The effect of a new Hausa language television station on attitudes about tolerance, violence, and women's empowerment in Northern Nigeria

Jake Bowers[*]   Chris Grady[†]   Graham Couturier[‡]   Annette Brown[§]

November 08, 2017

Abstract

We present preliminary results from a series of studies aiming to assess the impact of the Arewa24 Hausa language television channel on the attitudes of people in Northern Nigeria. In this particular paper, we report on the results of a matched design in which we compare the attitudes of people who reported watching the Arewa24 television channel in 2015 to the attitudes of people who either (1) were interviewed in 2013, before the channel launched, or (2) reported not having watched the channel in 2015. We also present pre-analysis plans for another matched study, using two different cross-sectional surveys, as well as a pre-analysis plan for an experimental encouragement design.

1 Introduction

In June of 2014, Equal Access International launched free to air Hausa language television station in Kano, Nigeria, Arewa24. In the context of growing support for violence in Northern Nigeria, Equal Access hoped that programming on the channel would diminish popular support for violence and associated religious intolerance or, at least, stem the tide of such support. This channel is the first of its kind, presenting culturally appropriate content in Hausa written, produced, and filmed by veterans of the Hausa-language film industry based in Kano, Nigeria. The intended audience for this channel includes all of Northern Nigeria. This paper presents a preliminary assessment of the impact of this attempt to use television to change (or resist change in) attitudes in the context of violence and social change in Northern Nigeria and the Muslim world.[1]

Although no television station can combat an armed insurrection, the hope of Equal Access and many other organizations devoting energy on media-based anti-violence programs is that popular support for such an insurrection may be moderated or at least not increased, at low cost, via information and compelling narratives. Media interventions, such as radio and television programs, offer a method to reach an enormous number of people with anti-violent, anti-prejudice messaging. Governments and

---

[*] jwbowers@illinois.edu, University of Illinois at Urbana-Champaign, Departments of Political Science and Statistics.

[†] cdgrady2@illinois.edu, University of Illinois at Urbana-Champaign, Department of Political Science.

[‡] gcouturier@equalaccess.org, Equal Access International

[§] ABrown@fhi360.org, FHI 360

[1] In the years just preceding the launch of Arewa24, Northern Nigerians suffered from the mass attacks and virtual secession of the North Eastern state of Borno at the hands of Boko Haram — a group now allied with ISIS — as well as attacks of various kinds, including the bombing of mosques, in Kano and other cities in the North Central and North West of the region.(cites)

1

non-governmental organizations routinely tap into the power of media to engage their populations about prejudice (Howard et al. 2003; Paluck and Green 2009b). The movement began with *telenovelas* in Mexico produced by Miguel Sabido, which launched an attitude-change methodology that soon spread to programs in India and Kenya before being adopted by media organizations worldwide (Singhal, Rogers, and Brown 1993). In 2014 media NGO Equal Access took this methodology to northern Nigeria. Because Arewa24 is a channel, and not a single program, Equal Access hoped to use one program to focus on moderating attitudes toward violence, another to liberalize attitudes toward women in society, and yet another to empower youth. All programs also aimed to increase tolerance toward people of different religions (mainly different varieties of Islam, but also toward Christians from southern Nigeria).[2]

This paper *will eventually* present data from 2 pairs of large before-vs-after cross-sectional surveys as well as a randomized experiment that encouraged people to increase their exposure to the soap opera Dadin Kowa. *For now*, we present a matched design for the first pair of cross-surveys (a pair done by the same organization in 2013 before the channel was launched and again in 2015 after about one year of operation). We also present pre-analysis plans for the second set of surveys and the randomized experiment. The two observational designs help shed light on the impact of the channel overall (i.e. the effect of any exposure to any part of the channel). The experimental design helps us learn about the effect of the content on the channel.

The effects that we report in the first set of studies are surprising and disappointing from the perspective of Equal Access: we will show that those people who reported watching the channel had less liberal attitudes toward women's role in society (both men and women respondents) and supported the use of violence more, than matched survey respondents who reported no exposure to the channel (or who were interviewed in 2013, before the channel existed). We will also show an *positive* difference in tolerance toward other groups.

Learning about the the impact of this attempt to use media to change attitudes and norms offers us knowledge about an important policy tool in common use in development. It also offers a chance to learn about the theory of vicarious learning that would explain and motivate the use of the media to change attitudes in the first place.

2 Why would television programming change attitudes?

Arewa24's programs might change viewer's attitudes because humans learn through observation as well as through direct experience (Bandura 1971). Learning through observation is not confined to physical behaviors such as riding a bike or throwing a baseball, but extends to thoughts and emotions, as well (Bandura 1971; Bandura, Bryant, and Zillmann 1994). Just as we learn about the desirability of our own behaviors through their ensuing results, we learn vicariously through observing the consequences that result from the behavior of others. Observations that can cause learning need not be of physically present models; television programming is especially conducive for observational learning because viewers naturally attend to and value the behavior and outcomes of the characters (Bandura and McClelland 1977, Bandura (2001)).

2.1 Learning about attitudes through observation

Bandura's theory of social/observational/vicarious learning upended the then dominant theory of learning, Behaviorism. In its simplest form, behaviorism argued that learning occurs when a stimulus,

---

[2]Arewa24 is a 24 hour channel, and so, it produces a total of about 12 original shoes and also translates appropriate content from abroad: for example, it carries a Turkish soap opera translated into Hausa.

normally our behavior, produces a response. If that response is favorable (a reward) we repeat the behavior; if that response is unfavorable (a punishment) we do not. This is the formulation proposed by first by Thorndike (Thorndike 1905) and later by Skinner (Skinner 1990), who named this type of learning *operant conditioning*. Bandura rejected the clean stimulus-response associations on the grounds that humans were not passive receivers of information and external punishment or reinforcement was not the only, or even the primary, method of human learning[3]. Bandura's Social Learning Theory diverged from past work in emphasizing the idea that cognitive processes, not behavioral processes, governed human learning. He suggested that it wasn't that a behavior was rewarded that made a person learn that the behavior was good, it was his or her own understanding of what that behavior would get us in the future. An immediate reward for the behavior might provide us information on future consequences of the behavior, but we could imagine consequences in the absence of a reward, and even imagine future consequences opposite the consequences we had just received.

What does Social Learning Theory have to do with attitude change? After all, a learner can only directly observe the behavior of others, not the underlying attitudes and motivations that generate that behavior. But to Bandura, what humans learned from viewing a behavior-reward situation was not the clear, unambiguous lesson that the behavior was good. Humans form a "theory of mind" about the intentions of those they observe (Wimmer and Perner 1983), attributing to them the mental states that could generate their behavior. An observed behavior is symbolic of other information — of the underlying attitudes that motivated the behavior, the planned action to satisfy that attitude, and the implementation of that plan. When humans observe a behavior-consequence situation, they learn about far more than behavior.

Humans are hard-wired to seek and label cause and effect relationships (cite) and so people attempt to identify not only the response the behavior caused, but what attitude caused the behavior. For example, imagine someone being congratulated for a failing effort. As an observer, we do not necessarily learn that failing is desirable, or that their failed plan was a good one. Rather, we may learn that that the underlying desire that motivated the behavior was worthy or at least functional. Thus, social learning theory provides both the mechanism for people to learn through observation and the mechanism for people to learn attitudes, not just behavioral reproductions from observing others.

2.2 When will observational learning of attitudes occur?

For observational learning to generate attitude change, compared to direct experiential learning, the presentation must (1) to connect the observed behavior and consequence to the viewer, and (2) ensure that the viewer interprets the mind of the model correctly.[4]

First, the social learner must believe that the observed behavior-consequence relationship is informative for the consequence the learner would receive for the behavior if the learner were to *do* the behavior. The relevance of observation to observer can be enhanced in three ways. First, the person that the learner observes — called a behavioral model — should be a good comparison to the learner; i.e. the learner and the model should be similar on traits important for the outcome (Bandura and McClelland 1977; Kazdin 1974a). For example, a person afraid of the influence of western values will learn more

---

[3]Bandura was not the first proponent of humans as active information processors. Edward Tolman argued that learning could occur without external reinforcement because, in essence, living things create their own rewards for their behavior, independent of rewards provided by the environment. And further, when a reward is received, the nature of the reward matters just as much as receiving a reward. Tolman's modification of behaviorism became known as Purposive Behaviorism. But Bandura certainly took the idea of an active learner further than Tolman.

[4]Cites and explanation

observing another person with the same fear overcome that fear than observing a non-fearful person remain unafraid. Second, the behavior-consequence relationship should be shown occurring for multiple models (Bandura and Menlove 1968; Kazdin 1975,Kazdin (1976)). Observing a relationship multiple times adds credibility to the relationship between behavior and outcomes. Third, the reward should naturally be caused by the behavior so that observers can identify that it is the behavior that causes the reward, not that the reward was "generic" and incidentally given for the behavior (Kazdin 1974b). Human causal reasoning looks for mechanisms when identifying causal relationships (Ahn and Kalish 2000, Blaisdell et al. (2006)), so the reward cannot just co-occur or covary with the behavior. To assist the learner with the belief that a reward will follow an observed attitude or behavior a television program should use many relatable behavioral models receiving benefits that are clearly caused by the behavior.

