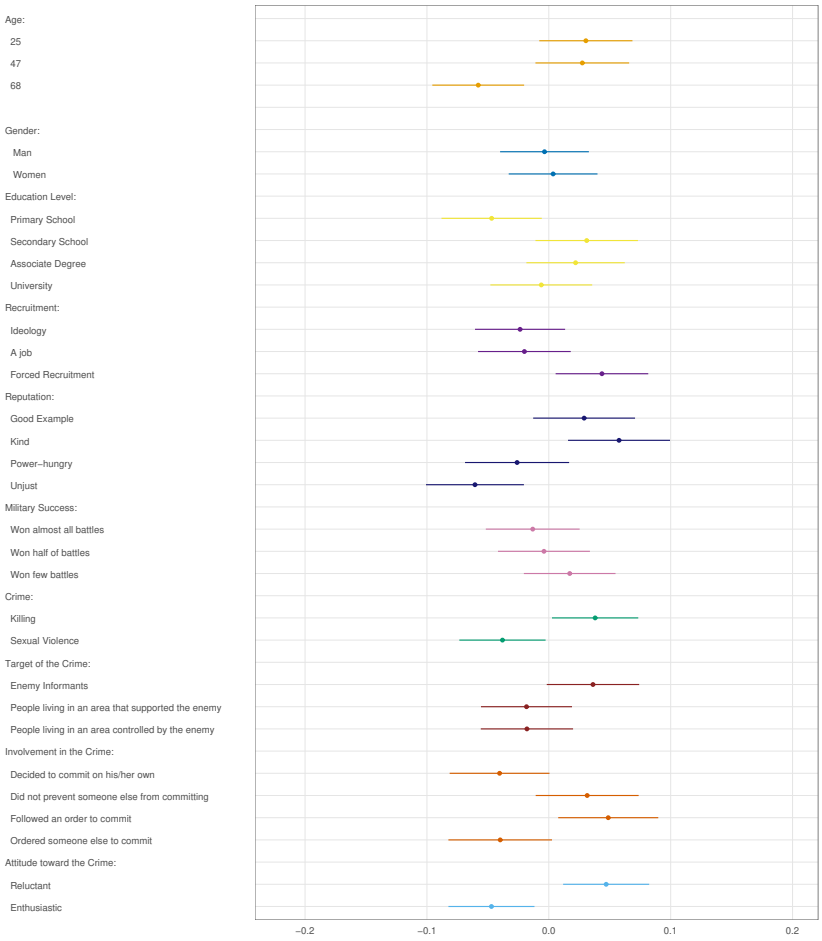


Figure 5: ACP estimates

## 5 Conclusion and discussion

Drawing on psychology and conducting a conjoint experiment, Levy (2022) investigates how voters choose in a common scenario where candidates were past violence perpetrators. I am able to replicate the main findings (Figure 1) in the original paper (original Figure 2 in Levy (2022)). The estimates and the statistical significance levels are identical to hers in the original paper. The

AMCE findings in Levy (2022), however, might not represent the majority preference that she implied. Abramson et al. (2022) suggests that AMCE does not entail majority preference when the direction and the intensity of preference are correlated. Also, this study's large amount of attributes and the values within each attribute further prevent a majority preference inference. Using the upper and lower bounds formula by Abramson et al. (2022), I construct the bounds for the proportion of the population that favors one feature over another. I find that none of the AMCE estimates in the paper can represent the majority preference for sure. Another issue with the AMCE estimates is its limitation with answering preference-related questions (Ganter, 2023). Following Ganter (2023) 's suggestion, I also estimate the ACP which is meant to uncover preference patterns better. The ACP results largely resemble those of the AMCE. In addition, I find that integrity is more important than security competence in the respondent's preference.

Puzzling questions arise from my extension: does the ACP also suffer from the majority preference inference issue as it does for the AMCE? Should we and could we construct bounds for the ACP? Moreover, do some integrity components matter more than others? Based on Figure 5, it is unclear whether any of the integrity components is more crucial than the other since all the estimates of the four attributes fall into a close range, and most of their error bars overlap. However, one can come up with other indicators of integrity and apply a similar research design to different scenarios. Since this empirical study of "perpetrators as candidates" and "public opinion on perpetrator's integrity" is relatively novel, more similar research can build up the external validity of the findings. With future research, we can also discover what kinds of integrity matter in what specific post-conflict settings.

## Replication and Extension Codes

The replication codes and the current draft are uploaded to GitHub:
https://github.com/yuantien605/Replication-of-Levy-2022-.git

## References


Abramson, S. F., Kocak, K., & Magazinnik, A. (2022). What do we learn about voter preferences from conjoint experiments? *American Journal of Political Science*, *66*(4), 1008–1020.

Balcells, L. (2012). The consequences of victimization on political identities: Evidence from spain [Publisher: SAGE Publications Inc]. *Politics & Society*, *40*(3), 311–347.

Cuesta, B. d. l., Egami, N., & Imai, K. (2022). Improving the external validity of conjoint analysis: The essential role of profile distribution [Publisher: Cambridge University Press]. *Political Analysis*, *30*(1), 19–45.

Daly, S. Z. (2019). Voting for victors: Why violent actors win postwar elections [Publisher: Cambridge University Press]. *World Politics*, *71*(4), 747–805.

Dumaine, L., Nordås, R., Gargiulo, M., & Wood, E. J. (2022). Repertoires of conflict-related sexual violence: Introducing the RSVAC data package [Publisher: SAGE Publications Ltd]. *Journal of Peace Research*, *59*(4), 611–621.

Fujii, L. A. (2013). The puzzle of extra-lethal violence [Publisher: [American Political Science Association, Cambridge University Press]]. *Perspectives on Politics*, *11*(2), 410–426.

Gallego, J. (2018). Civil conflict and voting behavior: Evidence from colombia [Publisher: [Sage Publications, Ltd., Sage Publications, Inc.]]. *Conflict Management and Peace Science*, *35*(6), 601–621.

Ganter, F. (2023). Identification of preferences in forced-choice conjoint experiments: Reassessing the quantity of interest [Publisher: Cambridge University Press]. *Political Analysis*, *31*(1), 98–112.

Gutiérrez-Sanín, F., & Wood, E. J. (2017). What should we mean by "pattern of political violence"? repertoire, targeting, frequency, and technique [Publisher: Cambridge University Press]. *Perspectives on Politics*, *15*(1), 20–41.

Hainmueller, J., Hopkins, D. J., & Yamamoto, T. (2014). Causal inference in conjoint analysis: Understanding multidimensional choices via stated preference experiments [Publisher: Cambridge University Press]. *Political Analysis*, *22*(1), 1–30.

Levy, G. (2022). Evaluations of violence at the polls: Civilian victimization and support for perpetrators after war [Publisher: The University of Chicago Press]. *The Journal of Politics*, *84*(2), 783–797.

Matanock, A. M. (2017). *Electing peace: From civil conflict to political participation*. Cambridge University Press.

Stanton, J. A. (2016). *Violence and restraint in civil war: Civilian targeting in the shadow of international law*. Cambridge University Press.

Weintraub, M., Vargas, J. F., & Flores, T. E. (2015). Vote choice and legacies of violence: Evidence from the 2014 colombian presidential elections [Publisher: SAGE Publications Ltd]. *Research & Politics*, *2*(2), 2053168015573348.

Wood, E. J. (2006). Variation in sexual violence during war [Publisher: SAGE Publications Inc]. *Politics & Society*, *34*(3), 307–342.