Improving Intergroup Relations Amid Group Conflict - Design

Christopher Grady

December 10, 2019

1 Research Design

We evaluate the effects of Engaging Communities for Peace in Nigeria (ECPN) with a site-level field experiment. Each site contains two communities, one of farmers and one of pastoralists. The communities within a site engaged in deadly clashes within one year of our scoping exercise.[1] We identified fifteen sites eligible for the study and surveyed ~50 randomly selected respondents per community. We then randomly selected the communities in ten of fifteen sites to receive the ECPN program, blocking by state so that an equal proportion of sites in Benue and Nassarawa received the program. After 18 months, we surveyed another ~50 randomly selected respondents and ~10 respondents from the baseline survey per community. In between the surveys, we monitored farmer-pastoralist interactions in markets and at social events.

This designs gives us two datasets to analyze. First, we aggregate the randomly-sampled individuals to compare communities before and after ECPN. Communities were randomly assigned to receive ECPN or function as a control group, which allows us to determine the causal effect of ECPN at a community-level. This comparison between communities that received or did not receive ECPN is our main analysis.

Second, we supplement the community-level analysis by creating a dataset of ~10 respondents per community before and after ECPN. From our baseline random sample, we identified and resurveyed (1) ECPN committee participants, (2) respondents who lived in intervention sites but did not participate in ECPN committees, and (3) respondents from the control group, who neither participated in ECPN committees nor lived in communities where ECPN was implemented. We then compare the change of participants and nonparticipants in intervention sites to the change in control respondents. The main goal of this analysis is to learn about the effect of participating directly in ECPN committees, and thus directly experiencing intergroup contact, relative to the effect of living in communities where ECPN was implemented but not participating in committees, and thus only experiencing indirect intergroup contact. Our ability to make generalizable causal claims about participation is limited, though, because individuals in intervention sites were not randomized into participation or nonparticipation with ECPN committees.[2]

[chris: figure showing sampling strategy, numbers per group, and timeline.]

1.1 Estimation

Here we describe our estimation procedure for the community-level analysis and the individual-level analysis. For both analyses we estimate one-tailed "greater than" tests because our hypotheses are that the change in outcomes for treatment units will be *greater than* control, not that the change in outcomes for treatment units will be *different* than control. Both analyses also use randomization inference for p-values and bootstrapping for standard errors. The specifics of each procedure are described in Appendix A.

We use the difference-in-differences framework to estimate the effect of ECPN. We have two observations per community: a baseline outcome and an endline outcome. To maximize precision, we generally predict endline outcomes while controlling for baseline outcomes.

---

[1]To identify eligible sites, we undertook a scoping exercise to determine if the two communities in an implementation site had a demonstrated need for a peacebuilding program and were willing to participate in one. We defined "demonstrated need" as the communities engaging in violent clashes within one year of the scoping exercise. Willingness to participate in the program was obtained through conversations with community leaders, none of whom refused the program.

[2]We initially randomly assigned baseline survey respondents to be part of ECPN committees, but random assignment proved difficult. Many people who were not selected wanted to be on the committees, and some people who were selected were not able to participate or could not be located when the committees were launched. As a result, people self-selected into committees.

$$Y_{i,j} = \beta_0 + \beta_1 Z_{i,j} + X_{i,j} + \delta_j + \epsilon_{i,j}$$

Where i is the community in state j, Z is the treatment indicator, X is the outcome at baseline, and Y is the outcome at endline. δ is a fixed effect for the state j in which the community belongs.

However, the "controlling-for" method is biased when treatment assignment correlates with baseline outcomes.[3] Therefore, when the baseline difference between treatment and control groups is greater than 0.20 standard deviations, we use the "differencing" method typical of difference-in-differences estimation. This method sacrifices power to ensure an unbiased estimate of the average treatment effect.

$$Y_{i,j} = \beta_0 + \beta_1 Z_{i,j} + \delta_j + \epsilon_{i,j}$$

Where i is the community in state j, Z is the treatment indicator, and Y is the change in outcome from baseline to endline. δ is a fixed effect for the state j in which the community belongs.