Second, the learner must form an accurate theory of the mind of the model and understands what about the behavior is being rewarded. For Arewa24, for example, viewers should understand that friendly interactions with out-group members are rewarded because positive affect for the out-group is good not because the out-group members are being effectively fooled. That is, the viewer needs to understand that their behavioral model has positive feelings for the out-group member. If the viewer observes peaceful intergroup interactions but interpret the lesson as "it is good to interact peacefully, even with people you hate," then they can model peaceful behavior, but not tolerant attitudes. This is one of ways that television programming may have an advantage over real-life observation: the television program can provide context and interpretation to clarify which lesson viewers should learn. The second way is that, in many dramas, like the Dadin Kowa drama on Arewa24, viewers observe the characters over a long period of time, and so can observe both short-term consequences for a behavior and long-term consequences if the script writers can plan for long narrative trajectories. Third, television can demonstrate attitudes not present in their normal social environment (Bandura 2001). Fourth, as previously discussed, television naturally captures the attention of viewers (Bandura and McClelland 1977).[5].

For these reasons, and because of the low cost per person exposed, television appears an ideal medium through which to foster social change. Based on this theory, "prosocial" programs have appeared in Mexico, Peru, India, and Kenya, and observational studies have shown them to be associated with adult literacy, family planning, female equality, and child development (Singhal et al. 2003; Singhal, Rogers, and Brown 1993). And field experiments in developing countries have found that media broadcasts can influence social norms against dissent (Paluck and Green 2009a), promote family planning activities (La Ferrara, Chong, and Duryea 2012; Piotrow et al. 1990; W. Vaughan 2000), improve women's status among women (Jensen and Oster 2009), and encourage financially responsible behavior (Berg and Zia 2013).[6]

In this study we focus on three attitudes in particular: attitudes towards religious and ethnic out-groups,

---

[5]There are many other ways in which learning can be encouraged through television. For example, viewers may form an emotional attachment with characters, and emotional activation improves learning (McGaugh, Cahill, and Roozendaal 1996). The timing and frequency of reinforcement also affects learning (Hogarth et al. 1991) and television can provide consistent feedback, ensuring that an encouraged behavior is always rewarded and a discouraged behavior is always punished.

[6]Numerous studies suggest television is effective at influencing the attitudes of the viewers towards out-groups. In a meta-analysis about the effects of Sesame Street on children's attitudes, Mares and Pan (Mares and Pan 2013) conclude that the program positively affects numerous attitudinal outcomes, including attitudes towards out-groups – even out-groups associated with "long-standing hostilities or stereotyping." Shatzer et al (Shatzer, Korzenny, and Griffis-Korzenny 1985), Schiappa et al (Schiappa, Gregg, and Hewes 2005), Ortiz and Harwood (Ortiz and Harwood 2007), and Mazziotta and Mummendey (Eller et al. 2011) all find that programs with out-groups or intergroup interactions positively affect attitudes and beliefs about the target out-group.

attitudes about violence, attitudes toward the role of women in society. In terms of observational learning for attitudes about violence, Bandura's (Bandura, Ross, and Ross 1963) seminal "Bobo doll experiment" launched the ideas of social learning theory into the academic consciousness by demonstrating that aggression could be learned through observation. Subsequent studies have confirmed observational learning of aggression through broadcast media (Drabman and Thomas 1974; Hearold and Comstock 1986 for a review; Molitor and Hirsch 1994). Learning about aggression and violence can go both ways, and numerous other studies demonstrate that prosocial television programming can reduce aggressive acts (Bankart and Anderson> 1979; Lovelace and Huston 1983). The effect is not confined to children, and both college-aged adults (Malamuth and Check 1981) and older adults (Loye, Gorney, and Steele 1977) become more aggressive and accepting of violence after viewing violence, and less aggressive after viewing prosocial media. Outside of the lab, there is little evidence for television, or any media, to reduce viewers' prejudice and aggression. The few and pioneering field studies that we have found show no real effects on the kinds of interpersonal attitudes that Arewa24 aimed to influence. For example, Paluck (2009) and Paluck and Green (2009a)'s radio show experiment in Rwanda influenced perceptions of social norms and social deference, but "did little to change listeners' personal beliefs" about prejudice, violence, and trauma (2009, 1). Gibson and Gouws (Gibson and Gouws 2005), employing written vignettes in South Africa, report that no amount of positive information about out-groups could override subject's level of out-group animosity.

2.3 Why might a television program not change attitudes?

Why is evidence for this theory, so abundant in the lab, so elusive in the field? We believe there are three practical reasons and four psychological reasons. Practically, field experimental interventions of media effects are difficult and expensive. Producing radio or television content is expensive. Creating a new radio or television channel is even more expensive. Most people do not live in environments empty of competing media content and so the broader media environment will likely mute the intervention's effect on attitude change, and requiring very large samples to detect the effect from the noise.

The second practical reason is that most media interventions center around radio, and radio may not be conducive like television. The reason for this, obviously, is that radio programs are much cheaper to produce and disseminate. However, radio communication precludes the communication of visual information, and physical observation is an important component of social learning theory. Life-like visual and audio sensory information should increase identification with a program's characters (Crigler, Just, and Neuman 1994) and increase the likelihood and effectiveness of observational learning.

The third practical reason is that the researchers studying the effects of media content rarely, if ever, control or influence the content. And social scientists are unlikely to have the training to produce an effective television or radio program on their own. The producers of television and radio programs generally have a stronger entertainment incentive than educational incentive, so mechanisms for observational learning may not be triggered. At the same time, prosocial programs cannot have an effect without viewers. If entertainment value and education value conflict, then media will only have a large effect on attitudes when entertainment and education align. This limitation should dampen the possibilities for prosocial programming to change attitudes.

The four psychological reasons call into question the theory's ability to change attitudes and may imply limits to its efficacy. First, viewers can only observe the behavior of a model. Attitude change requires that the viewer have a symbolic understanding of the attitude that motivates the behavior, but it is difficult to control viewers' interpretations of the behavior — especially if television viewing

is done in groups. In theory, it should be far easier to change ideas about social norms and other perceptions of society, since that does not require a theory of mind about the model. A viewer can understand that people behave a certain way in their society without understanding why. Indeed, research shows that social norms are successfully conveyed through media interventions (Paluck and Green 2009a), and other prejudice reduction interventions display similar results (Scacco and Warren 2016).

Second, it is easier to teach attitudes that are in the direct material interest of the viewer. Past research shows that television viewing of prosocial programming has a positive effect on literacy, family planning, and financial responsibility, but all of these behaviors provide direct material benefits for the viewer. The benefits of more positive attitudes towards out-groups, or violence, is much less direct. The learner must connect the attitude/behavior to future desirable outcomes for any learning to occur, but what individually desirable outcomes occur because of reduced prejudice? In many cases, reducing prejudice is only good if (1) rest of the social group reduces prejudice as well, so the viewer is not ostracized for becoming less prejudiced, and (2) another social group reduces prejudice against a viewers own group at the same time as the viewers group reduces prejudice so that neither group are not at strategic disadvantage if only one stops discriminating.[7]

Third, the final process of Bandura's theory involves motivation and reinforcement. Attitude change of the type encouraged by NGO and government led efforts often aim to push the attitude to move counter to an existing group norm, and so the group may punish the encouraged attitude. This is especially likely when attempting to change attitudes towards an out-group, since the in-group's identity may involve disdain and hostility towards the out-group (Brewer 1999). Further, behavioral modeling is much more effective when the learner relates to the behavioral model and holding an attitude that is strange to the learner might work against the relatability of the model. For example, if the viewer believes that any member of their in-group who interacts with a disliked out-group is a traitor, they are unlikely to model behaviors and attitudes of traitor in-group members.

Fourth, even in the absence of negative social pressure, two psychological mechanisms work against individual attitude change: motivated reasoning (Kunda 1990) and cognitive dissonance (Festinger 1962). Motivated reasoning should prevent attitude change for two reasons. First, people are likely to attend to media that echoes their current beliefs and their interpretations of new and perhaps contrary information will tend to involve ignoring or framing the information to bolster existing beliefs (Vallone, Ross, and Lepper 1985). Second, even if intolerant people understand and interpret the new television show as pro-tolerance, they may reject that information and develop a counter-narrative. For example, if shown a narrative where a member of a hated out-group interacts amicably with a member of their in-group, the viewer may counter that by thinking of a time when he/she encountered hostility with a member of the out-group. Finally, even if the intolerant viewer accepts the message, the resulting psychological discomfort and cognitive dissonance may drive the viewer *further* from tolerance (Gubler 2013).

2.4 What can Arewa24 teach us about television and attitude change?

Our current observational study uses a matched design to compare those who reported viewing Arewa24 with those who did not report viewing the channel (in 2015) and with those who had no chance to view the channel (in 2013). If those who view the channel report more prosocial attitudes than those who do not, we might conclude that (1) the programming successfully triggered

---

[7]This is also a problem with writing compelling content: how to show people gaining tangible rewards of behavior that is not directly in their interest.

observational learning or (2) the social context of watching the channel (i.e. discussion) caused social learning or (3) that prosocial attitudes either drive cause Arewa24 viewing or are associated with the characteristics of people that we did not observe or control in the matched design. If those who view the channel who view the channel report less prosocial attitudes than those who do not, we might conclude that (1) contrary to the theoretical expectations, observational learning might not effect individual attitudes, or might not effect individual attitudes of this type; or (2) the shows may have been designed poorly for attitude change; or (3) the channel might not be watched by many of the type of people who would change attitudes (for example, if it only attracted older viewers, or displayed behavioral models and contexts that were not relatable), or (4) the shows on Arewa24 do not affect attitudes towards out-groups or violence — Arewa24 produces a dozen original programs and not all of them focus on intergroup contact or nonviolence, or (5) some attitudes may change, but the attitudes we measure may remain the same[8], (6) the effect may take more than two years be manifest given our sample size.

