For peer review in Journal of Peace Research

What Drives Attitudes towards the Reintegration of Former Fighters? Insights from a Conjoint Experiment in Nigeria

Online Appendix

Contents

1 Survey Information[1]

1.1 A Three-Stage Sampling Process

In 2014, [redacted] set up a large-scale longitudinal panel study with Nigerian university students in order to evaluate the extent to which the National Youth Service Corps (NYSC) contributed to improved intergroup relations and stronger national identities [redacted]. When this project was finished, the panel was updated and all original respondents were contacted again to participate in a new project unraveling public opinion concerning the Boko Haram (BH) insurgency. In the remainder of this Appendix, the original evaluation project is referred to as "NYSC survey", while the data in the main manuscript comes from the "BH Survey". In what follows, we explain the three-stage sampling mechanism used to sample the respondents and outline how we updated the panel in 2017. Our sampling procedure largely builds on the insights gained by Langer and colleagues (2017) in Ghana.

Although the importance and application of large-scale surveys in developing countries is growing, contextual factors continue to confront survey researchers with a series of pressing methodological and practical challenges (Bulmer, 1993; Bulmer and Warwick, 1993). One particular challenge to recruit respondents for the initial NYSC survey was the absence of a reliable sampling frame. Both the Nigerian universities and the NYSC secretariat informed us that they could only provide the names of their students or prospective participants, respectively, but not their contact details. We therefore implemented a multistage sampling mechanism. In the first stage (see Figure A1), students from a diverse array of randomly selected Bachelor programs in a number of purposefully selected universities were recruited by means of an off-line paper-administered self-interview (PASI). Five federal universities were selected to be included in study: the University of Nigeria (UNN) in Nsukka, University of Abuja, University of Lagos, Usmanu Danfodiyo University (UDU) in Sokoto, and University of Port-Harcourt. These universities guaranteed an ethno-regional and -religious mix of students while simultaneously took into account safety precautions. We did not visit a North-Eastern university due to security concerns. Official permission to survey the students was obtained from the central University Administration and Heads of Department of all selected Bachelor programs. The classes were visited in May-September 2015, and the in-class survey of 12–15 minutes was usually conducted at the beginning or ending of a lecture. The main goals

[1] This section hugely benefited from the fieldwork conducted by and doctoral thesis of [redacted].

of this in-class survey were to (1) obtain email addresses and phone numbers to create an online panel and (2) build trust as to the sincerity, relevance, and political independence of our project. A total of 6,830 students across the five universities completed this in-class questionnaire.

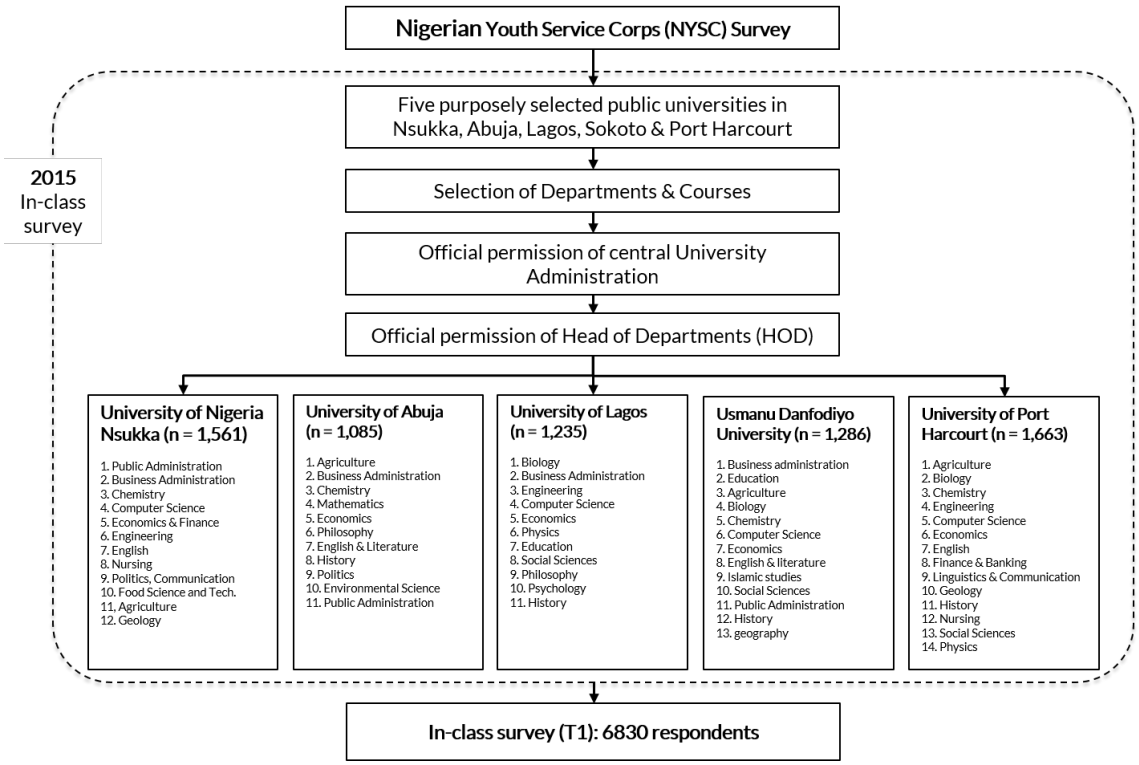

Figure A1. First Stage of the Sampling Process.

In the second stage (see Figure A2), using the contact information respondents provided during the in-class survey, the students were invited to participate in a web self-administered questionnaire (WSAQ; Callegaro et al., 2015). There were a several reasons why we chose to employ a WSAQ. First, the target population was widely dispersed and highly mobile, and a WSAQ allowed us to keep track of our participants. Second, given the relatively sensitive nature of various topics addressed in the survey (e.g., inter-ethnic and -religious relations, corruption, political violence, ...), an interviewer-mediated approach could increase social desirability biases (du Toit, 2016). Third, the cost effectiveness of WSAQ surveys allowed for a larger sample size within the available budget. A convenient byproduct is that a WSAQ enabled us to conduct the conjoint experiment in the 2018 wave, given that Qualtrics randomized the conjoint attributes by recalling a PHP script hosted on a server and piping the attributes back into an HTML-created 2*7-table.

The first online survey took place before the students went for national service (Oct. 2015–Febr. 2016), the second one after they finished their service (April or Aug.–Sept. 2017, depending on the batch in which the students were sent out for service). Students were invited to participate in the online surveys by email, while text messages were also sent to respondents' mobile phones informing them about the invitation and incentive they receive after participation (see Section 1.2.1 as well). Via text messages, 921 missing or wrong e-mail addresses were retrieved. Reminders to participate were sent after one, two, and three weeks again both by email and text message. A total of 2,585 students completed the first online questionnaire (plus 720 partial completions), while 2,302 respondents competed the second online questionnaire (plus 135 partial completions).

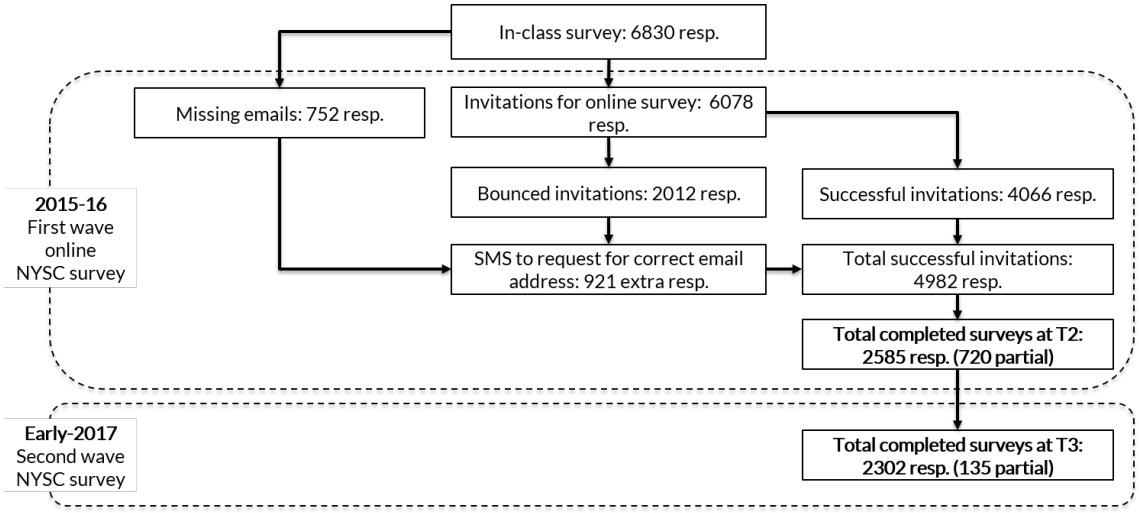

Figure A2. Second Stage of the Sampling Process.

In 2017, the project evaluating the NYSC terminated, while another research project on the impact of the Boko Haram violence commenced. Given the unique access we gained in the years before, we decided to update all contact information and to re-contact the original sample. We had three sources of information to update our contact entries. Each year, after completion of the study, participants were asked to leave their phone number in order to receive their phone top-ups. They could also leave their (new) e-mail address at the end of the survey. Last, several respondents contacted us throughout the years to inform us that they had a new e-mail address or phone number. This strategy left us with 6,199 e-mail addresses that were re-contacted in November 2017. A total of 2,492 students completed the 2017 survey (plus 92 partial completions), while 2,124 students completed the 2018 survey (plus 31 partial completions).

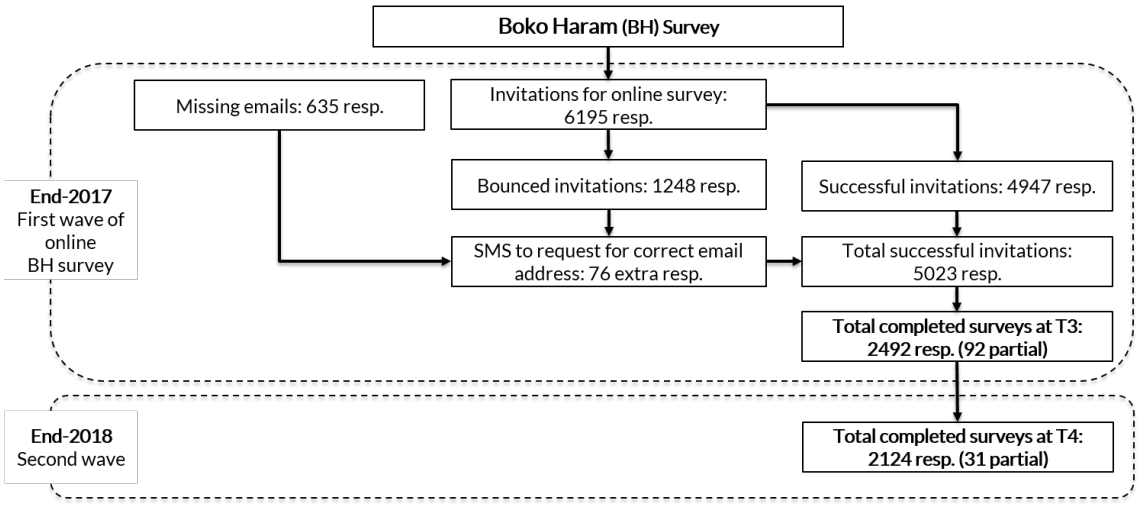

Figure A3. Third Stage of the Sampling Process.