We use randomization inference for p-values and bootstrapping for standard errors because our units of analysis, communities and individuals, are clustered in sites and we have only fifteen sites. Analytic standard errors may underestimate the uncertainty of our causal estimate [cite]. With randomization inference, we first shuffle the treatment variable to break the relationship between treatment and outcomes. Next we regress outcomes on treatment using the equations specified above. We then store the resulting coefficient. Lastly, we repeat that process 10,000 times to create the distribution of coefficients we would observe if treatment had no effect on outcomes – the null hypothesis. Our p-value is the proportion of the null distribution that is greater than or equal to our observed coefficient. Bootstrapping for standard errors is similar, but instead of shuffling the treatment indicator we resample units with replacement. Bootstrapping yields a distribution of possible treatment effects given the observed data, and the 95% confidence interval is between the coefficients at the 2.5th percentile and the 97.5th percentile. More details can be found in Appendix A.[4]

1.2 Outcomes

We measured three outcomes to estimate the effect of ECPN: (1) intergroup attitudes, (2) intergroup contact, and (3) insecurity. If ECPN improved intergroup relations, we would expect respondents to report better attitudes towards the outgroup, more intergroup contact and willingness to engage in intergroup contact, and reduced insecurity due to violence. We also measured three mechanisms from the contact literature through which contact could affect outcomes: (1) empathy/perspective-taking, (2) perceived threat, and (3) ingroup expansion. Lastly, we measured a placebo outcome that may be affected by social desirability: attitudes about violence. We measured these outcomes with survey self-reports, survey experiments, a natural-field behavioral game, and monitoring of farmer-pastoralist interaction in markets and social events.

For most survey self-reports, we combine together several survey questions to create an index. We create both additive indices and inverse-covariance weighted indices. Inverse-covariance weighting constructs an index by down-weighting index questions that are correlated with other index questions and up-weighting those that are uncorrelated with other questions. This approach maximizes the amount of unique information the index takes from each question and prevents "double counting" when two questions measure the same thing. We report results using inverse-covariance weighted indices, but results hold with additive indices. Results with additive indices are included in Appendix 2.

1.2.1 Primary outcomes

Intergroup affect: Our first outcome is affect towards the other side. A primary goal of our contact intervention, and of much previous contact research, was for individual's attitudes to improve, Changing attitudes towards the other side is one pathway towards improving intergroup relations and changing behavior,

---

[3]For a comparison between the controlling-for method and the differencing method, see https://declaredesign.org/blog/2019-01-15-change-scores.html.

[4]Randomization inference: We mimic our randomization process by randomizing the intervention to communities in site-level clusters and within state blocks. This means that both communities in an implementation site (farmers and pastoralists) will always be treated together and that assignment to the intervention is conducted separately in Nasarawa and Benue, just as the intervention was assigned in this study. This procedure ensures that our null distribution is created by randomizing the intervention between exchangeable units.

though recent research shows that attitude change is not necessary for behavioral change [@paluck2009jsp; @scacco2018nigeria].

We measure intergroup affect with survey self-reports and an endorsement experiment. The survey questions include two measures of intergroup trust and a five item social distance scale created for the farmer-pastoralist context. We create an index measuring intergroup affect with these seven questions; the index alpha is [chris: alpha].

In an endorsement experiment, respondents are asked how much they support a hypothetical policy. In the treatment condition, the policy is 'endorsed' by a group that the respondent has a positive or negative opinion about. In the control condition, the policy is not endorsed by any group. The average difference in support between the endorsed and unendorsed policy represents the change in support for the policy because of the group's endorsement. In our case, we asked respondents how much they would support a water policy if it was endorsed by a farmer organization (asked of pastoralists), if it was endorsed by a pastoralist organization (asked of farmers), or if no endorsement was mentioned (the control condition posed to both pastoralists and farmers). Support was measured on a 5-point scale, where high values indicated support and low values indicated opposition.

Intergroup contact: Our second outcome is intergroup contact that occurs outside of the intervention. Natural, voluntary intergroup contact provides behavioral evidence that farmer-pastoralist relations are improving. We measure intergroup contact with survey self-reports, monitoring of farmer-pastoralists interactions in markets and social events, and a survey experiment.[5][6]

The self-reports and behavioral observations tell us the real, descriptive change in intergroup contact. The survey self-reports ask if and how often the respondent interacted with the other group in the past month. The respondents are asked about interaction in markets, at public social events, in the respondent's own home, at the home of a member of the other group, and in any other way. The responses are then ranked, scaled from 0-1, and combined into an index. The behavioral observations provide a measure of contact independent of response biases.