Though we do not know how the attitudes and behaviors relevant to intergroup interactions, violence, or women's role in society are rewarded or interpreted in the show.[9], the writers were trained on attitude change via television programming by a media consultant.[10] We do know that programs like Dadin Kowa, Arewa24's flagship and award winning soap opera, involves actors modelling intergroup tolerance and nonviolence. Most of the characters, like the population of northern Nigeria, are Hausa Muslims. And the program exhibits a wide range of Muslims and Christians interacting amicably and solving problems nonviolently.

We do not expect the effect of viewing Arewa24 to be large. Arewa24 is one channel in a diverse and sophisticated media environment. And we do not believe the effect will be homogeneous. First, with an eye to psychological mechanisms like motivated reasoning and cognitive dissonance, we predict the largest effects among youth and people with more moderate attitudes, since both should have with less ingrained attitudes. Programming to promote tolerance and nonviolence is most likely to move moderate people towards tolerance than intolerant people towards moderate opinions.[11]

3 Design

We use an observational design to assess the impact of the television channel by comparing survey respondents who reported watching the new channel in 2015 to other respondents who reported not watching the show in 2015 and other respondents who were interviewed in 2013, before the channel was launched. We also present a very brief pre-analysis plan for a randomized experiment studying the effects of an encouragement to watch the content of the channel. And a pre-analysis plan for the comparison of two cross-sectional surveys that we ran in 2014 and early 2017.

3.1 Sample

The data used in this current paper come from two household surveys conducted by ORB International, one in 2013 and one in late 2015. In total, 17,196 respondents in 1,642 neighborhood clusters completed the 2015 survey and 10,482 respondents in 1,283 neighborhood clusters completed the

---

[8]Here we are thinking of the Uganda abortion attitudes media study (cite to Green et al) in which attitudes about abortion did not change, but willingness to ostracize someone for getting an abortion decreased.

[9]We are working to get English translation transcripts of each episode of Dadin Kowa, but currently only have episode synopses.

[10]We are working to get details about this training.

[11]We did not pre-register these expectations before the current paper, although they appear as a part of the pre-analysis plans for the subsequent 2 surveys and the RCT.

2013 survey. Both survey samples arose from a two-stage cluster stratified design, using states as strata and neighborhoods as clusters. Neighborhoods from every state in Nigeria were selected into the survey, neighborhoods within states were randomly selected for enumeration[12], 8 or 10 households per PSU were randomly selected for enumeration in 2013 or 2015 respectively, and then one person per household was randomly selected as the respondent.[13].

3.2 Outcomes

Arewa24 designed its programming to affect four general attitudes: (1) intergroup tolerance, (2) support for violence, (3) support for women's empowerment, and (4) support for youth empowerment. The 2015 survey asks about all of these attitudes, but the 2013 survey includes no questions about youth empowerment. The measurement of support for women's empowerment is also complicated because, though the surveys both ask questions about women's empowerment, they ask *different* questions to measure support for women's empowerment. We are on firmer ground comparing questions on tolerance and violence. Both surveys contain six identical questions to measure intergroup tolerance and ten identical questions to measure support for violence. In the analyses we present below, we use scales constructed by taking the mean response to the questions about violence and tolerance. We compare attitudes that either in favor of, or against, women's empowerment in 2013 and 2015 by collapsing categories of the different questions.[Insert question text.]

3.3 A matched design with two control groups

We want to know how viewing Arewa24 affects the attitudes of viewers, but Arewa24 viewers and non-viewers differ in many ways that could affect attitudes. We are not interested in differences between people who do and do not watch Arewa24 that would exist even in the absence of Arewa24, only those differences which arise as a result of viewing Arewa24. Our goal, then, is to design a study that isolates the effects of viewing Arewa24 from other pre-existing differences between Arewa24 viewers and non-viewers. Since this is an observational study, we cannot isolate the Arewa24 based differences from all observed and unobserved pre-existing differences, rather we do our best to isolate the Arewa24-based differences from differences in covariates that we did observe in both 2013 and 2015.

Our procedure for clarifying the viewer to non-viewer comparison (for making the case that, say, the differences are not due to gender, or region, or education), is to create a matched design comparing self-reported viewers to self-reported non-viewers (in 2015) and non-viewers by construction (in 2013). Non-viewers from 2013 and non-viewers from 2015 are good comparisons to viewers in 2015 in different ways. A 2015-2013 comparison matches people who watch Arewa24 with very similar people who do not watch because the channel did not yet exist: that is, the 2015–2013 comparison makes 2013 appear, prima facie, as a good counterfactual version of the 2015 viewers. And 2013 non-viewers are people we would predict to become viewers once the channel airs. However, they are displaced in time, and so differences may be the result of Arewa24 viewing, or they may be the result of a two-year time difference[14]. Between 2013 and 2015 the relationship between Boko Haram and ISIS become more formalized, violence ostensibly in the name of Islam intensified around the world, a presidential election occurred in Nigeria, and the Nigerian army invaded the state of Borno and took

---

[12]In most states, the neighborhoods are randomly selected, but in 2015 some states, like Borno, the neighborhoods are restricted only to neighborhoods in urban areas due to safety concerns.

[13]In both surveys enumerator error sometimes resulted in more than 8 or 10 surveys to be conducted in a PSU.

[14]To account for this, we plan to also match 2015 non-viewers with 2013 non-viewers, and compare their change with the 2015 viewers-2013 non-viewers change in a difference-in-differences design. And we plan to create two sets of matches for the same 2015 viewers — one to 2015 non-viewers and one to 2013 respondents in general (all are non-viewers).

it back from control by Boko Haram. A 2015–2015 comparison matches very similar viewers and non-viewers in the same time context. However, the non-viewers may have selected out of viewing Arewa24 even though it is available and/or may be exposed to the effects of Arewa24 indirectly via other members of their social network or advertising.

In this paper we allow the 2015 Arewa24 viewers to form blocks with 2013 and 2015 non-viewers simultaneously. That is, each block can contain one or more Arewa24 viewer and one or more non-Arewa24 viewer, and the non-viewers can come from both 2013 and 2015. This allows us to generate a large number of very similar blocks. The benefit of this approach is that we only use each Arewa24 viewer once — sometimes compared with similar others from 2013, sometimes compared with similar others from 2015, and sometimes compared to a mixed set of non-viewers. As we report below, this leads to a well balanced research design. However, in the last minutes of writing this paper, we assessed the similarity of the 2013 only sets and the 2015 only sets and found that, considered alone, these sets were not well balanced. In order to compare what we learn from the two sets of comparisons, (that 2015 viewers and 2013 non-viewers may differ in a different way than 2015 viewers and 2015 non-viewers) we plan to create separate matched blocks for a 2015–2015 comparison and a 2013–2015 comparison. If Arewa24 viewers in 2015 and non-viewers in 2015 differ in the same way that Arewa24 viewers in 2015 and non-viewers in 2013 differ, then we are on firmer footing to claim that viewer–non-viewer differences are the consequence of Arewa24 viewing. However, if the 2015–2015 analysis has a different set of results than the 2015-2013 analysis, it is possible that differences arise from non-Arewa24 causes. The 2015-2013 differences could arise from the two year time difference, and the 2015–2015 differences could arise from selection into viewing Arewa24 where selection is driven by factors that we have not measured and are not strongly correlated with the factors that we measure.

We began by choosing covariates measured in both surveys to construct our matched sets. We identified 25 individual-level covariates that might affect viewing of Arewa24 but should not be influenced by viewing the channel itself, ranging from typical demographic questions like age, gender and socioeconomic status, to attitudes such as social identity and concern for the welfare of Muslims around the world. Rather than dropping observations that are missing data for one or more variables, we set missing observations to the covariate mean and create a separate variable for missingness on each covariate — thereby allowing us to match respondents on missingness as well as on the values taken by the covariates. Since television watching is a social activity in Nigeria (in part because not every family has a television and satellite dish), and because we expect that social interpretation of the channel and programming plays a role in attitude formation, we also included the neighborhood average of each of those variables calculated for each respondent after excluding his or her own score; that we also matched on the social context of the respondent. This yields 81 total traits for matching[15]. Of those 81 matching covariates, we believe two, gender and region of Nigeria (North West, North Wast, and North Central), are especially important for capturing unobservable and potentially confounding traits. We therefore required exact matches on gender and region of Nigeria.

Of the 81 possible covariates, a group had virtually no variation (i.e. about 25 terms had 90% or more of their values constant) and we dropped those terms (mostly consisting of the terms recording the proportion of missing responses) from consideration. We then further pre-screened the covariates that we included in the matching by doing a preliminary variable selection step using an adaptive lasso using 10-fold cross-validation, and model selection using the minimum mean squared error (cite).[16]

---

[15]Appendix 1 contains a full list of matched covariates.