1.2 Response and Attrition Rates

1.2.1 Maximizing Response Rates

To maximize the response rate and minimize the attrition rate, we followed several recommendations as set out in the Tailored Design Method (TDM; Dillman, 1978; Dillman et al., 2014) and by Langer, Meuleman, and colleagues (Langer et al., 2017; Meuleman et al., 2018). First, we sent invitations and reminders via both e-mail and text messages. Using text messaging proved to be a crucial component to reach the target population as several e-mail addresses bounced back the first time. Second, reminders were sent after one, two, and three weeks (again via e-mail and text messages). The survey usually closed after 28 days. Third, time investment was minimized by keeping the survey as short as possible, while a clear and minimal layout was used to facilitate comprehension of the question. This latter strategy was also important to minimize the time needed to load a survey page as bandwidth in Nigeria is more limited. Fourth, for each online survey, respondents were offered a monetary incentive conditional on completion of the survey. Because direct cash incentives were not feasible, we gave respondents a mobile top-up of 1,500 naira (i.e., about 4$). Our preparatory fieldwork confirmed that mobile phone credits function as an easily transmittable quasi-cash incentive in Nigeria, given that our students informed us that mobile phones credited via a pay-as-you-go system are widely used. Mobile top-ups were sent using the electronic payment platform from E-tranzact, which allowed us to distribute top-ups to all phone providers in Nigeria.

1.2.2 Response and Attrition Rates

In what follows, several response rates are calculated using AAPOR guidelines (AAPOR, 2016). Almost 40% of the baseline in-class sample completed the first online survey (i.e., the minimum response rate or RR1 = $\frac{2,585}{6,830} = 37.8\%$). This number gives a rather pessimistic estimation, however, since we encountered a relatively a higher number of bounced emails and failed text messages. When taking non-contact into account, the so-called minimum cooperation rate (COOP1) constitutes 51.9% (i.e. 2,585 complete responses divided by all 4,982 eligible units ever contacted). When also considering partial responses, the cooperation rate further increases to 66.8% (i.e., 3,305 fully and partially completed responses divided by 4,982 contacted students). The completion rate for the first online survey is 78.2% (i.e., $\frac{2,585}{3,305}$). Table A1 checks whether some respondents were significantly more likely to participate in the online survey by comparing the minimum cooperation rates by gender, religion, and university for the first online survey. As one can see, men, Christians, and students from the University of Lagos were more likely to participate in the online survey (probably because of better access to the internet).

Table A1. Cooperation Rates by Gender, Religion, and University for First Online Survey.

	No/Partial response		Complete response		χ^2-test
Total (N=4,982)	2,397	48.1%	2,585	51.9%	
Gender					$\chi^2(1) = 27.076$, $p < .001$
Male	1,237	44.8%	1,527	55.2%	
Female	1,148	52.2%	1,052	47.8%	
Religion					$\chi^2(2) = 8.805$, $p = .012$
Christian	1,796	47.5%	2,010	52.8 %	
Muslim	504	52.5%	456	47.5%	
Other/No	61	50.0%	61	50.0%	
University					$\chi^2(4) = 78.638$, $p < .001$
Nsukka	568	46.7%	647	53.3%	
Abuja	415	53.2%	365	46.8%	
Lagos	378	37.1%	640	62.9%	
Sokoto	445	55.5%	357	44.5%	
Port-Harcourt	591	50.6%	576	49.4%	

After we updated and re-contacted the sample, 40.2% (i.e., 2,492 surveys completed divided by 6,195 invited students) respondents fully completed the 2017 online survey, which increases to 49.6% if you take into account the number of bounced emails (i.e., 2,492 complete responses divided by 5,023 eligible units contacted) and to 51.5 % if you also consider the partial completions (i.e., 2,584 fully and partially completed responses divided by 5,023 contacted students). The completion rates in these last two surveys are remarkably high (i.e., 96.3% and 99%), which suggests that respondents are quite interested in the topic of this new project. Table A2 summarizes the number of complete and partial responses in all survey waves. Last, on average, the item non-response across survey waves for all questions was 2.5%, which is similar to results from Western web-based surveys (Lesser et al., 2012; Millar and Dillman, 2012).

Table A2. Overview of Sample Size Across Survey Waves.

	Offline	NYSC Online Survey		BH Online Survey	
	Baseline	($N = 4,982$ contacts)		($N = 5,022$ contacts)	
	2015	2015-16	Early-2017	End-2017	End-2018
All responses	6,630	3,305	2,437	2,584	2,155
Complete responses	6,630	2,585	2,302	2,492	2,124
Partial responses	NA	720	135	92	31

1.3 Survey Items

Table A3. Question Wording, Response Options, and Additional Information for Descriptives and Moderators (Based on 2018 BH Survey).

Religion	
Question wording	What is your religion? I am
Response options	• christian • muslim • traditional-African religious • not religious • other religion, please specify: [open text box]
Add. information	Traditional/other/no religion are aggregated in Table 1. For the moderation analyses, the 23 respondents in this bulk category were excluded.
Place of origin	
Question wording	In what state were you born?
Response options	List of all 36 states and the Federal Capital Territory.
Add. information	This variable was recoded into geo-political zones: • North-Central • North-East • North-West • South-East • South-South • South-West • Not in Nigeria/missing For the moderation analyses, these categories were further combined into people born in the North (=1) and those born in the South (=0).
Perceived victimization	
Question wording	In general, to what extent do you consider yourself a victim of the Boko Haram crisis?
Response options	• Not at all • Very little • Somewhat • A great deal

Add information	For the moderation analyses, this variable was recoded into respondents who do not at all consider themselves a victim (=0) and those who do, to some extent, consider themselves a victim (=1).[2]
Worry of Boko Haram violence	
Question wording (1)	In general, how worried are you about another large-scale attack in Nigeria by Boko Haram in the near future?
Question wording (2)	How worried are you about becoming victim of a violent attack by Boko Haram?
Question wording (3)	How worried are you about a family member or friend becoming victim of a violent attack by Boko Haram?
Response options	0 to 10 Likert scale
Add information	These items form a composite scale with sufficient reliability ($\alpha = .86$). Factor score based on the mean reported in Table 1. For the moderation analyses, a composite variable (using Bartlett scores; DiStefano, Zhu & Mîndrilă, 2009) was dichotomized based on the mean value into respondents with low levels (=0) and high levels (=1) of concern for Boko Haram violence.

[2] In the main analyses reported in the paper, we use these *perceptions* of victimhood, as individuals tend to react to their perception of situations and facts, rather than to situations and facts as such (Thomas and Thomas, 1929). In other words, we assume that perceiving yourself a victim is sufficient to shape attitudes and behavior. Nevertheless, the results are robust to using a more objective, self-reported exposure measure (see Appendix §4).

1.4 Research Ethics

Post-conflict settings are challenging contexts in which to conduct research, and raise unique ethical challenges due to both the high stakes of peace and the sensitivity of subjects (Matanock, 2021). Therefore, several measures were taken to ensure the safety of both respondents and researchers:

- Official permission to survey the students was obtained from the central University Administration and all Heads of Department responsible for the selected bachelor programs. Our fieldwork scheme and strategy was also approved by the Department Head of our university, but we did not get permission to visit a North-Eastern university due to safety concerns.
- All respondents were recruited in class where we introduced the research team and institution, extensively explained the purpose and course of our research project, and gave students the opportunity to ask questions and express concerns (which they did).
- For all subsequent online surveys, participants electronically signed a standard informed consent form. Here, they were informed that their participation was voluntary, they could exit the survey at any time, and that all information would be treated with full confidentiality.
- To guarantee confidentiality, we separated people's answers to our questions from identifying information (such as email addresses and phone numbers), and all data is currently stored on an encrypted server provided by our university. To ensure the replicability and transparency of our findings, a shorter and fully anonymized dataset will be made publicly available.
- Before seeing any profile, the experiment was was introduced with the following sentence: "*In what follows, we will show you two profiles of hypothetical Boko Haram fighters.* [...]", aimed at minimizing deception. Furthermore, we also stressed that all profiles pertained to *former* fighters (see Figure 2A in the main paper) to address the concern that respondents might assume the profiles relate to people still involved in the violent insurgency.
- Qualitative evidence (through cognitive interviews after the pilot study and emails received after participation) also suggests that many of our respondents were eager to tell their story and to report their experiences with and opinions about the Boko Haram insurgency. In a similar vein, item non-response is not significantly higher for the terrorism-related questions and the overall completion rate is remarkably high as well (99%).
- Respondents were compensated for their participation via a mobile top-up of 1,500₦($\pm$4$).

In sum, this study has been conducted conform APSA Principles of Ethics for Human Subjects Research (2019) or the guidelines prescribed by a typical Institutional Review Board (IRB).[3]

[3] At the time we embarked on the research project on which this paper draws, our university did not require formal

2 Quantities of Interest

2.1 Interpretation[4]

The *average treatment effect* (ATE) is probably one of the most common quantity of interest within experimental designs (Gerber and Green, 2012) and, in the context of conjoint experiments, represents the difference in responses (or cell means, $E[Y \mid \overrightarrow{\mathrm{X}} = \overrightarrow{\mathrm{x}}]$) for two different sets of profiles. Estimating this quantity is an unfeasible task considering that the number of observations in a typical conjoint experiment is much lower the number of possible profiles ($N < C$). In other words, the means for many—and often for most—experimental cells remain unobserved. In our experiment, assuming full randomization, there are about 43,200 possible profiles while we have 11,904 rated profiles clustered within 1,984 respondents, so more than 70 percent of all experimental cells remain unobserved ($N \ll C$). Fortunately, the main strength of a conjoint experiment is that it enables estimating an alternative quantity of interest, being the *average marginal component effect* (AMCE; Hainmueller et al., 2014). AMCEs quantify the overall effect of an individual treatment component (or feature; F) rather than the effect of an entire set of components given that, in most conjoint experiments, $N > F$. Formally, AMCEs are defined as the marginal effect of a certain attribute averaged over the joint distribution of the remaining attributes (or, in other words, differences in conditional/marginal means, $E[Y \mid \overrightarrow{\mathrm{X}_{\mathrm{f}}} = \overrightarrow{\mathrm{x}_{\mathrm{f}}}]$). While a comparison of marginal means allows *causal* interpretation of conjoint experiments, Leeper and colleagues (2020) shown how the marginal means alone are of direct interest for the common *descriptive* use of conjoint designs. *Marginal means* (MMs) are defined as the "favorability on the scale of the outcome over alternative values of each feature independent of the features in the design" (Leeper et al., 2020, SI p.5). As we explain in the main article, the difference between marginal means (MMs) and a comparison of marginal means (AMCEs) becomes especially important when estimating heterogeneous treatment effects, which is also a quantity of interest in our study.

Noteworthy, Abramson, Kocak, and Magazinnik (2019) recently demonstrate how conjoint experiments (and more specifically, AMCEs and MMs) do not capture the underlying majority preference, but they are often interpreted as doing so. In other words, conjoints do not reflect preferences in the strict economic sense of the term. This is not a consequence of uncertainty or bias, but is inherent to the AMCEs properties as an aggregation mechanism. More specifically, AMCEs and, by extension, MMs are influenced by both preference *direction* (i.e., what feature respondents

ethical approval for these types of projects. As a result, this project does not have an official IRB number.