In the markets, we measured interactions related to buying and selling market goods, such as the number of farmer and pastoralist sellers present and the number of farmer and pastoralist buyers. We then create a farmers index and a pastoralist index to measure the presence of farmers and pastoralists in the market. At social events, we measured the number of members of the other group in attendance and the number who ate or drank anything[7], both in absolute numbers and as a percentage of total attendees. We then create measures for the number of farmers and pastoralists attending social events and the number of farmers and pastoralists eating at social events.[8]

A survey experiment, which we are calling the *percent experiment*, tells us about respondents' willingness to engage in contact. It asks respondents two questions about their willingness to interact with members of the other side. We asked respondents if they would (1) join a group and (2) live in a community with

---

[5]Much of the self-reports and the observations are overdispersed count data. We recode all count data as rank.

[6]We also attempted to measure willingness to engage in contact with a second survey experiment, a list experiment. List experiments are used to provide anonymity to respondents and encourage them to give honest answers to sensitive questions. In a list experiment, the researcher randomly assigns respondents to one of two (or more) conditions. Individuals in the control condition are presented with a list of three items; individuals in the treatment condition see the same list plus an additional item, which is the item of interest and the one on which the experimenter wants to ensure the respondent of anonymity. Subjects are asked how many items apply to them. The average difference between the treatment and control conditions represents the percentage of respondents who responded to the sensitive item. In our case, the sensitive item read "When you have to interact with a member of [the other group] in the market." Our list experiment failed. The endline difference between the 3-item list and the 4-item list is negative in almost half of all communities, implying that the presence of the 4th item slightly decreased the number of reported items. This decrease violates the assumptions of the list experiment. The average difference between the 3-item list and the 4-item list should not be negative; the addition of a 4th item should not decrease the propensity of subjects to respond to other items. We therefore exclude the list experiment as a measure.

[7]Taking food or beverages at a social event is a sign of closeness and intimacy in these contexts. Casual attendees would not take food or beverages

[8]Observations were made in two periods: July 2016 – February 2017, immediately after the project commenced but before joint project committees convened, and September 2017 – December 2017, after project committees convened but before the endline survey began. Events that occurred February 2017 or earlier are baseline measurements; events occurring September 2017 or later are endline measurements.

some percentage of the other group. The percentage is randomized between 5%, 25%, 50%, and 75%; the percentage is the same for those two questions but varies across individuals. We take the mean response so that a respondent saying yes to both is assigned a 1, a respondent saying yes to one is assigned a 0.5, and a respondent saying no to both is assigned a 0. These questions allow us to determine if treatment communities become more willing to interact with outgroup members and if treatment communities become less sensitive to higher proportions of the outgroup.[9]

Insecurity: Our third outcome is feelings of insecurity due to conflict. The end goal of ECPN is to reduce conflict between farmers and pastoralist. The disaggregated and diffuse nature of the conflict makes obtaining an accurate measure of violent conflict extremely difficult.[10] Instead, we measured the effect that violent conflict has on individuals. We ask respondents if they avoid any areas during the day or night due to insecurity and if insecurity restricted them from engaging in various activities, such as grazing their animals, working on their farms, fetching water for their families, and working for wages. We combined these ten insecurity questions into an index, with high values indicating low perceptions of insecurity and low values indicating high perceptions of insecurity. The index alpha is [chris: alpha].

Violence Placebo: Several of our outcomes are survey self-reports, and all self-reports could be affected by social desirability bias. Our survey results are suspect if respondents in treatment communities learned the "correct" answers better than respondents in control communities. If social desirability accounts for the effect in survey self-reports, we would also expect differences between treatment and control for other normatively desirable attitudes. To test social desirability effects, we conduct a placebo analysis using attitudes about violence as a placebo. Attitudes about violence are a good candidate for a placebo because intergroup contact should not affect attitudes about violence, but respondents may feel social pressure to answer violence questions in a desirable way. We measure attitudes about violence with a six question index asking respondents if it is always, sometimes, rarely, or never justified to use violence in certain situations, such as retaliating against violence or bringing criminals to justice.