[16]We chose an adaptive lasso because of its oracle property (cite) and its good performance in a number of areas (cite to Efron on standard errors, and to work on the elastic net, etc..). The first stage of the adaptive lasso was a lasso itself.

The lasso left 41 covariates, mainly dropping more of the missingness variables we had created at both the individual and PSU-level. We then created three distance matrices to record similarity between the roughly 2250 viewers and the roughly 18640 non-viewers (divided into sub-groups by gender and region): (1) a propensity score distance matrix using a Bayesian logistic regression (`bayesglm`) with all 41 covariates (cite `bayesglm`), (2) a propensity score distance matrix using the final adaptive lasso fit, and (3) a Mahalanobis distance matrix created from rank transforming the 41 covariates in an effort to give them all equal weight in this assessment of (dis)similarity on the covariates themselves without taking Arewa24 watching into account. The final match used calipers at the 10th percentile for the two different (but very similar) propensity score matrices and a caliper to exclude a few bad matches on the Mahalanobis distance matrix (using the 75th percentiles). The matching yields an effective sample size of about 2100, after dropping 14993 non-viewers and 2 viewers. The structure of the sets is shown in table 1: with "1:0" referring to a 1 treated unit matched with 0 control units (i.e. excluded) and 0:1 referring to number of dropped control units, 710 pairs ("1:1" sets), 117 sets with two treated and 1 control, and 154 sets with more than 5 controls (up to about 15) per treated unit.

Table 1: Structure of Matched Sets from Optimal, Full Matching on Propensity Score within Propensity Score and Mahalnobis Distance Calipers (cite)

Trted:Control	1:0	5:1	4:1	3:1	2:1	1:1	1:2	1:3	1:4	1:5+	0:1
# Sets	2	72	30	60	117	710	252	142	95	154	14993

Note: Matched sets from the 'fullmatch' R command in the 'optmatch' package.

Did this process produce sets that looked similar enough to allay concerns that, say, by comparing Arewa24 viewers and non-viewers we were really only telling a story about income or education (i.e. comparing those who can afford a television and a satellite dish versus those who cannot)? Figure 1 below shows the Arewa24 viewer/non-viewer differences before and after matching.

Before matching, Arewa24 viewers and non-Arewa24 viewers differ on nearly every covariate we measure. The Arewa24 viewers appear far more affluent on a variety of traits. For example, Arewa24 viewers tend to have more education and yearly income, and score higher on a socioeconomic status index asking about utilities they can access in their home[17]. Consequently, Arewa24 viewers are also more likely to live amongst other more affluent people in urban areas and engage in activities that require affluence, such as accessing the internet or, perhaps obviously, watching television. Less obviously, Arewa24 viewers differ in ways that are not clearly related to the ability to watch television. Arewa24 viewers report systematically different primary social identities than non-Arewa24 viewers and are more likely to be concerned with the welfare of Muslims around the world. Any of these differences could confound the effect of viewing Arewa24.

After matching, the blocks containing 2015 viewers are well balanced with non-viewers. Our standard for a good comparison in an observational study is the randomized experiment. We know that this is *not* a randomized experiment, however, if we can create a matched research design that is statistically indistinguishable from such an experiment, then we have a warrant to talk about the comparisons between viewers and non-viewers on tolerance, violence, and women's empowerment without fear that these differences should be driven by any of the *observed* variables that we matched on.

The idea being that we would prefer not to force balance on variables, especially binary variables, for which only a small number of cases had any non zero values. We used the `glmnet` package for R for this purpose. (cite glmnet)

[17]Running water, electricity, waste collection, mobile phones, etc...

3.4 Outcome Measures

Arewa24's programming is designed to affect four attitudes in particular: (1) intergroup tolerance, (2) support for violence, (3) support for women's empowerment, and (4) support for youth empowerment. The 2015 survey asks about all of these attitudes, but the 2013 survey includes no questions about youth empowerment. The measurement of support for women's empowerment is also complicated because, though the surveys both ask questions about women's empowerment, they ask *different* questions to measure support for women's empowerment. We are on firmer ground comparing questions on tolerance and violence. Both surveys contain six identical questions to measure intergroup tolerance and ten identical questions to measure support for violence. Those question lists are included in Appendix 2 and Appendix 3.
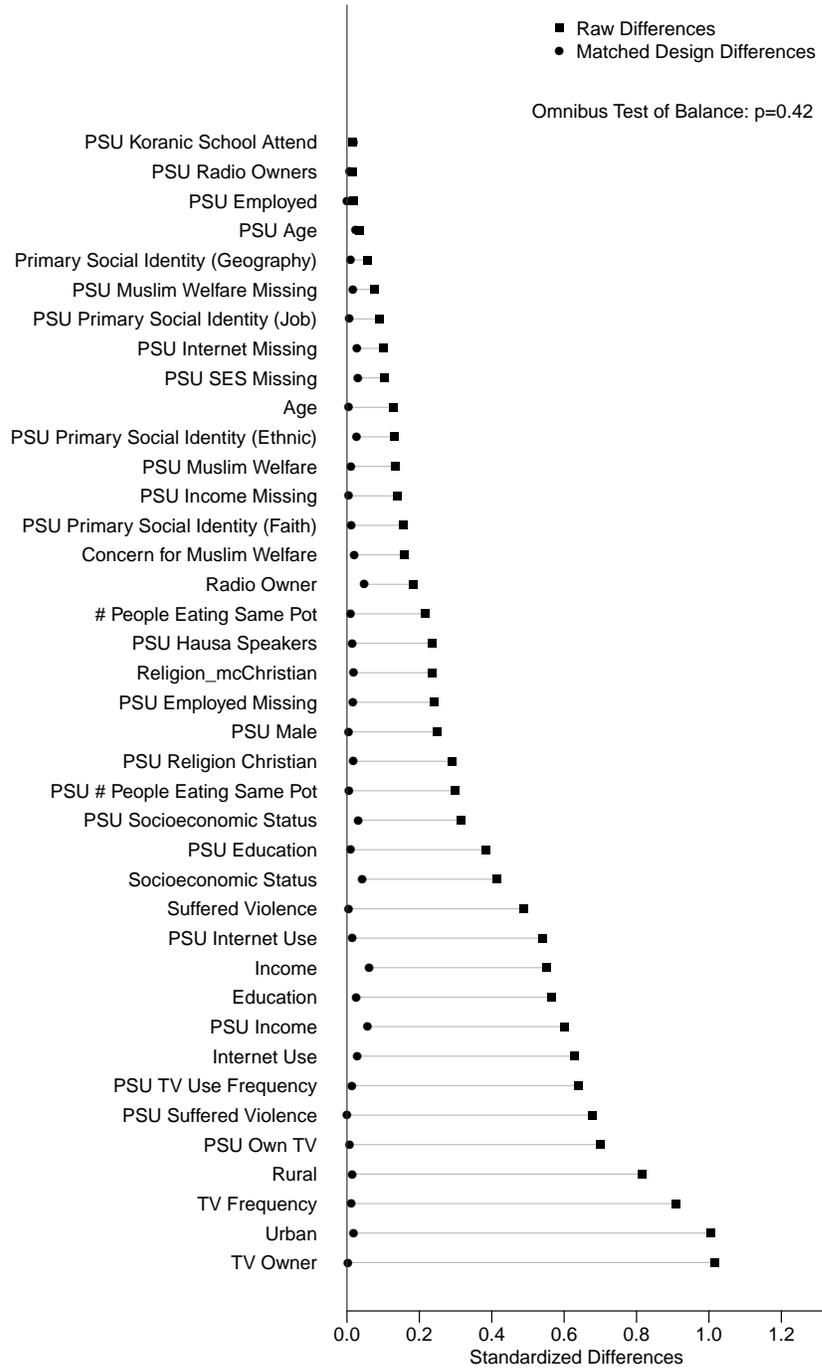
Figure 1: Standardized differences between Arewa24 watchers and non-watchers conditional on matched set (circles) or without matching but restricted to the matched sample (squares). The omnibus balance test assesses the hypothesis that the linear combination of all of these differences arose from a randomized experiment (Hansen and Bowers 2008).