[4] This section hugely benefited from and applies the notation used in the Supplementary Information provided by Thomas Leeper, Sara Hobbolt, and James Tilley in their 2020 Political Analysis article.

prefer) and preference *intensity* (i.e., how much weight respondents give to that feature's attribute relative to the other attributes). For example, imagine that a majority of voters has a *weak* preference for a male candidate, whereas a minority of voters has a *strong* preference for a female candidate. This minority will therefore always prefer a female candidate over a male candidate (e.g., even when a female candidate does not align with the respondent's own party identity), while the majority will sometimes vote for a man but might equally vote for a women with other good attributes (e.g., when a female candidate aligns with the respondent's own party identity and party identity has given more weight than gender). Since this minority preference gets weighted more, the estimands might seem to favor women (i.e., the AMCE for women is positive of the MMs for woman in higher than for men) while the majority will still vote for a male candidate. In short, the larger the correlation between direction and intensity, the more misleading AMCEs and MMs will be (Abramson et al., 2019).

As a result, a more accurate interpretation of the quantities of interest implies that conjoint experiments capture *revealed* preferences or observed choices rather than *underlying* or unobserved majority preferences. As an example, MMs are thus more correctly interpreted as "X% of the time respondents chose feature A (e.g., women) compared to choosing feature B (e.g. men) only (1-X)% of the time" rather than "X% of the respondents prefer feature A (e.g., women)". In addition, while AMCEs and MMs might therefore be less appropriate to make claims about voters' preferences or election outcomes (given that elections are won by the median voter), the intensity of preferences can still be relevant or even the outcome of interest for other research questions—especially when assuming that higher intensity is more likely to translate into actual behavior. This, we argue, is an important point to consider in our study because, even when a small minority strongly opposes reintegration because of one or several feature(s) of former fighters, this might still cause a serious (non-violent or violent) public backlash.

2.2 Estimation: Main Model

Following best practices, we prohibited profiles that gave raise to counterfactuals that were impossible or too unrealistic to be evaluated in a meaningful way in order to increase external validity and decrease skewed results. Specifically, we imposed three restrictions. First, those fighters that were born within Boko Haram were constrained to be forced to join Boko Haram given that they did not deliberately choose to join Boko Haram. Second, a female fighter could have coordinated attacks and conducted killings, bombings, and kidnappings while with Boko Haram, but not rape. Last, fighters that mainly targeted the police, military, or government officials were also constrained to

have conducted any atrocity, but rape. By applying these restrictions, some cells were by design unobservable. Leeper and colleagues (2020) argue for more care when defining and estimating AMCEs in these commonly used, constrained designs. More specifically, constrained designs require more transparency about what features are being marginalized over when estimating AMCEs or, preferably, an analysis of only the complete and comparable subset of the design (i.e. the fractions of the design where all cells are observed). When attributes are constrained, one compares dissimilar subsets of profiles along these restricted dimensions. For instance, our quantities of interest for Boko Haram fighters who joined at age 30 are averaged over all possible reasons to join, while the estimates for those who were born within Boko Haram are only average over those forced to join because all other reasons were restricted by design. To robustly and transparently estimate our results, we calculate three quantities of interest (and report discrepancies in the main paper):

1. over the levels of the feature of interest only (i.e., the results reported in the main paper):

$$Y_{\text{Model 1}} = \beta_0 + \beta_1 X_1 + \epsilon,$$

with Y respondents' reintegration attitudes and X_1 the different levels of the first feature of ex-fighters. Thus, the equation to get the AMCE for the feature *gender*, which has two levels (with male fighters as the reference category), becomes:

$$Y = \beta_0 + \beta_1 X_{Female} + \epsilon$$

2. over subsets of the design that are conditionally unconstrained. More specifically, we partitioned the data to form two complete and overlapping experimental designs:

   (a) Subset 1 excluding child soldiers born within Boko Haram and fighters who have mainly conducted rape. The equation to estimate the Marginal Means thus becomes:

$$\begin{aligned} Y_{\text{Model 2}} = \beta_0 &+ \beta_{1.1}\text{Male} + \beta_{1.2}\text{Female} \\ &+ \beta_{2.1}\text{10y} + \beta_{2.2}\text{15y} + \beta_{2.3}\text{20y} + \beta_{2.4}\text{25y} + \beta_{2.5}\text{30y} \\ &+ \beta_{3.1}\text{Forced} + \beta_{3.2}\text{Poverty} + \beta_{3.3}\text{Livelihood} + \beta_{3.4}\text{Corruption} + \beta_{3.5}\text{Caliphate} \\ &+ \beta_{4.1}\text{Kidnappings} + \beta_{4.2}\text{Coordination} + \beta_{4.3}\text{Bombings} + \beta_{4.4}\text{Killings} \\ &+ \beta_{5.1}\text{Christians} + \beta_{5.2}\text{Muslims} + \beta_{5.3}\text{Civilians} + \beta_{5.4}\text{Government} + \beta_{5.5}\text{Military} + \beta_{5.6}\text{Police} \\ &+ \beta_{6.1}\text{Remorse} + \beta_{6.2}\text{Disappointment} + \beta_{6.3}\text{Uncertainty} + \beta_{6.4}\text{Injured} + \beta_{6.5}\text{Captured} \\ &+ \beta_{7.1}\text{Help} + \beta_{7.2}\text{Forgiveness} + \beta_{7.3}\text{Apologies} + \beta_{7.4}\text{Deradicalization} + \beta_{7.5}\text{Paid} + \beta_{7.5}\text{Nothing} + \epsilon \end{aligned}$$

(b) Subset 2 excluding female fighters and fighters who mainly targeted government officials, military and police personnel:

$$
\begin{aligned}
Y_{\text{Model 3}} = \beta_0 &+ \beta_{1.1}\text{Male} \\
&+ \beta_{2.0}\text{0y} + \beta_{2.1}\text{10y} + \beta_{2.2}\text{15y} + \beta_{2.3}\text{20y} + \beta_{2.4}\text{25y} + \beta_{2.5}\text{30y} \\
&+ \beta_{3.1}\text{Forced} + \beta_{3.2}\text{Poverty} + \beta_{3.3}\text{Livelihood} + \beta_{3.4}\text{Corruption} + \beta_{3.5}\text{Caliphate} \\
&+ \beta_{4.1}\text{Kidnappings} + \beta_{4.2}\text{Coordination} + \beta_{4.3}\text{Bombings} + \beta_{4.4}\text{Killings} + \beta_{4.5}\text{Rape} \\
&+ \beta_{5.1}\text{Christians} + \beta_{5.2}\text{Muslims} + \beta_{5.3}\text{Civilians} \\
&+ \beta_{6.1}\text{Remorse} + \beta_{6.2}\text{Disappointment} + \beta_{6.3}\text{Uncertainty} + \beta_{6.4}\text{Injured} + \beta_{6.5}\text{Captured} \\
&+ \beta_{7.1}\text{Help} + \beta_{7.2}\text{Forgiveness} + \beta_{7.3}\text{Apologies} + \beta_{7.4}\text{Deradicalization} + \beta_{7.5}\text{Paid} + \beta_{7.5}\text{Nothing} \\
&+ \epsilon
\end{aligned}
$$

3. over all feature levels with the explicit caveat that the comparison happens across dissimilar subsets of profiles for the attributes Gender, Age, Joining, Atrocity, and Victim.

$$Y_{\text{Model 4}} = \alpha + \beta_1 X_1 + \beta_2 X_2 + ...\beta_\eta X_\eta + \epsilon,$$

with X_1, X_2, and X_η all the different levels of all seven ex-fighter features. Here, the regression model also included the appropriate interaction terms to account for the dependency across certain features. By including interaction terms between conditionally dependent attributes, we obtain unbiased estimates (see Hainmueller et al., 2014 for more details).

As a result, Model 1 and 4 in the Tables A4 (AMCEs) and A5 (MMs) below are based on all 11,940 observations derived from all 1,984 respondents. However, Model 2 draws on 10,779 observations from the 1,984 respondents after omitting the feature levels "Age==0 year" and "Atrocity==Rape" and Model 3 on a substantially reduced sample of 3,558 observations from 1,732 respondents after omitting the feature levels "Gender==Female" and "Victim=Government Officials—Police Personnel—Military Personnel". Furthermore, the feature Gender is removed entirely when estimating the AMCEs based Subset 2.

Table A4. Average Marginal Component Effects (AMCEs) for Forced-Choice Design.

Attribute Level	Model 1	Model 2	Model 3	Model 4
Gender				
Male	NA	NA	NA	NA
Female	0.0267 (0.009)**	0.0208 (0.009)*	NA	0.0172 (0.009)
Age				
30 years	NA	NA	NA	NA
25 years	0.0183 (0.015)	0.0170 (0.015)	0.0235 (0.026)	0.0163 (0.015)
20 years	0.0353 (0.015)*	0.0434 (0.015)*	0.0125 (0.026)	0.0395 (0.015)*
15 years	0.0757 (0.015)***	0.0718 (0.015)***	0.0766 (0.026)***	0.0701 (0.015)***
10 years	0.1384 (0.015)***	0.1458 (0.015)***	0.1459 (0.026)***	0.1416 (0.015)***
0 year	0.1685 (0.028)***	NA	0.1131 (0.061)$^{\cdot 06}$	0.0963 (0.033)**
Joining				
Establishing a Caliphate	NA	NA	NA	NA
Corruption	0.0607 (0.015)***	0.0635 (0.016)***	0.0330 (0.026)	0.0660 (0.015)***
Securing a livelihood	0.0887 (0.015)***	0.0927 (0.015)***	0.1002 (0.026)***	0.0922 (0.015)***
Poverty in the North	0.1224 (0.015)***	0.1218 (0.015)***	0.1010 (0.027)***	0.1229 (0.015)***
Forced to join	0.2078 (0.015)***	0.2148 (0.015)***	0.2158 (0.027)***	0.2160 (0.015)***
Atrocity				
Rape	NA	NA	NA	NA
Killings	0.0107 (0.021)	NA	-0.0048 (0.026)	-0.0140 (0.028)
Bombings	0.0224 (0.021)	0.0107 (0.013)	-0.0008 (0.026)	-0.0038 (0.028)
Coordinating	0.0512 (0.021)*	0.0425 (0.013)**	0.0385 (0.026)	0.0284 (0.028)
Kidnappings	0.0692 (0.021)**	0.0656 (0.014)***	0.0600 (0.026)*	0.0594 (0.023)**
Victim				
Christians	NA	NA	NA	NA
Muslims	0.0281 (0.015)	0.0295 (0.016)*	0.0277 (0.020)	0.0248 (0.016)
Civilians	0.0388 (0.016)*	0.0303 (0.016)	0.0398 (0.020)*	0.0346 (0.016)*
Police personnel	0.0464 (0.016)**	0.0331 (0.016)*	NA	0.0400 (0.016)*
Military personnel	0.0634 (0.016)***	0.0577 (0.017)***	NA	0.0606 (0.016)***
Government officials	0.0725 (0.016)***	0.0656 (0.017)***	NA	0.0675 (0.016)***
Leaving				
Captured by the military	NA	NA	NA	NA
Injured and hospitalized	0.0180 (0.014)	0.0230 (0.015)	0.0198 (0.026)	0.0177 (0.014)
Uncertain about survival	0.0435 (0.015)**	0.0416 (0.015)**	0.0226 (0.026)	0.0405 (0.014)**
Disappointed in ideology	0.1698 (0.014)***	0.1733 (0.015)***	0.1642 (0.026)***	0.1673 (0.014)***
Remorse of violence	0.2087 (0.014)***	0.2157 (0.015)***	0.2245 (0.027)***	0.2088 (0.014)***
Reconciliation				
No conciliatory actions	NA	NA	NA	NA
Paid compensation	0.1189 (0.016)***	0.1277 (0.016)***	0.1393 (0.028)***	0.1263 (0.016)***
Deradicalization programme	0.1394 (0.016)***	0.1273 (0.016)***	0.1649 (0.028)***	0.1315 (0.016)***
Offered apologies	0.1442 (0.016)***	0.1409 (0.017)***	0.1592 (0.027)***	0.1448 (0.016)***
Asked for forgiveness	0.1434 (0.015)***	0.1405 (0.016)***	0.1870 (0.027)***	0.1466 (0.015)***
Helped police and military	0.2390 (0.016)***	0.2405 (0.017)***	0.2720 (0.027)***	0.2369 (0.016)***

Note. Clustered standard errors are in parentheses. Models 1-4 average over different joint distributions of feature levels (see Section 2.2 above). Model 1 reported in the main manuscript. $^{*}p < 0.05$. ** $p < 0.01$. $^{***}p < 0.001$.