1.2.2 Mechanisms

The primary outcomes of intergroup affect, intergroup contact, and insecurity tell us if ECPN worked but provide no evidence for how the program worked. Previous work on contact specified three mechanisms through which contact affects attitudes: empathy/perspective-taking, threat/anxiety, and ingroup expansion [@pettigrew2008does; @al2013intergroup; @dovidio2017reducing]. We do not manipulate these mechanisms directly, and so cannot make causal claims about the mediating role of these variables for ECPN. But we can provide exploratory evidence that these mechanisms played a role if (1) ECPN affects these mechanisms and (2) these mechanisms affect intergroup affect, intergroup contact, and insecurity.

Threat: We use three self-report survey questions to measure threat felt by the outgroup. These questions ask if the outgroup is a threat to the respondent's community, believe in different morals than the respondent's community, and overly influence the respondent's community.[11].

Empathy/Perspective-taking: We measure empathy with two questions and perspective-taking with one question. For empathy, one question asks if the respondent's group would help a member of the other side if something unfortunate happened to that person, like a serious illness or the death of a parent. The second questions is the same but asks if someone from the other group would help someone from the respondent's group. For perspective-taking, the question asks who the respondent believes is responsible for the violence between their community and the other community: the other group or both groups.[12]

---

[9]This experiment was based on a question from the GSS asking respondents if they would favor or oppose living in a neighborhood that was half white/black.

[10]Asking respondents to recount the number of violent events does not accurately measure the scale of the conflict because those answers are determined by the awareness and memory of the community members. Awareness of individual violent events is low because many of the violent events occur in fields and grazing routes far from the town center and residential areas. In addition, ECPN sought to increase awareness of violent events through its conflict forums. The type of event that all community members are aware of – large massacres, burning of homes, etc... – generally lead to the disintegration of both communities as community members flee the area fearing further violence or reprisals. These large-scale events are rare and none occurred in intervention or control communities during the study.

[11]These threat questions are based on questions from @van2007testing

[12]We planned to create an index with these three questions, but their alpha was below 0.70 and improved to 0.90 without the

Ingroup expansion: We measured respondents' recategorization of their ingroup to include outgroup members with eight survey self-reports and a public goods game. Five survey questions ask respondents to answer questions about "people in this area, including people from the other group", such as if the groups share the same morals and if the groups work together to achieve common goals. Three more questions ask the respondent about the groups working together on specific goals, such as repairing a road or solving a water supply problem.

We also used a natural-field public goods game to measure the ability of the groups to cooperate to achieve a common goal. If ECPN causes respondents to incorporate the former outgroup into their ingroup, then we expect those communities to better cooperate in a public goods game. Compared with lab-based behavioral games, whose choice-making situations are necessarily artificial, the choice-making situation of a natural-field game is akin to the choices people make in their lives (Harrison and List 2004; Winking and Mizer 2013). Because these communities often decide how to contribute to some public good, such as repairing a borehole or a market, we chose to use a natural-field public goods game (PGG) as a realistic behavioral measure of cooperation.[13]

These designs and measurements put us in a strong position to identify effects if effects exist. First, we have data at the community-level and individual-level. If the two analyses show similar relationships, we can be more sure that those relationships are not spurious. Second, both community and individual-level analyses use a baseline/endline + control group design to differentiate a secular trend from a treatment effect. Many things change in the social environment between the beginning and the end of ECPN that could deteriorate intergroup relations, especially an economic downturn in Nigeria and the anti-grazing law in Benue. By comparing the *change* in the treatment group to the *change* in the control, we are more certain that differences are due to ECPN and not other factors. Third, outcomes are measured using survey self-reports, survey experiments, a behavioral game, and monitoring of social behavior. If we observe similar relationships across multiple modes we can be more certain that the relationship is not spurious.

---

perspective-taking question.

[13]This game is similar to the one implemented by Fearon, Humphreys, and Weinstein (2009) as part of a similar study on community-driven development in Liberia.