4 Results

Figure 2 displays the estimated average effect of viewing Arewa24 conditional on matched set and with randomization justified standard errors and large sample confidence intervals (Lin 2013). The estimated treatment effects are shown over all of the matched sets, for the exclusively 2015 and 2013 subsets of the matched sets, and for the male and female subsets of the matched sets. Note, however, that the matched sets are only balanced in the aggregate. The year and gender subsets function to show the heterogeneity in the matched sets, but since they are not themselves balanced on covariates we should not take too seriously the observed estimates within subsets. The fact that there is substantial heterogeneity by year implies we should have separate matches for 2015-2013 controls and 2015–2015 controls: thereby using the same treated units in two different matched designs.
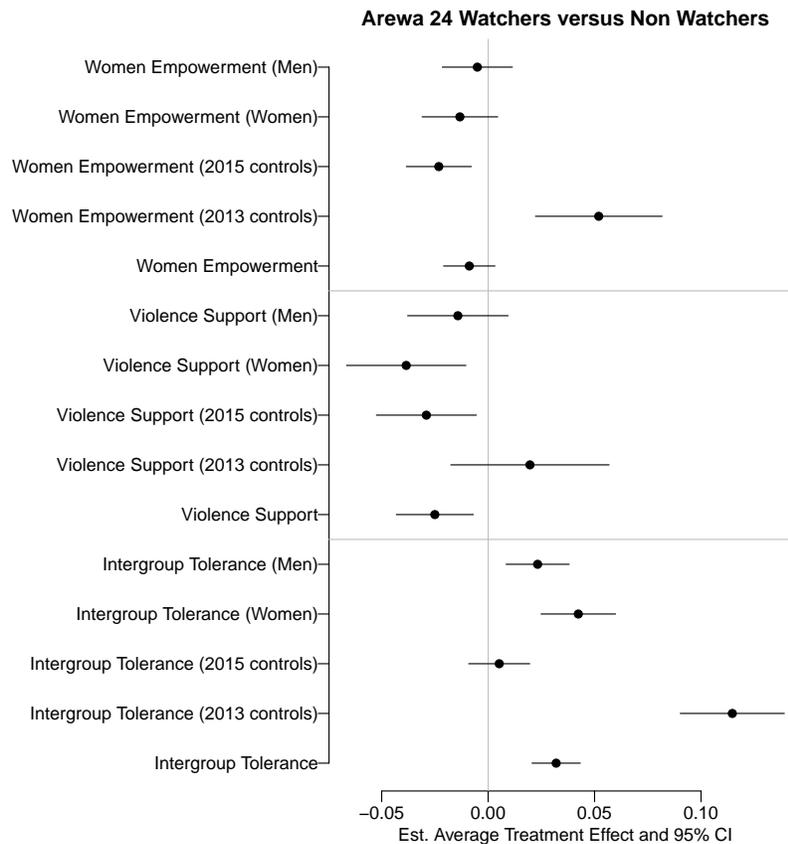

Figure 2: Estimated average effects of reporting Arewa24 watching (circles) and 95% confidence intervals. The analyses are all conditional on matched set and use randomization justified standard errors.

This analysis suggests that viewing Arewa24 affects both intergroup tolerance and support for violence, with no overall effect on support for women's empowerment. Arewa24 viewers in our matched sets are more tolerant than non-viewers, but viewers are also more supportive of violence. The effect on tolerance is driven by the 2015-2013 matches: 2015 viewers are much more tolerant than 2013 non-viewers. The effect on support for violence is driven by the 2015–2015 matched sets and by the female matched sets. Though there is no effect overall on women's empowerment, Arewa24 viewers in the exclusively 2015-2013 matched sets are significantly more tolerant than non-viewers, and Arewa24 viewers in the exclusively 2015–2015 matched sets are significantly *less* tolerant than

non-viewers.

The overall effect for intergroup tolerance could mean that Arewa24's pro-tolerance messaging is effective. But the tolerance heterogeneity by year could also imply that a shift in tolerance has taken place over time in northern Nigeria. Matched 2015 viewers are substantially more tolerant than 2013 non-viewers, but no more tolerant than 2015 non-viewers. It could also be that characteristics of the survey and survey enumeration lead to systematic differences in agreement or disagreement the outcome measurement questions[18].

The overall effect on support for violence could reflect a failure of Arewa24's anti-violence messaging. But the violence heterogeneity by year could also imply that people viewing Arewa24 are ex ante more supportive of violence despite their similarities displayed in Figure 1. The viewers in the exclusively 2015–2015 matched sets are more supportive of violence, but in the 2015-2013 matched sets viewers and non-viewers are equally supportive of violence. It could also be the case that characteristics of the survey confound the 2015-2013 comparison[19].

The lack of an overall effect for the women's empowerment questions, coupled with differences between 2015-2013 matched sets and 2015–2015 matched sets, precludes a strong conclusion about the effects of Arewa24 on support for women's empowerment. Based on the overall null effect we hesitantly conclude that viewing Arewa24 does not affect attitudes towards women's empowerment.

As a caveat, watching the Arewa24 channel should not necessarily affect any of these attitudes: Arewa24 airs many programs that are not designed to affect attitudes, though it is careful not to air programs that would inflame tensions or model poor behavior. Still, only one or two programs on Arewa24 seek to influence each of these attitudes, respectively, and without a measure of which programs people are watching, we must employ viewing Arewa24 as a reasonable proxy for viewing the attitude change programs, acknowledging the noise inherent in this measure.

5 Sensitivity Analysis

In any observational study, unobservables may correlate with both the outcome and the causal variable, and we may mis-attribute the effect of those unobservables as effects of the predictor — the non-confounding assumption. Sensitivity analysis helps us determine if those unobservables are strong enough to wipe out our treatment effect. How strongly correlated would unobservables have to be with the treatment, and how predictive of the outcome, for them to completely explain our treatment effect?

We will use the Hosman, Hansen, and Holland (2010) method to calibrate the aspects of the confounder with other covariates that predict the outcome.

6 Two Preanalysis Plans

We will support these observational results with two additional studies: one encouragement experiment, and one pre-post survey study. In the encouragement design we will randomly encourage

---

[18]Here we believe question ordering effects are unlikely, since the set of tolerance questions are preceded by identical question sets in both surveys. However, the 2013 tolerance series has three extra questions asking about intergroup marriage that do not appear in the 2015 survey. These could have changed how the other questions were answered, though we have no strong theoretical reason to believe they would make the respondent less tolerant when asked other questions. If there is a survey difference confounding our results, it's more likely from enumerator effects or, potentially, differences in sampling strategy.

[19]Again, we believe question ordering effects unlikely, because the survey's format is largely unchanged from 2013–2015 and because violence questions are preceded by the same question sets in both surveys.

subjects in and around the major cities of northwestern Nigeria to watch Arewa24's flagship soap opera (the treatment group) or a rival northern-Nigerian network (the placebo group); a control group will not be encouraged to watch any TV channel or program. In the pre-post survey study, 60 neighborhoods were surveyed in the first months Arewa24 began airing programs and those same 60 neighborhoods were surveyed again after Arewa24 has aired programs for two years; in some cases, the same people were surveyed.

6.1 A preanalysis plan for an Encouragement Design

For this analysis we employ an "intent to treat" framework, where the treatment is being encouraged to view Arewa24's flagship soap opera, Dadin Kowa. We call this "intent to treat" because we focus on the effect of being *assigned* to the treatment group of this study, rather than analyzing the effect of actually watching the shows. After a preliminary round of of interactive voice response (IVR) calls to collect information such as gender, geographic location, previous exposure to Arewa24, and consent to participate in the study, we will randomly assign each of the subjects to one of the three arms of the study within blocks defined by gender[20]. Respondents will complete one short survey of about 4–5 questions per week, administered by interactive voice response system (IVR). The surveys will contain 1–2 questions to measure outcomes of interest per week and 3–4 other questions designed to distract respondents from the purpose of the study and to gather other useful information. The study will run for approximately 25 weeks.

We have four topics: (1) Tolerance, (2) Violence, (3) Women, and (4) Youth. Key questions about each topic will be asked twice during the 25 week study. We will conduct two analyses for each question: First we merely compare the responses to the second question between the treatment arms. Second, we compare change overtime from the first to the second question across the treatment arms. This second analysis has the advantage of reducing noise in the comparisons across treatment arms and by focusing on change due to the experiment. The disadvantage is that not all respondents will answer all of the outcome questions every week, so we will lose observations, and thus lower statistical power and raise concerns about non-random attrition. In general we will present estimates of average treatment effects along with tests of the hypothesis of no effects.

In a pilot test and in other surveys we have conducted in Africa, we find strong acquiescence bias for survey questions. To counter this, we will randomly assign subjects to receive reverse-scale questions in which the "socially desirable" response comes last instead of first. Subjects will receive the same scale the first and second time the question is asked[21].

The questions will be analyzed individually and combined into indices to measure the four main concepts we test (tolerance, violence, women, and youth attitudes). We will need to account for multiple outcomes, since we will have over a dozen outcome variables. We will need to do that in a way that can handle missing outcomes, since many respondents will not reply every week. We currently plan to use Rosenbaum-style sensitivity analysis to engage with the problem of missingness. The idea of double-sampling (Aronow et al. 2013) is attractive here, but we are not sure how to implement it when we are gathering outcomes at many time points via the mobile phone and recorded

---

[20]We have not determined the number of subjects. Ideally we would have 10,000 respondents to detect small differences between groups based on a pilot study with about 100 subjects completed in late 2016.

[21]We think we will pre-assign these conditions to respondents at the beginning of the study so that we can treat them as blocks. However, with all the permutations that can arise from 18 normal or reverse scale questions, that will result in a very large number of blocks. We could also randomly assign a normal or reverse scale the first time the question is asked, but then the groups may not be balanced on normal/reverse scale, or subjects will have unequal probability of being assigned to a normal scale or reverse scale. We are thinking of other ways of randomly assigning these questions.

voice based response system.

We plan a few analyses for these data. First, we we will use the Caughey, Dafoe, and Seawright (2017) non-parametric combinations (NPC) method to generate global-p values to assess the null of no effects *on any outcome*. It accounts for both the probability of generating a p-value < 0.05 by chance when testing multiple outcomes, and the probability of generating multiple p-values near to 0.05 by chance when testing multiple outcomes.