Table A5. Marginal Means (MMs) for Forced-Choice Design.

Attribute Level	Model 1	Model 2	Model 3	Model 4
Gender				
Male	[0.4791, 0.4960]	[0.4758, 0.4956]	NA	[0.4791, 0.4960]
Female	[0.5046, 0.5238]	[0.5019, 0.5220]	NA	[0.5046, 0.5238]
Age				
30 years	[0.4248, 0.4617]	[0.4247, 0.4627]	[0.3715, 0.4420]	[0.4248, 0.4617]
25 years	[0.4434, 0.4796]	[0.4436, 0.4814]	[0.4075, 0.4790]	[0.4434, 0.4796]
20 years	[0.4599, 0.4971]	[0.4623, 0.5009]	[0.3989, 0.4734]	[0.4599, 0.4971]
15 years	[0.5007, 0.5373]	[0.5032, 0.5409]	[0.4582, 0.5314]	[0.5007, 0.5373]
10 years	[0.5635, 0.5998]	[0.5663, 0.6038]	[0.5194, 0.5917]	[0.5635, 0.5998]
0 year	[0.5623, 0.6611]	NA	[0.5242, 0.7111]	[0.5623, 0.6611]
Joining				
Establishing a Caliphate	[0.3784, 0.4191]	[0.3810, 0.4239]	[0.3389, 0.4143]	[0.3784, 0.4191]
Corruption	[0.4409, 0.4780]	[0.4411, 0.4795]	[0.3697, 0.4424]	[0.4409, 0.4780]
Securing a livelihood	[0.4689, 0.5060]	[0.4716, 0.5104]	[0.4441, 0.5172]	[0.4689, 0.5060]
Poverty in the North	[0.5027, 0.5396]	[0.5035, 0.5416]	[0.4518, 0.5260]	[0.5027, 0.5396]
Forced to join	[0.5906, 0.6225]	[0.5923, 0.6289]	[0.5583, 0.6249]	[0.5906, 0.6225]
Atrocity				
Rape	[0.4282, 0.5005]	NA	[0.4282, 0.5005]	[0.4282, 0.5005]
Killings	[0.4588, 0.4912]	[0.4521, 0.4853]	[0.4054, 0.4767]	[0.4588, 0.4912]
Bombings	[0.4706, 0.5029]	[0.4670, 0.5000]	[0.4060, 0.4745]	[0.4706, 0.5029]
Coordinating	[0.4995, 0.5315]	[0.4942, 0.5272]	[0.4584, 0.5294]	[0.4995, 0.5315]
Kidnappings	[0.5168, 0.5504]	[0.5168, 0.5511]	[0.4779, 0.5502]	[0.5168, 0.5504]
Victim				
Christians	[0.4407, 0.4793]	[0.4405, 0.4829]	[0.4150, 0.4691]	[0.4407, 0.4793]
Muslims	[0.4686, 0.5075]	[0.4683, 0.5112]	[0.4437, 0.4986]	[0.4686, 0.5075]
Civilians	[0.4789, 0.5188]	[0.4719, 0.5149]	[0.4729, 0.5288]	[0.4789, 0.5188]
Police personnel	[0.4856, 0.5272]	[0.4762, 0.5188]	NA	[0.4856, 0.5272]
Military personnel	[0.5021, 0.5446]	[0.4976, 0.5408]	NA	[0.5021, 0.5446]
Government officials	[0.5106, 0.5544]	[0.5107, 0.5554]	NA	[0.5106, 0.5544]
Leaving				
Captured by the military	[0.3944, 0.4297]	[0.3877, 0.4248]	[0.3566, 0.4270]	[0.3944, 0.4297]
Injured and hospitalized	[0.4118, 0.4482]	[0.4099, 0.4489]	[0.3731, 0.4420]	[0.4118, 0.4482]
Uncertain about survival	[0.4372, 0.4739]	[0.4307, 0.4698]	[0.3774, 0.4474]	[0.4372, 0.4739]
Disappointed in ideology	[0.5637, 0.5999]	[0.5636, 0.6019]	[0.5111, 0.5816]	[0.5637, 0.5999]
Remorse of violence	[0.6028, 0.6386]	[0.6051, 0.6427]	[0.5640, 0.6396]	[0.6028, 0.6386]
Reconciliation				
No conciliatory actions	[0.3496, 0.3891]	[0.3501, 0.3918]	[0.2831, 0.3549]	[0.3496, 0.3891]
Paid compensation	[0.4679, 0.5087]	[0.4712, 0.5144]	[0.4154, 0.4972]	[0.4679, 0.5087]
Deradicalization programme	[0.4886, 0.5289]	[0.4809, 0.5235]	[0.4520, 0.5323]	[0.4886, 0.5289]
Offered apologies	[0.4934, 0.5337]	[0.4866, 0.5297]	[0.4342, 0.5130]	[0.4934, 0.5337]
Asked for forgiveness	[0.4934, 0.5321]	[0.4881, 0.5294]	[0.4688, 0.5439]	[0.4934, 0.5321]
Helped police and military	[0.5886, 0.6281]	[0.5916, 0.6340]	[0.5521, 0.6295]	[0.5886, 0.6281]

Note. 95% confidence intervals reported. Models 1-4 average over different joint distributions of feature levels (see Section 2.2 above). Model 1 reported in the main manuscript.

2.3 Estimation: Mediator Models

2.3.1 What Drives Assessments of Reintegration Success?

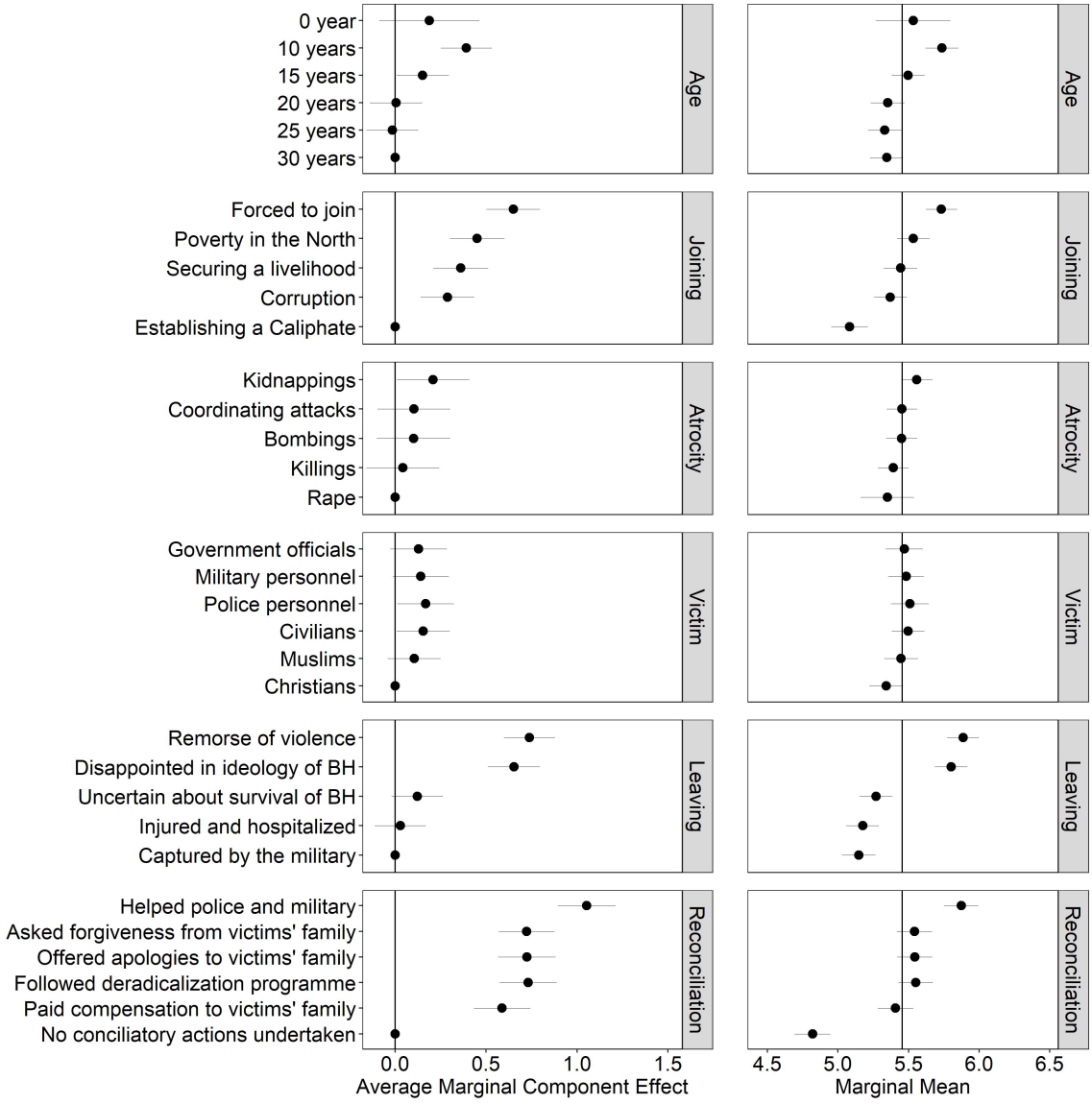

Figure A4. Average Marginal Component Effects (AMCEs) and Marginal Means (MMs) for 'Assessments of Success' Outcome (Paired, No Forced-Choice Design).