Then we estimate average treatment effects using randomization justified standard errors, covariance adjustment using Rosenbaum's (2002??) method, covariance adjustment using the Lin & Green (year??) method from the Green Lab SOP, and a Tukey test to account for the multiple experimental groups we have[22].

We have run a pilot study, and we imagine that the pilot data are a good approximation of the data we will obtain in a larger study. In the code chunks below we demonstrate our analyses on our pilot data. Before conducting these analyses, we confirm that the coverage rates we use are correct by showing that our estimators produce p-values below 0.05 less than 5% of the time on when the null hypothesis of no effects is true.

R code examples for our Encouragement Design preanalysis plan

```
# scan in cleaned pilot data (need to make this not my computer only)
load("C:/Users/cdgra/Dropbox/chris_wrk/NigeriaTV/IVRDesign/Analysis/df.rda")
df$row <- rownames(df)

# load libraries for hc2, NPC, others
library(lmtest)
library(sandwich)
library(NPC)
library(multcomp)
```

First, we evaluate coverage of a simple linear model and the canned p-value.

```
######
# Coverage Tests
######
# lm clt/iid coverage
lm.shuf <- function(dat=df,outcome,treatment='exp_group')
{
  dat[["newz"]] <- sample(dat[[treatment]])
  lm1 <- lm(reformulate('newz', response=outcome),data=dat)
  pval <- summary(lm1)$coefficients["newzdadin_ivr",4]
  return(pval)
}
cov.test2 <- function()
{
  cov.test <- replicate(1000,lm.shuf(outcome="trust_rel_diff"))
  return(mean(cov.test>=0.05))
```

---

[22]We expect to have a large N and so for the central limit theorem to apply, and we can test this with randomization tests if needed.

```
}
testing.lm <- replicate(10,cov.test2())
mean(testing.lm)
```

Next we evaluate coverage of a simple linear model using randomization justified standard errors. As expected, the coverage also generates p-values at or above 0.05 about 95% of the time.

```
# hc2 coverage
hc2.shuf <- function(dat=df,outcome,treatment='exp_group')
{
  dat[["newz"]] <- sample(dat[[treatment]])
  lm1 <- lm(reformulate('newz', response=outcome),data=dat)
  pval <- coeftest(lm1,vcov=vcovHC(lm1,type="HC2"))['newzdadin_ivr',4]
  return(pval)
}
hc2.test <- function()
{
  cov.test <- replicate(1000,hc2.shuf(outcome="trust_rel_diff"))
  return(mean(cov.test>=0.05))
}
testing.hc2 <- replicate(10,hc2.test())
mean(testing.hc2)
```

Moving on from a simple linear model, we evaluate coverage of NPC, which tests multiple hypotheses simultaneously. This code is extremely time intensive and is still running on a small cluster computer.

```
# NPC coverage (note: this takes a long time to run.  Like, days.)

# global-p with harmonic weighted means
npc.shuf <- function(dat=df,outcome,treatment='exp_group')
{
  dat[["newz"]] <- sample(dat[[treatment]])
  npc=NPC(dat, tr.var="newz",tr.label='dadin_ivr',
          block.var="blocks",
          y.vars=outcome,seed=round(rnorm(1,1000,500)),
          n.perms=500, test.statistic="HarmonicWtdMean", alternative="greater",return.matrix=
  pval <- npc$p.values["NPC"]
  return(pval)
}
#system.time(replicate(1,npc.shuf(outcome=c("trust_ethnic_second","trust_denom_second","trust
#                     "marry_ethnic_second","marry_denom_second","marry_rel_second"))))
npc.test <- function()
{
  cov.test <- replicate(600,npc.shuf(outcome=c("trust_ethnic_second","trust_denom_second","tr
                        "marry_ethnic_second","marry_denom_second","marry_rel_second")))
  return(mean(cov.test>=0.05))
}
system.time(testing.npc <- replicate(2,npc.test()))
mean(testing.npc)
```

Finally, we use a Tukey test to account for our multiple comparisons. We have three experimental groups (treatment, placebo, control) and so three differences of means being tested for each question. The Tukey test generates p-values of 0.05 or above about 98% of the time.

```
# Tukey coverage
tuk.shuf <- function(dat=df,outcome,treatment='exp_group')
{
  dat[["newz"]] <- sample(dat[[treatment]])
  tr.aov  <-  aov(dat[[outcome]]~newz+tr.blockvar, dat)
  tr.tukey  <-  glht(tr.aov, linfct = mcp(newz = "Tukey"))
  pval <- summary(tr.tukey)$test$pvalues[1:3]
  return(pval)
}
#test <- replicate(1000,tuk.shuf(outcome='trust_rel_diff'))
tuk.test <- function()
{
  cov.test <- replicate(1000,tuk.shuf(outcome="trust_rel_diff"))
  return(mean(cov.test>=0.05))
}
system.time(testing.tukey <- replicate(10,tuk.test()))
mean(testing.tukey)
```

Here we include the code we plan to use for analyses.

```
# global-p with harmonic weighted means
npc.tol=NPC(df, tr.var="exp_group",tr.label='dadin_ivr',
        block.var="blocks",
    y.vars=c("trust_ethnic_second","trust_denom_second","trust_rel_second",
            "marry_ethnic_second","marry_denom_second","marry_rel_second"),
    n.perms=1000, test.statistic="HarmonicWtdMean", alternative="greater",return.matrix=TRUE)
npc.tol$p.values
```

Next we have Rosenbaum style covariance adjustment.

```
# lm predicting outcomes with exp_group + blocks
lm.tr <- lm(trust_rel_diff~exp_group+tr.blockvar,df)
coeftest(lm.tr,vcov=vcovHC(lm.tr,type="HC2"))

# rosenbaum method for covariance adjustment in experiments
## add residual as a column in the df
lm.trust_rel_resid <- resid(lm(trust_rel_diff~gender2+trust_rel_group+state,df))
df[names(lm.trust_rel_resid),"trust_rel_resid"] <- lm.trust_rel_resid
stopifnot(df[df$row==8,"trust_rel_resid"]==lm.trust_rel_resid["8"])
#with(df,tapply(trust_rel_resid,tr.blockvar,mean,na.rm=TRUE)) # block averages are 0.
lm.tr_covar <- lm(trust_rel_resid~exp_group+gender2,data=df)
coeftest(lm.tr_covar, vcov = vcovHC(lm.tr_covar, type = "HC2"))
```

And the Lin & Green style covariance adjustment.

```r
# winston lin method for covariance adjustment
dat$Ymd  <-  dat$Y - ave(dat$Y,dat$BlockFactor)
dat$Zmd  <-  dat$Z - ave(dat$Z,dat$BlockFactor)
lm1a  <-  lm(Ymd~Zmd, dat)
lm1b  <-  lm(Y~Z+BlockFactor, dat)
stopifnot(require(lmtest))
stopifnot(require(sandwich))
coeftest(lm1, vcov(lm1, "HC2"))
### from green lab sop
# Mean-center the covariates
X1_c  <-  X1 - mean(X1)
X2_c  <-  X2 - mean(X2)
# Conduct Estimation
fit_adj  <-  lm(Y_obs ~ Z + Z*(X1_c + X2_c), singular.ok = FALSE)
# Robust Standard Errors
coeftest(fit_adj, vcov = vcovHC(fit_adj, type = "HC2"))
# Compare to unadjusted model
fit_unadj  <-  lm(Y_obs ~ Z, singular.ok = FALSE)
coeftest(fit_unadj, vcov = vcovHC(fit_unadj, type = "HC2"))

# tukey test for multiple comparisons
tr.aov  <-  aov(trust_rel_diff~exp_group+tr.blockvar, df)
tr.tukey  <-  glht(tr.aov, linfct = mcp(exp_group = "Tukey"))
summary(tr.tukey)
```

Compliance analysis for our Encouragement Design preanalysis plan

In any encouragement design, we cannot guarantee that subjects randomly assigned to treatment actually receive the treatment. Thus, we will measure *compliance* with treatment. What percentage of the treatment group actually received the treatment? In this longitudinal ITT experiment, compliance patterns are complex because subjects have the opportunity to comply or not comply for 25 weeks. We will measure compliance with a self-reported yes/no question each week and with a question about the content of the shows every fifth week.

In this study, compliance itself is a policy relevant outcome: we wonder whether those encouraged to what Dadin Kowa were more likely to report actually watching it and/or answer the factual show content question more accurately than those in the other conditions.

In some cases, we can use a placebo to learn about the effect of actually watching the assigned show: if those in the placebo group comply at the same rates as those in the treatment group, and the outcome in the placebo group are the same as the outcomes in the control group, then the difference in outcomes among compliers in the treated and placebo group tell us about the effect of watching. There are a lot of "if" statements that we will have to confront before we can feel comfortable using the placebo group in this way. However, we design the study with a placebo group so that we have the opportunity to ask this kind of question.

We do not plan to estimate the effect of treatment on the compliers in this study because there are so many kinds of compliance pattens, each with a very small number of cases.

```
# measuring compliance
lmY1<-lm(ismissingonY1~exp_group,df)
anovaY1<-anova(Y1)
anovaY1$p.value
```

6.2 A preanalysis plan for an Observational Design of a Baseline versus an Endline design

In our pre/post survey study, we will compare (1) neighborhoods at the onset of Arewa24 programming with neighborhoods after two years of Arewa24 programming, (2) the same individual over time (if we are lucky enough to find such people), and (3) viewers to non-viewers over time. We also believe that community leadership could be important in northern Nigeria, and so we surveyed the traditional leader of each community for which we surveyed.

Since we have GPS coordinates for every PSU and for most surveys, we will also test if there seems to be diffusion between individuals and between PSUs.

7 Conclusion

[To Do]

8 Appendices

8.1 Appendix 1: List of covariates

Initial Covariates (variable names, not nice labels) are shown here. The covariates selected for use in matching are listed in Figure 1.

```
load("../Soccom_Code/socom_2013/matchingdata.rda")
paste(names(psdat))
```

```
 [1] "agenum"                                "educ"
 [3] "employment2Working"                    "genderMale"
 [5] "hausa"                                 "identityA.person.from.your.ethnic.grou
 [7] "identityA.person.from.your.faith"      "identityA.person.from.your.town.or.reg
 [9] "identityA.person.of.your.job.or.occupation" "identityNone.of.these"
[11] "inc2"                                  "int"
[13] "koranic"                               "muslimwelfare"
[15] "radio"                                 "regionNE"
[17] "regionNW"                              "religion_mcChristian"
[19] "rural.urbanRural"                      "rural.urbanUrban"
[21] "samepot"                               "ses"
[23] "suffvio"                               "tv"
[25] "tvfreq"                                "agenum.NA"
[27] "educ.NA"                               "employment2.NA"
[29] "identity.NA"                           "inc2.NA"
[31] "int.NA"                                "koranic.NA"
[33] "muslimwelfare.NA"                      "radio.NA"
[35] "religion_mc.NA"                        "samepot.NA"
```

```
[37] "ses.NA"                                              "suffvio.NA"
[39] "tv.NA"                                               "tvfreq.NA"
[41] "psu"                                                 "agenum_psu"
[43] "educ_psu"                                            "employment2Working_psu"
[45] "genderMale_psu"                                      "hausa_psu"
[47] "identityA.person.from.your.ethnic.group_psu"         "identityA.person.from.your.faith_psu"
[49] "identityA.person.from.your.town.or.region_psu"       "identityA.person.of.your.job.or.occupa
[51] "identityNone.of.these_psu"                           "inc2_psu"
[53] "int_psu"                                             "koranic_psu"
[55] "muslimwelfare_psu"                                   "radio_psu"
[57] "regionNE_psu"                                        "regionNW_psu"
[59] "religion_mcChristian_psu"                            "rural.urbanRural_psu"
[61] "rural.urbanUrban_psu"                                "samepot_psu"
[63] "ses_psu"                                             "suffvio_psu"
[65] "tv_psu"                                              "tvfreq_psu"
[67] "agenum.NA_psu"                                       "educ.NA_psu"
[69] "employment2.NA_psu"                                  "identity.NA_psu"
[71] "inc2.NA_psu"                                         "int.NA_psu"
[73] "koranic.NA_psu"                                      "muslimwelfare.NA_psu"
[75] "radio.NA_psu"                                        "religion_mc.NA_psu"
[77] "samepot.NA_psu"                                      "ses.NA_psu"
[79] "suffvio.NA_psu"                                      "tv.NA_psu"
[81] "tvfreq.NA_psu"                                       "ps2"
[83] "a24"                                                 "region"
[85] "fm0"
```

8.2 Appendix 2: 2015 Survey Question List

Tolerance

Six questions measure tolerance in the 2015 SOCOM survey. Higher scores indicate a greater level of tolerance for those with different religious and ethnic backgrounds.

Please tell me how comfortable or uncomfortable you would be in accepting someone from a **religious group different than your own** as each of the following:

1. A co-worker/classmate
2. A neighbor

Please tell me how comfortable or uncomfortable you would be in accepting someone from **your same religion, but a different sect/denomination** as each of the following:

3. A co-worker/classmate
4. A neighbor

Please tell me how comfortable or uncomfortable you would be in accepting someone from an **ethnic group different than your own** as each of the following:

5. A co-worker/classmate
6. A neighbor

Violence

Twelve questions measure violence in the 2015 SOCOM survey. Higher scores indicate less acceptance of violent means. Note that questions 5-6 from this survey are not included in the 2013 survey, and that the 2013 survey has an additional question (13).

Please tell me if you strongly agree, agree, somewhat agree, somewhat disagree, disagree, or strongly disagree with each of the following statements :

1. The use of violence is justified when defending ones religion
2. The use of violence is justified to maintain our society's culture and traditions
3. The use of violence is justified to bring criminals to justice when the government fails to act
4. The use of violence is justified to force the government to change their policies or actions
5. Followers of a religion should not react with violence if violence is used against its members
6. People of a different religion than mine are responsible for more violence in this country
7. I would support a family member's decision to join a group that uses violence to defend my religion.
8. I would support a close friends decision to join a group that uses violence to defend my religion
9. Those who use violence in the name of religion, if killed, are guaranteed personal salvation.
10. While the current level of violence in Nigeria is regrettable, violence will eventually produce necessary changes in society.
11. It is important to support guerrilla fighters in Islamic countries fighting on behalf of Muslims
12. It is important to support charities that fund groups like al Qaeda

Women's Empowerment

Five questions measure support for women's empowerment on the 2015 SOCOM survey. Higher score indicates more support for empowerment of women. There is no overlap with the 2015 questions about women's empowerment.

I am now going to read you some statements about women and youth. Please tell me if you strongly agree, agree, somewhat agree, somewhat disagree, disagree, or strongly disagree with each of the following statements:

1. Women and girls are denied opportunities based on their gender in my community
2. The lack of opportunities for women and girls is a problem in my community.
3. There is enough respect for women's ideas and decisions in my community
4. Too much credence is given to women's ideas in my community.
5. Education is more important for a boy than for a girl.

Youth Empowerment

Three questions measure support for youth empowerment on the 2015 SOCOM survey. Higher score indicates more awareness of the negative position of youth/more support for youth empowerment.

I am now going to read you some statements about women and youth. Please tell me if you strongly agree, agree, somewhat agree, somewhat disagree, disagree, or strongly disagree with each of the following statements:

1. Youth are denied opportunities in my community.
2. The lack of opportunities for youth is a problem in my community
3. There is enough respect for youth ideas and decisions in my community.

8.3 Appendix 3: 2013 Survey Question List

Tolerance

Nine questions measure tolerance in the 2013 SOCOM survey. Higher scores indicate a greater level of tolerance for those with different religious and ethnic backgrounds. Note that the third section in each sequence was not included in the 2015 SOCOM survey.

Please tell me how comfortable or uncomfortable you would be in accepting someone from a **religious group different than your own** as each of the following:

1. A co-worker/classmate
2. A neighbor
3. A close relative by marriage

Please tell me how comfortable or uncomfortable you would be in accepting someone from **your same religion, but a different sect/denomination** as each of the following:

4. A co-worker/classmate
5. A neighbor
6. A close relative by marriage

Please tell me how comfortable or uncomfortable you would be in accepting someone from an **ethnic group different than your own** as each of the following:

7. A co-worker/classmate
8. A neighbor
9. A close relative by marriage

Violence

Eleven questions measure violence in the 2013 SOCOM survey. Higher scores indicate less acceptance of violent means. Note that questions 5-6 from the 2015 survey are not included here, and that the 2013 survey has an additional question (13).

Please tell me if you strongly agree, somewhat agree, somewhat disagree, or strongly disagree with each of the following statements :

1. The use of violence is justified when defending ones religion
2. The use of violence is justified to maintain our society's culture and traditions
3. The use of violence is justified to bring criminals to justice when the government fails to act
4. The use of violence is justified to force the government to change their policies or actions
5. I would support a family member's decision to join a group that uses violence to defend my religion.
6. I would support a close friends decision to join a group that uses violence to defend my religion
7. Those who use violence in the name of religion, if killed, are guaranteed personal salvation.
8. While the current level of violence in Nigeria is regrettable, violence will eventually produce necessary changes in society.
9. It is important to support guerrilla fighters in Islamic countries fighting on behalf of Muslims
10. It is important to support charities that fund groups like al Qaeda
11. Do you think that using violence in the name of your religion is always justified, sometimes justified, rarely justified, or never justified?

Women's Empowerment

Two questions measure support for women's empowerment on the 2013 SOCOM survey. Higher score indicates more support for empowerment of women. There is no overlap with the 2015 questions about women's empowerment.

I am now going to read you some statements about women and youth. Please tell me if you strongly agree, somewhat agree, somewhat disagree, or strongly disagree with each of the following statements:

6. Women should be allowed to work outside the home
7. It is a man's right and duty to maintain order in his family by whatever means necessary

References

Ahn, Woo-kyoung, and Charles W Kalish. 2000. "The Role of Mechanism Beliefs in Causal Reasoning." *Explanation and cognition*: 199–225.

Aronow, Peter M, Alan S Gerber, Donald P Green, and Holger Kern. 2013. "Double Sampling for Missing Outcome Data in Randomized Experiments." *Typescript, Yale University. URL: http://papers. ssrn. com/sol3/papers. cfm.*

Bandura, Albert. 1971. "Vicarious and Self-Reinforcement Processes." *The nature of reinforcement* 228278.