2.3.2 What Drives Granting Ex-Fighters Amnesty?

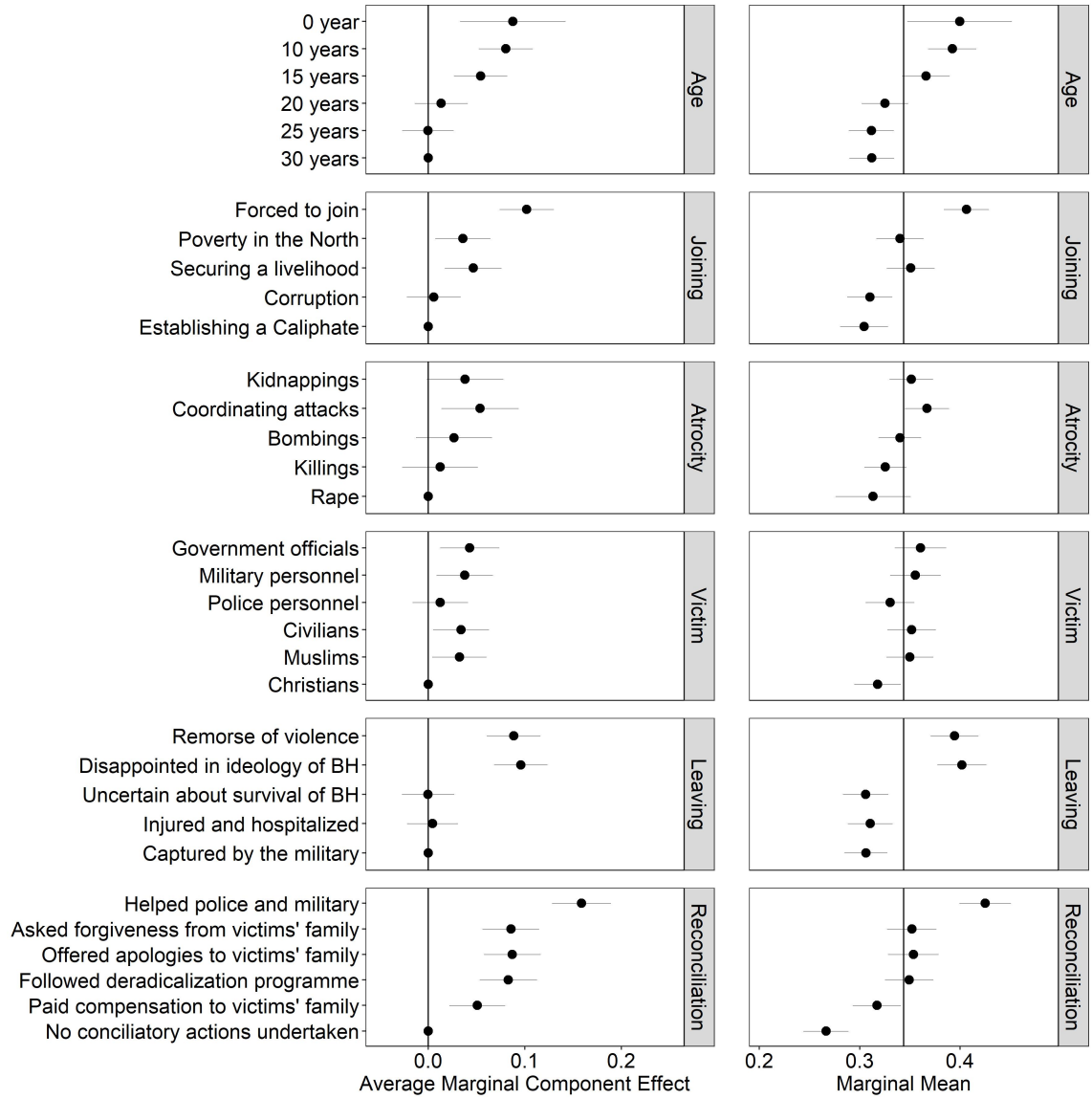

Figure A5. Average Marginal Component Effects (AMCEs) and Marginal Means (MMs) for 'Amnesty' Outcome (Paired, No Forced-Choice Design).

2.3.3 What Drives Sentencing Ex-Fighters to Death?

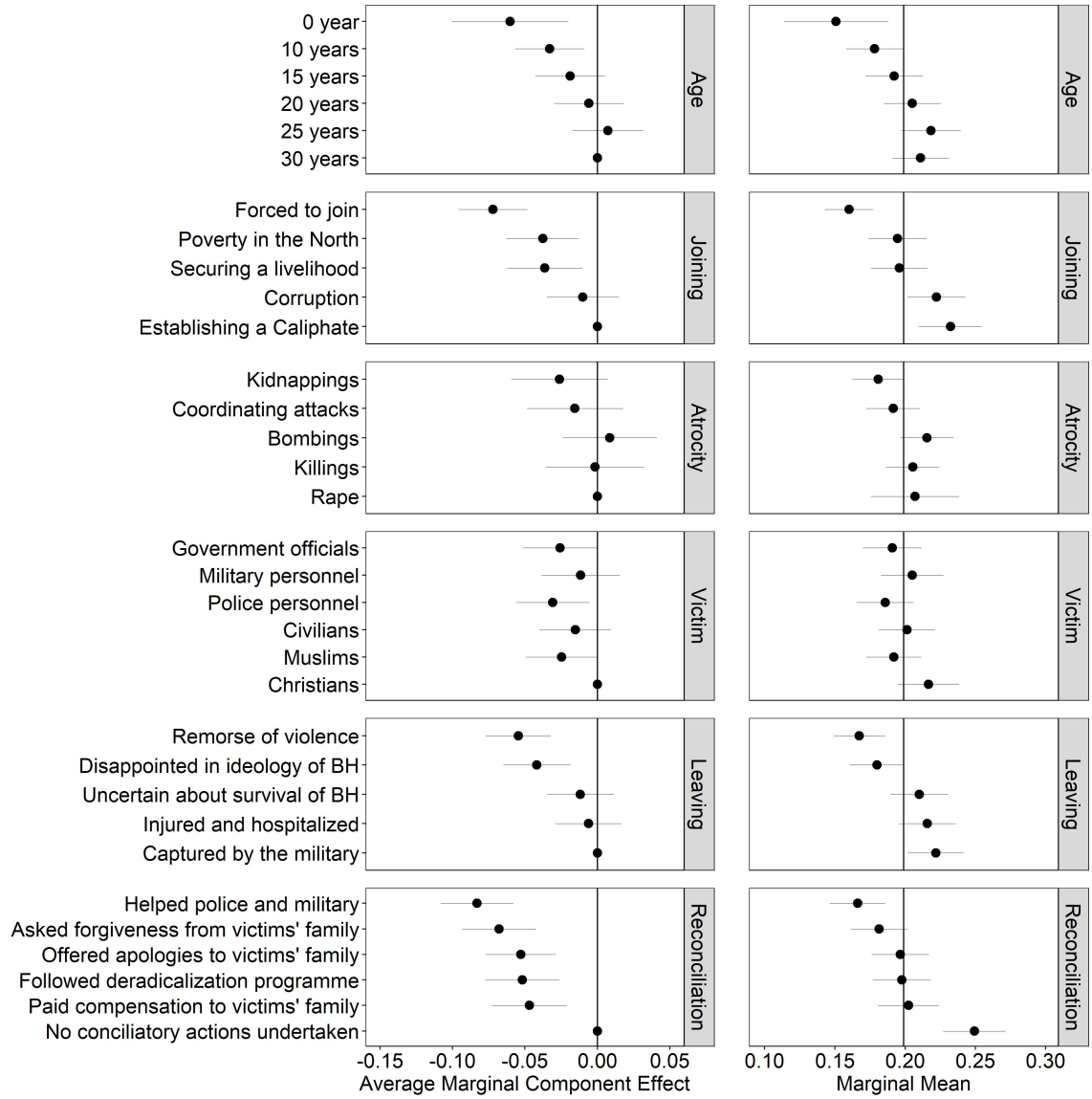

Figure A6. Average Marginal Component Effects (AMCEs) and Marginal Means (MMs) for 'Death Penalty' Outcome (Paired, No Forced-Choice Design).

2.4 Estimation: Interaction Models

The Results section in the main manuscript presents Average Marginal Component Effects (AMCEs) and Marginal Means (MMs). In this Appendix section, we further explore four interactions between sets of treatments. Particularly, given that the entry and exit phase were found to be most consequential in driving attitudes towards reintegration, we explore how they work in combination with one another. In other words, what are the preferences of our respondents vis-a-vis, e.g., an ex-combatant born within the insurgency but then leaving involuntarily or showing no signs of remorse? Figure A7 shows that those ex-fighters born within the insurgency and captured by the military are still more likely to be selected for reintegration compared to their older counterparts. In a similar vein, of all those ex-fighters captured by the military, those who were forced to join are still more likely to be selected (Figure A8). Furthermore, Figure A7 also reveals that leaving out of remorse is beneficial for all former fighters, regardless of the age they joined the insurgency. Lastly and more generally, information about the entry and exit phase display a cumulative effect on the observed selection probabilities. These cumulative effects are quite substantial: Fighters who were forced to join and left out of remorse were most often selected for reintegration (MM = 71.85%), whereas fighters who joined to fight for the Caliphate and undertook no conciliatory action were least often selected (MM = 26,74%). Nevertheless, while there are cumulative effects, the patterns of preferences remain the same in all four figures (A7-A10). That is, younger fighters and those forced to join are always more likely to be selected (regardless of why and how they left the insurgency). By contrast, fighting for the caliphate and showing no conciliatory gestures is always detrimental for public acceptance of reintegration. In sum, the interactive models replicate the conclusions presented in the main manuscript.

2.4.1 Interaction Between Age at Joining and Reason to Leave

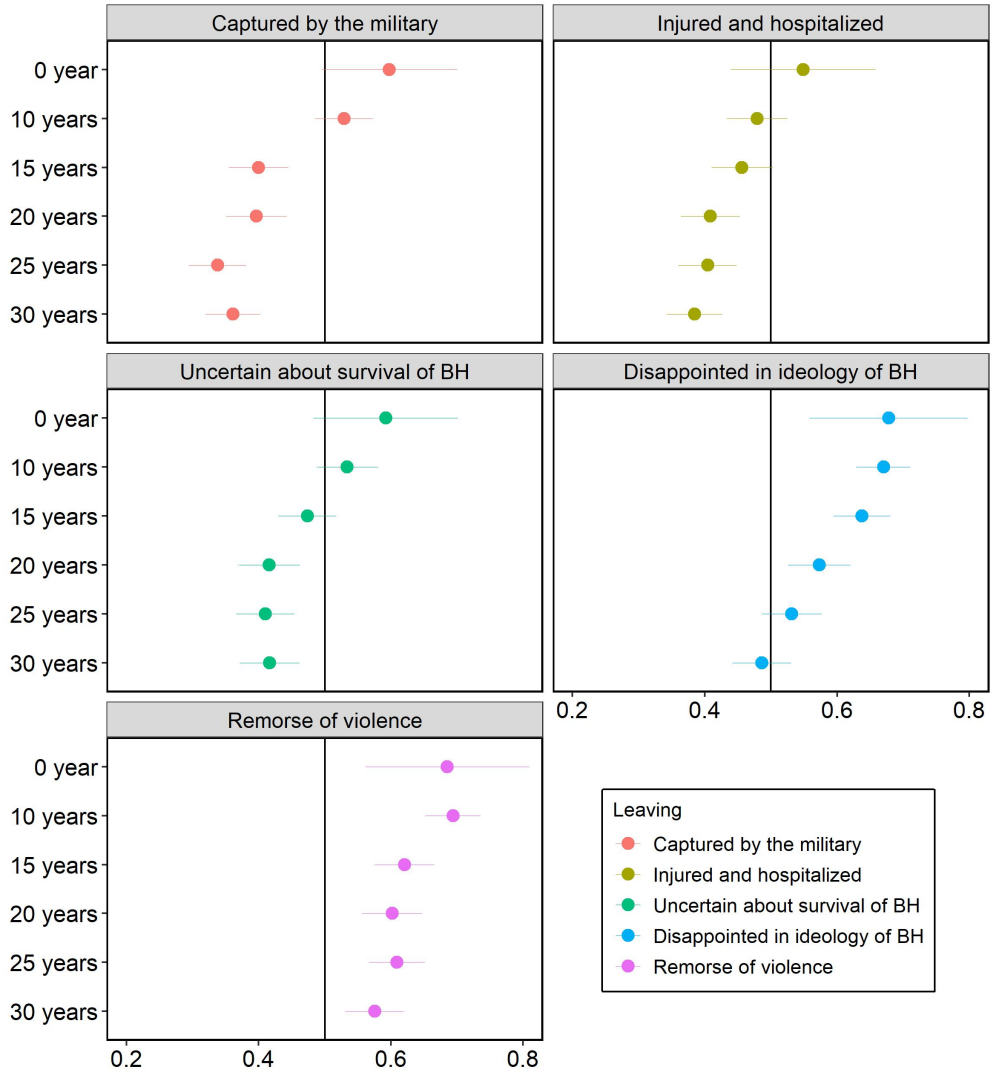

Figure A7. Marginal Means (MMs) for Forced-Choice Design for Age at Joining, by Reason to Leave.

2.4.2 Interaction Between Reason for Joining and Reason to Leave

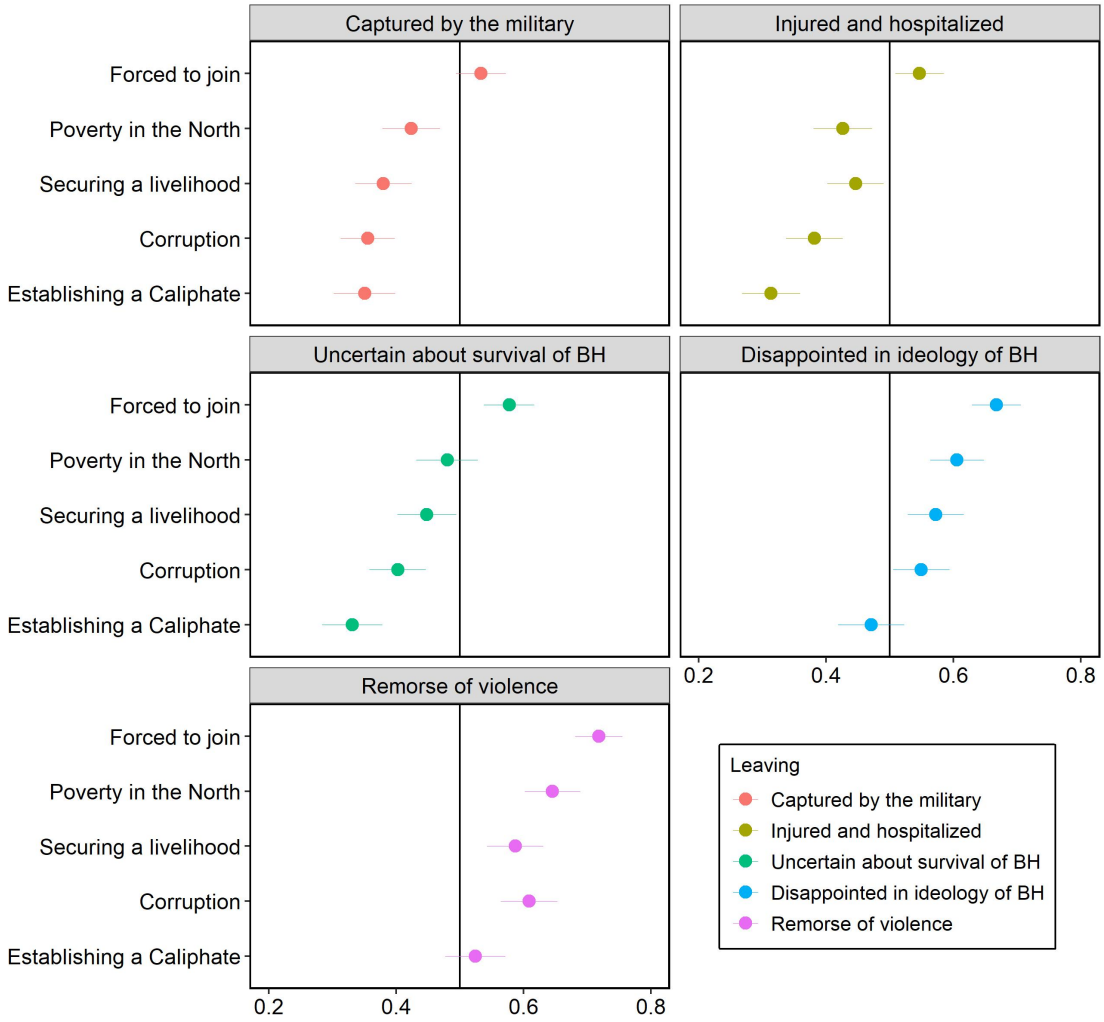

Figure A8. Marginal Means (MMs) for Forced-Choice Design for Reason to Join, by Reason to Leave.

2.4.3 Interaction Between Age at Joining and Conciliatory Gestures

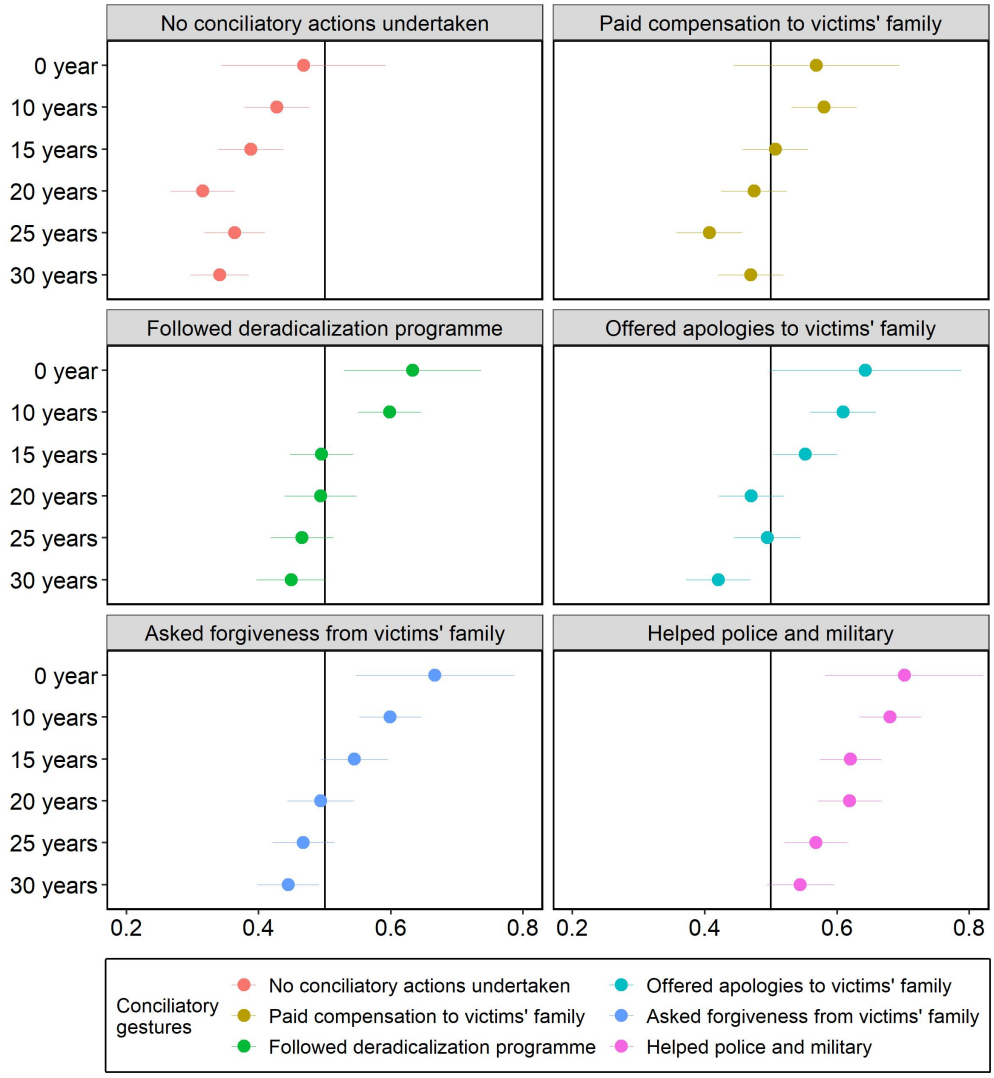

Figure A9. Marginal Means (MMs) for Forced-Choice Design for Age at Joining, by Conciliatory Gestures.

2.4.4 Interaction Between Reason for Joining and Conciliatory Gestures

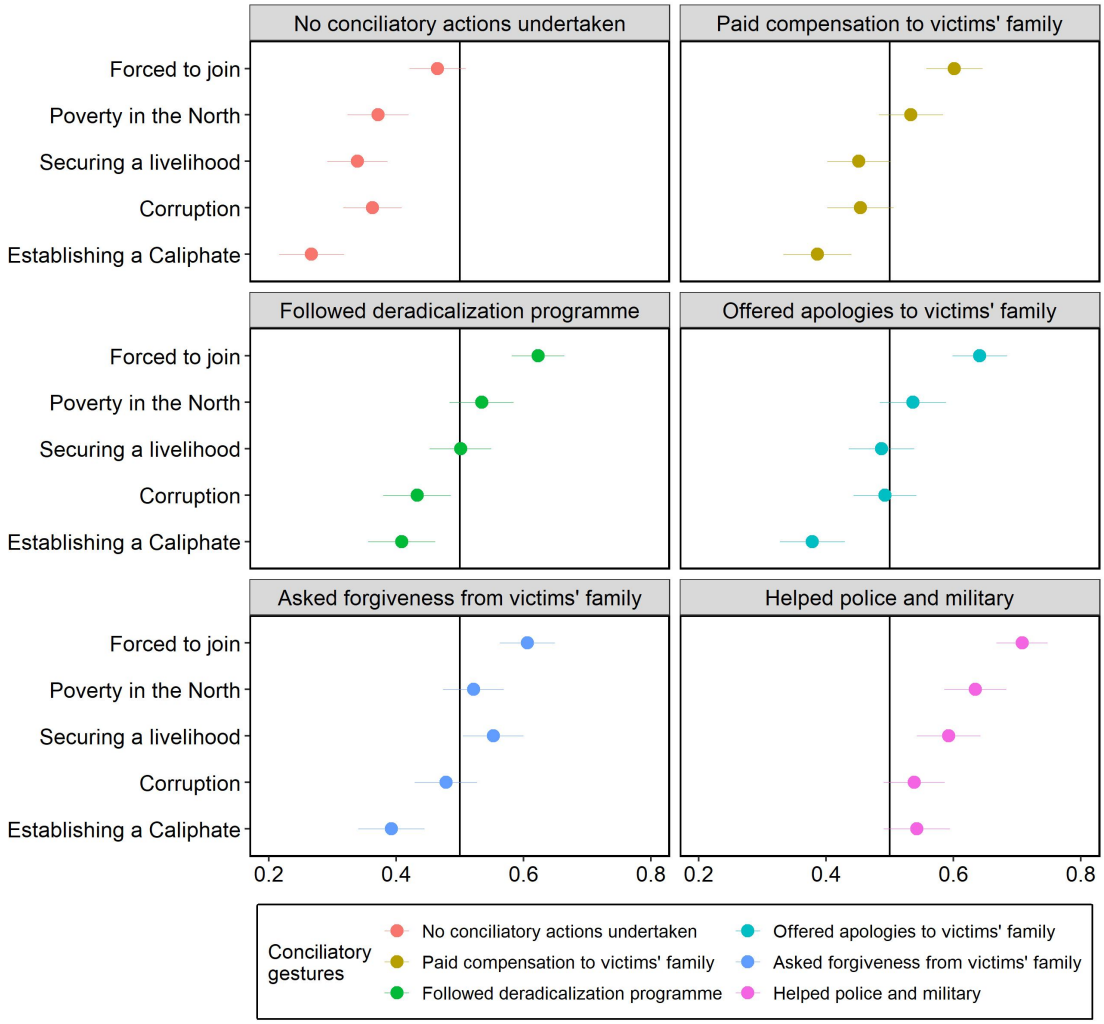

Figure A10. Marginal Means (MMs) for Forced-Choice Design for Reason to Join, by Conciliatory Gestures.