———. 2001. "Social Cognitive Theory of Mass Communication." *Media psychology* 3(3): 265–99.

Bandura, Albert, and David C McClelland. 1977. "Social Learning Theory."

Bandura, Albert, and Frances L Menlove. 1968. "Factors Determining Vicarious Extinction of Avoidance Behavior Through Symbolic Modeling." *Journal of personality and social psychology* 8(2p1): 99.

Bandura, Albert, Jennings Bryant, and Dolf Zillmann. 1994. "Media Effects: Advances in Theory and Research." *Social Cognitive Theory of Mass Communication. Hillsdale, NJ: Lawrence Erlbaum Associates*: 61–90.

Bandura, Albert, Dorothea Ross, and Sheila A Ross. 1963. "Imitation of Film-Mediated Aggressive Models." *The Journal of Abnormal and Social Psychology* 66(1): 3.

Bankart, C Peter, and Clark C Anderson>. 1979. "Short-Term Effects of Prosocial Television Viewing on Play of Preschool Boys and Girls." *Psychological Reports* 44(3): 935–41.

Berg, Gunhild, and Bilal Zia. 2013. "Harnessing Emotional Connections to Improve Financial Decisions: Evaluating the Impact of Financial Education in Mainstream Media."

Blaisdell, Aaron P, Kosuke Sawa, Kenneth J Leising, and Michael R Waldmann. 2006. "Causal Reasoning in Rats." *Science* 311(5763): 1020–2.

Brewer, Marilynn B. 1999. "The Psychology of Prejudice: Ingroup Love and Outgroup Hate?" *Journal of social issues* 55(3): 429–44.

Caughey, Devin, Allan Dafoe, and Jason Seawright. 2017. "Nonparametric Combination (Npc): A Framework for Testing Elaborate Theories." *The Journal of Politics* 79(2): 000–000.

Crigler, Ann N, Marion Just, and W Russell Neuman. 1994. "Interpreting Visual Versus Audio

Messages in Television News." *Journal of communication* 44(4): 132–49.

Drabman, Ronald S, and Margaret H Thomas. 1974. "Does Media Violence Increase Children's Toleration of Real-Life Aggression?" *Developmental Psychology* 10(3): 418.

Eller, Anja et al. 2011. "Vicarious Intergroup Contact Effects: Applying Social-Cognitive Theory to Intergroup Contact Research." *Group Processes & Intergroup Relations* 14(2): 255–74.

Festinger, Leon. 1962. 2 *A Theory of Cognitive Dissonance*. Stanford university press.

Gibson, James L, and Amanda Gouws. 2005. *Overcoming Intolerance in South Africa: Experiments in Democratic Persuasion*. Cambridge University Press.

Gubler, Joshua R. 2013. "When Humanizing the Enemy Fails: The Role of Dissonance and Justification in Intergroup Conflict." In *Annual Meeting of the American Political Science Association*,

Hansen, B.B., and J. Bowers. 2008. "Covariate Balance in Simple, Stratified and Clustered Comparative Studies." *Statistical Science* 23: 219.

Hearold, Susan, and Gerald Comstock. 1986. "A Synthesis of 1043 Effects of Television on Social Behavior." *Public communication and behavior* 1: 65–133.

Hogarth, Robin M, Brian J Gibbs, Craig R McKenzie, and Margaret A Marquis. 1991. "Learning from Feedback: Exactingness and Incentives." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 17(4): 734.

Hosman, Carrie A, Ben B. Hansen, and Paul W. Holland. 2010. "The Sensitivity of Linear Regression Coefficients's Confidence Limits to the Omission of a Confounder." *The Annals of Applied Statistics* 4(2): 849–70.

Howard, Ross, Francis Rolt, Hans Van de Veen, and Juliette Verhoeven. 2003. "The Power of the Media." *A handbook for peacebuilders. Utrecht: European Centre for Conflict Prevention*.

Jensen, Robert, and Emily Oster. 2009. "The Power of Tv: Cable Television and Women's Status in India." *The Quarterly Journal of Economics* 124(3): 1057–94.

Kazdin, Alan E. 1974a. "Covert Modeling, Model Similarity, and Reduction of Avoidance Behavior." *Behavior Therapy* 5(3): 325–40.

———. 1974b. "Effects of Covert Modeling and Model Reinforcement on Assertive Behavior." *Journal of Abnormal Psychology* 83(3): 240.

———. 1975. "Covert Modeling, Imagery Assessment, and Assertive Behavior." *Journal of Consulting and Clinical Psychology* 43(5): 716.

———. 1976. "Effects of Covert Modeling, Multiple Models, and Model Reinforcement on Assertive Behavior." *Behavior Therapy* 7(2): 211–22.

Kunda, Ziva. 1990. "The Case for Motivated Reasoning." *Psychological bulletin* 108(3): 480.

La Ferrara, Eliana, Alberto Chong, and Suzanne Duryea. 2012. "Soap Operas and Fertility: Evidence from Brazil." *American Economic Journal: Applied Economics* 4(4): 1–31.

Lin, Winston. 2013. "Agnostic Notes on Regression Adjustments to Experimental Data: Reexamining Freedman's Critique." *The Annals of Applied Statistics* 7(1): 295–318.

Lovelace, Valeria O, and Aletha C Huston. 1983. "Can Television Teach Prosocial Behavior?"

Prevention in Human Services 2(1-2): 93–106.

Loye, David, Roderic Gorney, and Gary Steele. 1977. "An Experimental Field Study." *Journal of Communication* 27(3): 206–16.

Malamuth, Neil M, and James VP Check. 1981. "The Effects of Mass Media Exposure on Acceptance of Violence Against Women: A Field Experiment." *Journal of Research in Personality* 15(4): 436–46.

Mares, Marie-Louise, and Zhongdang Pan. 2013. "Effects of Sesame Street: A Meta-Analysis of Children's Learning in 15 Countries." *Journal of Applied Developmental Psychology* 34(3): 140–51.

McGaugh, James L, Larry Cahill, and Benno Roozendaal. 1996. "Involvement of the Amygdala in Memory Storage: Interaction with Other Brain Systems." *Proceedings of the National Academy of Sciences* 93(24): 13508–14.

Molitor, Fred, and Kenneth William Hirsch. 1994. "Children's Toleration of Real-Life Aggression After Exposure to Media Violence: A Replication of the Drabman and Thomas Studies." *Child Study Journal*.

Ortiz, Michelle, and Jake Harwood. 2007. "A Social Cognitive Theory Approach to the Effects of Mediated Intergroup Contact on Intergroup Attitudes." *Journal of Broadcasting & Electronic Media* 51(4): 615–31.

Paluck, Elizabeth Levy. 2009. "Reducing Intergroup Prejudice and Conflict Using the Media: A Field Experiment in Rwanda." *Journal of personality and social psychology* 96(3): 574.

Paluck, Elizabeth Levy, and Donald P Green. 2009a. "Deference, Dissent, and Dispute Resolution: An Experimental Intervention Using Mass Media to Change Norms and Behavior in Rwanda." *American Political Science Review* 103(04): 622–44.

———. 2009b. "Prejudice Reduction: What Works? A Review and Assessment of Research and Practice." *Annual review of psychology* 60: 339–67.

Piotrow, Phyllis T et al. 1990. "Mass Media Family Planning Promotion in Three Nigerian Cities." *Studies in family planning* 21(5): 265–74.

Scacco, Alexandra, and Shana S Warren. 2016. "Youth Vocational Training and Conflict Mitigation: An Experimental Test of Social Contact Theory in Nigeria."

Schiappa, Edward, Peter B Gregg, and Dean E Hewes. 2005. "The Parasocial Contact Hypothesis." *Communication Monographs* 72(1): 92–115.

Shatzer, Milton J, Felipe Korzenny, and Betty Ann Griffis-Korzenny. 1985. "Adolescents Viewing Shogun: Cognitive and Attitudinal Effects." *Journal of Broadcasting & Electronic Media* 29(3): 341–46.

Singhal, Arvind, Michael J Cody, Everett M Rogers, and Miguel Sabido. 2003. *Entertainment-Education and Social Change: History, Research, and Practice*. Routledge.

Singhal, Arvind, Everett M Rogers, and William J Brown. 1993. "Harnessing the Potential of Entertainment-Education Telenovelas." *International Communication Gazette* 51(1): 1–18.

Skinner, Burrhus Frederic. 1990. *The Behavior of Organisms: An Experimental Analysis*. BF Skinner

Foundation.

Thorndike, Edward L. 1905. "The Functions of Mental States."

Vallone, Robert P, Lee Ross, and Mark R Lepper. 1985. "The Hostile Media Phenomenon: Biased Perception and Perceptions of Media Bias in Coverage of the Beirut Massacre." *Journal of personality and social psychology* 49(3): 577.

W. Vaughan, Arvind Singhal Ramadhan M. Swalehe Peter, Everett M. Rogers. 2000. "Entertainment-Education and Hiv/Aids Prevention: A Field Experiment in Tanzania." *Journal of health Communication* 5(sup1): 81–100.

Wimmer, Heinz, and Josef Perner. 1983. "Beliefs About Beliefs: Representation and Constraining Function of Wrong Beliefs in Young Children's Understanding of Deception." *Cognition* 13(1): 103–28.