3 Model Diagnostics

3.1 Check of Restrictions and Randomization

Figure A11 confirms that our restrictions worked as intended (see unequal display frequencies) and that non-restricted feature levels were correctly randomized (see equally display frequencies).

Figure A11. Frequencies.

3.2 Balance Tests for Covariates Used

In Figure A12, we test whether the randomization procedure produced experimental groups that are well-balanced across the covariates used in the main article. We test this by regressing respondents' gender, religious background, perceptions of victimization, and levels of concern about Boko Haram violence on the profile features. The confidence intervals hover closely around the grand mean which suggests little to no imbalance.

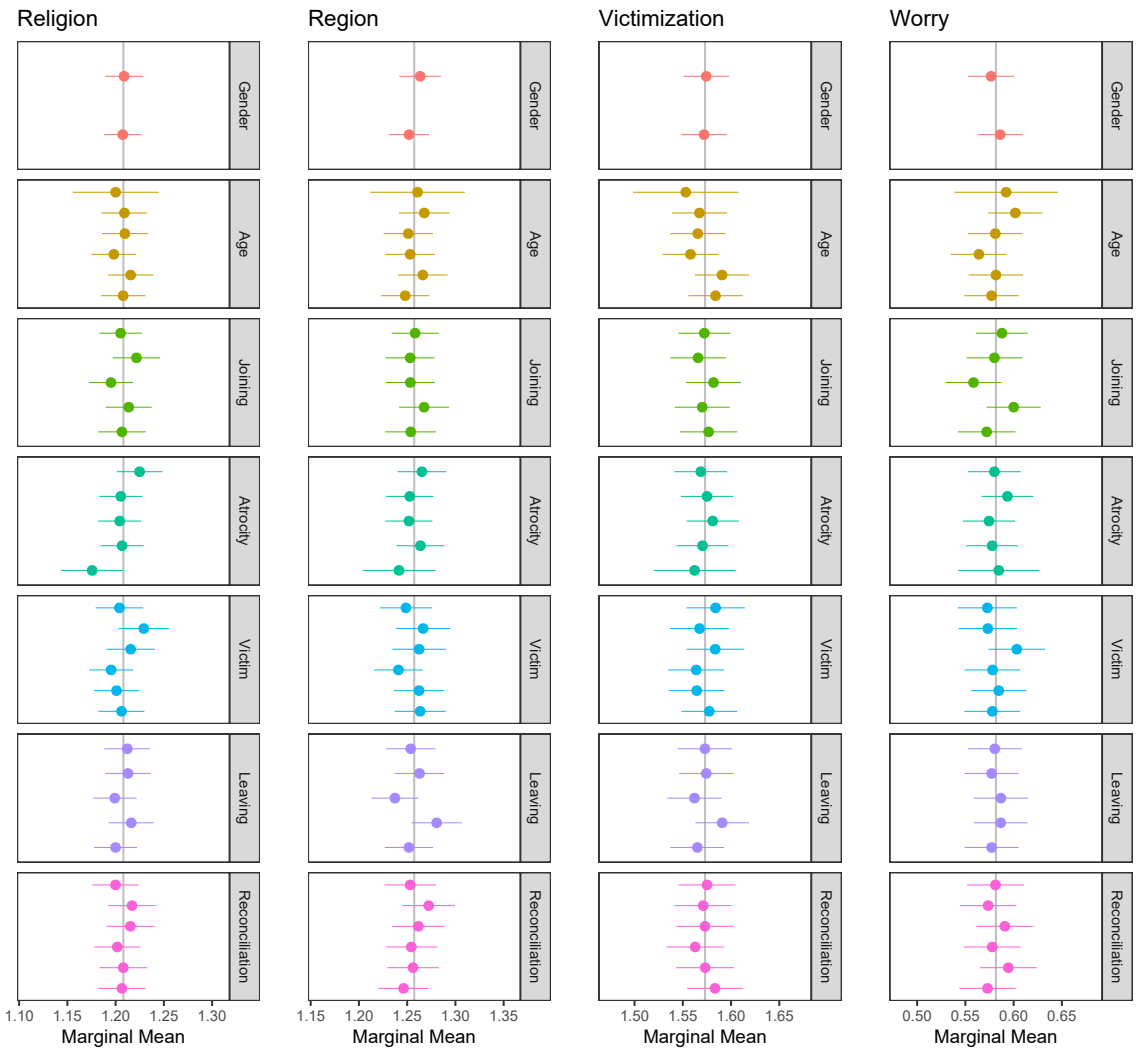

Figure A12. Balance Test for Covariates.

3.3 Carryover and Left/Right Assumption Tests

One of the assumptions underlying conjoint analyses states that "the potential outcomes always take on the same value as long as all the profiles in the same choice task have identical sets of attributes" (Hainmueller and Hopkins, 2014, p. 8). We therefore evaluate whether there the task round or profile placement affect the quantities of interest for both the choice- and rating-based outcomes. Table A6 below indicates that there are no carryover problems, but that the the left/right placement affects the results of the choice-based design. Figure A13 reveals that respondents are, on average, more likely to select the left profile, but this does not alter the substantive results of the paper. More specifically, regarding the experiment and its *causal* interpretation, Figure A13 indicates that there is not much reason to worry with respect to the quality of the experiment. Regardless of the position of the profile, the same attributes increase or decrease people's reintegration preferences compared to the baseline former fighter. Regarding the *descriptive* interpretation, the observed selection probabilities of left profiles are consistently higher, but the patterns of selection probabilities again remain the same (e.g., people showing no conciliatory intend are always chosen less often, regardless of the L/R position of the profiles). In short, the substantive interpretation of the results remains the same.

Table A6. Assumption Tests.

	F-value	p-value
Carry-over assumption	1.445	.205
Left/Right assumption	1.681	.001

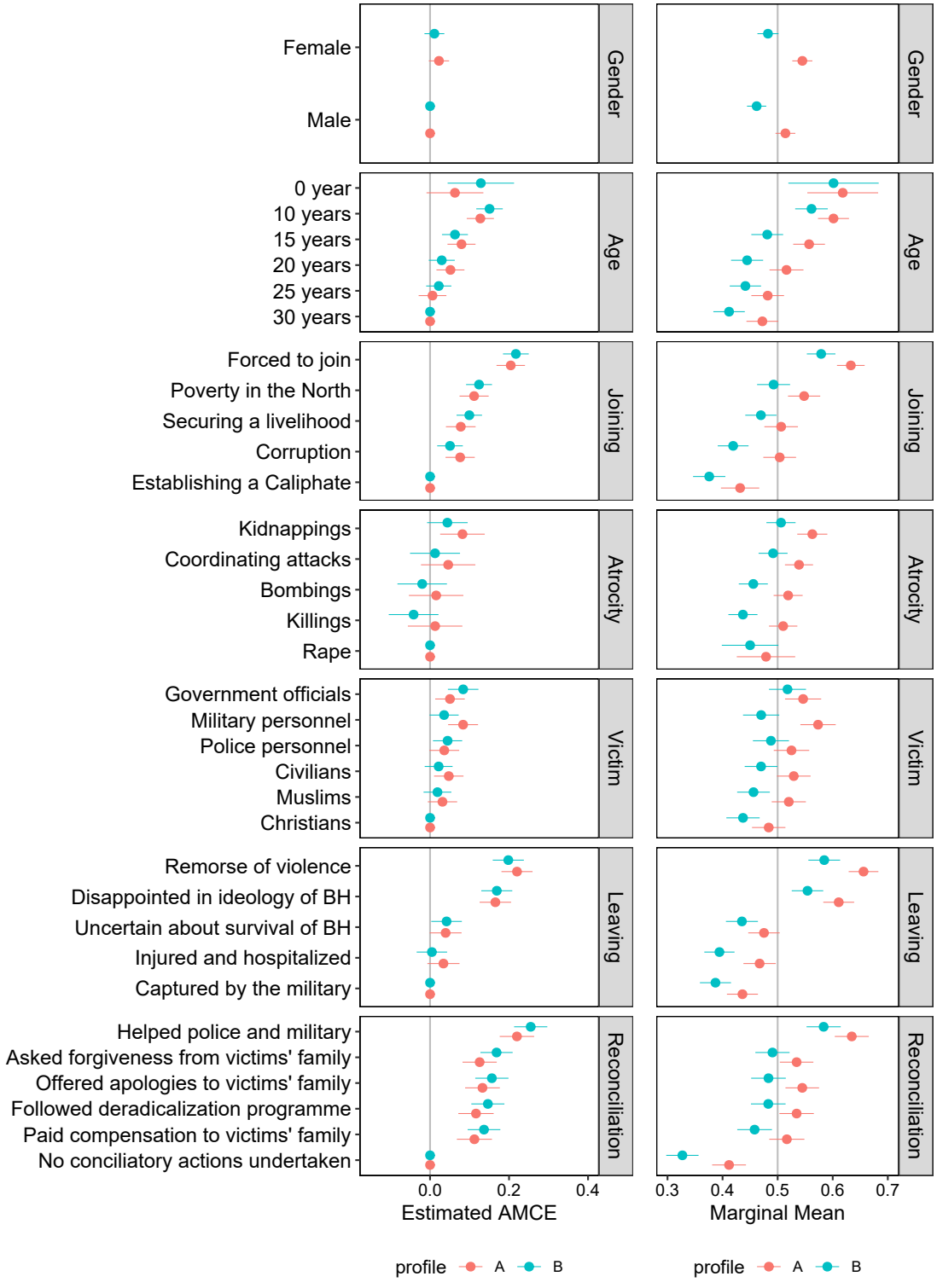

Figure A13. Left/Right Assumption Test for Choice-Based Design.

3.4 Post-Hoc Power Analysis

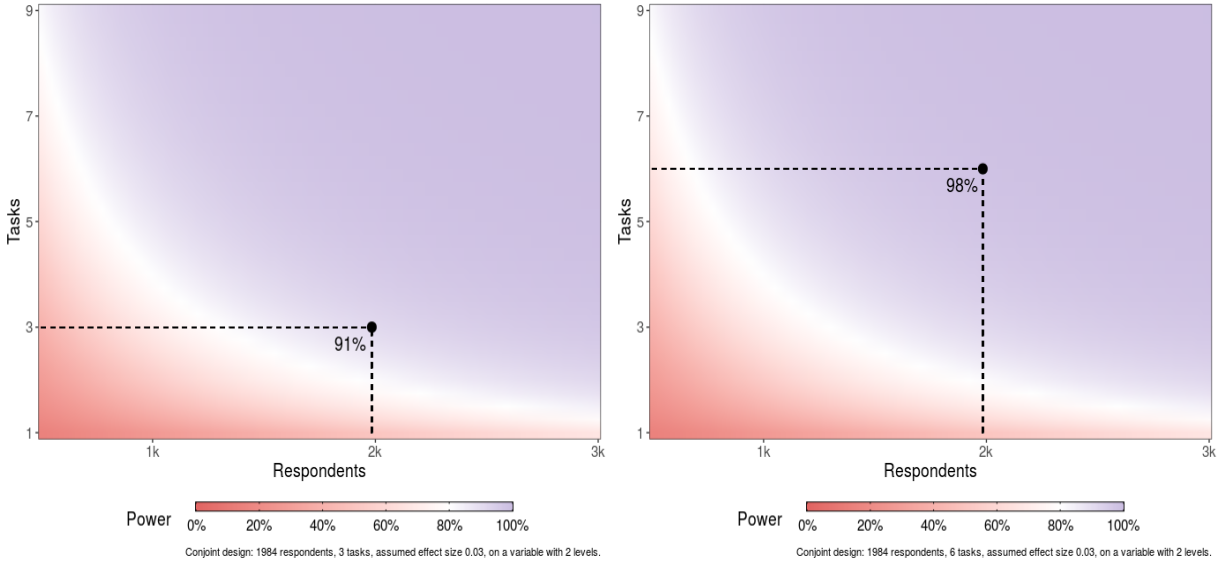

Figure A14. Post-Hoc Power Analysis for Primary (i.e., Left = 3 Tasks) and Secondary (i.e., Right = 6 Tasks) Outcomes based on Gender Attribute (i.e., Attribute with Lowest AMCE).

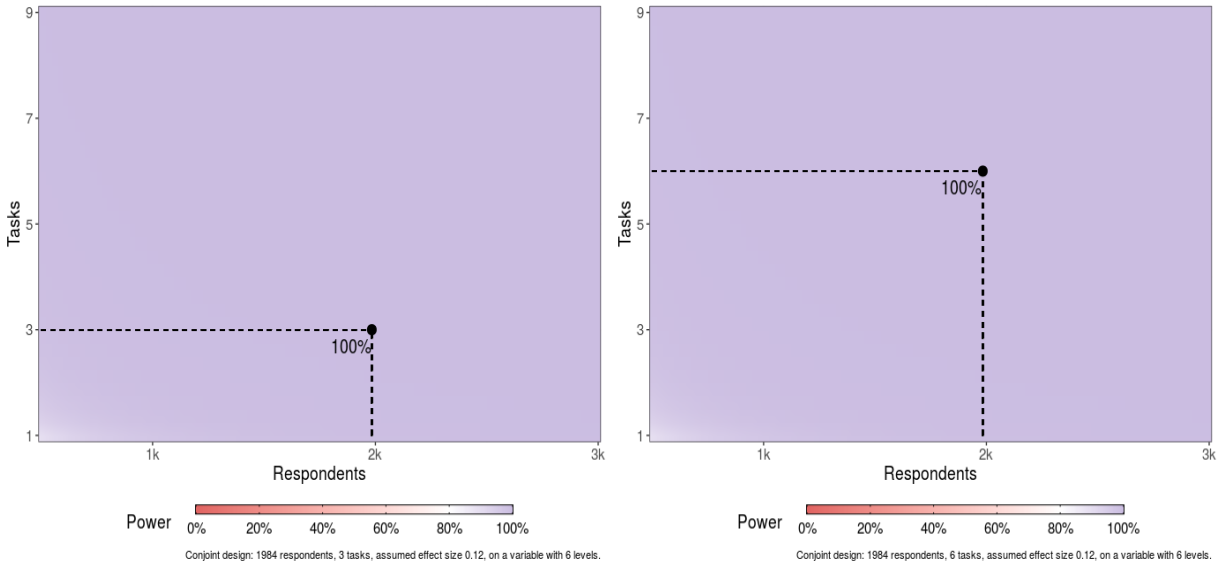

Figure A15. Post-Hoc Power Analysis for Primary (i.e., Left = 3 Tasks) and Secondary (i.e., Right = 6 Tasks) Outcomes based on Reconciliation Attribute (i.e., Attribute with Most Levels).

4 Robustness Check for Figure 5

Finally, we examine whether respondents who reported some level of exposure to Boko Haram violence (i.e., witnesses, injuries, loss of a relative, or loss of income) differ in their responses to the experiment compared to those who did not experience such harms. In line with the results in the main manuscript, self-reported exposure is not a significant moderator (see Figure A16).

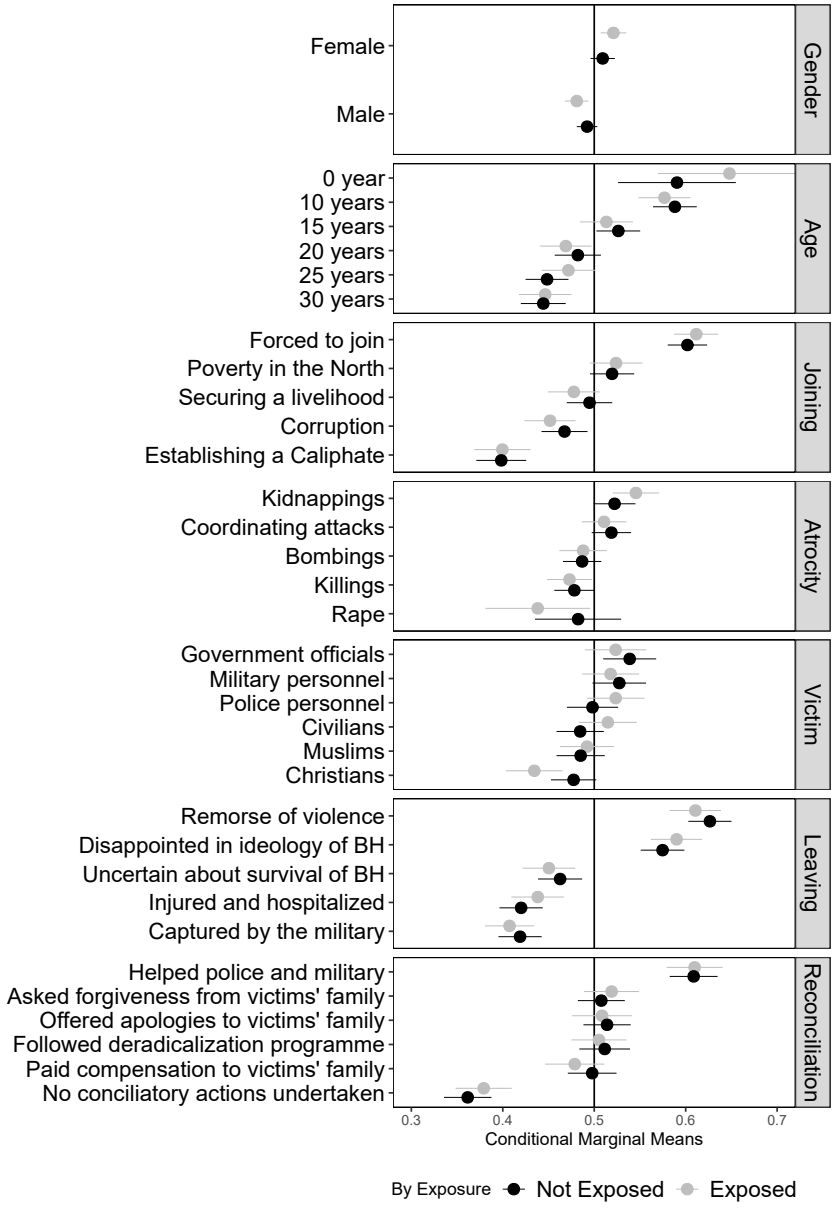

Figure A16. Marginal Means Conditioned by Respondents' Self-Reported Exposure Level.

References

AAPOR. (2016). Standard Definitions - AAPOR.

Abramson, S. F., Koçak, K., & Magazinnik, A. (2019). *What Do We Learn About Voter Preferences From Conjoint Experiments?* https://scholar.princeton.edu/sites/default/files/kkocak/files/conjoint_draft.pdf

APSA. (2019). *Report Of The APSA Ad-Hoc Human Subjects Committee* (tech. rep.). American Political Science Association.

Bulmer, M. (1993). Sampling. In M. Bulmer & D. P. Warwick (Eds.), *Social Research in Developing Countries. Surveys and Censuses in the Third World* (pp. 93–102). Routledge.

Bulmer, M., & Warwick, D. P. (Eds.). (1993). *Social Research in Developing Countries. Surveys and Censuses in the Third World*. Routledge.

Callegaro, M., Manfreda, K. L., & Vehovar, V. (2015). *Web Survey Methodology*. SAGE Publications, Inc.

Dillman, D. A. (1978). *Mail and Telephone Surveys*. John Wiley & Sons, Inc.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method* (4th). Wiley.

du Toit, J. (2016). Can We Augment Web Responses With Telephonic Responses to a Graduate Destination Survey? *Assessment & Evaluation in Higher Education*, *41*(4), 560–574.

Gerber, A. S., & Green, D. P. (2012). *Field Experiments: Design, Analysis and Interpretation*. W. W. Norton & Company.

Hainmueller, J., & Hopkins, D. J. (2014). Public Attitudes Toward Immigration. *Annual Review of Political Science*, *17*, 225–249.

Hainmueller, J., Hopkins, D. J., & Yamamoto, T. (2014). Causal Inference in Conjoint Analysis: Understanding Multidimensional Choices via Stated Preference Experiments. *Political Analysis*, *22*(1), 1–30.

Langer, A., Meuleman, B., Oshodi, A.-G. T., & Schroyens, M. (2017). Can Student Populations in Developing Countries Be Reached by Online Surveys? The Case of the National Service Scheme Survey (N3S) in Ghana. *Field Methods*, *29*(2), 154–170.

Leeper, T. J., Hobolt, S. B., & Tilley, J. (2020). Measuring Subgroup Preferences in Conjoint Experiments. *Political Analysis*, *28*(2), 207–221.

Lesser, V. M., Newton, L. A., & Yang, D. (2012). Comparing Item Nonresponse Across Different Delivery Modes in General Population Surveys. *Survey Practice*, *5*(2).

Matanock, A. M. (2021). Advances in Experimental Political Science. In J. N. Druckman & D. P. Green (Eds.), *Advances in experimental political science*. Cambridge University Press.

Meuleman, B., Langer, A., & Blom, A. G. (2018). Can Incentive Effects in Web Surveys Be Generalized to Non-Western Countries? Conditional and Unconditional Cash Incentives in a Web Survey of Ghanaian University Students. *Social Science Computer Review*, *36*(2), 231–250.

Millar, M. M., & Dillman, D. A. (2012). Do Mail and Internet Surveys Produce Different Item Nonresponse Rates? An Experiment Using Random Mode Assignment. *Survey Practice*, *5*(2), 1–6.