

Building Tolerance: Intergroup Contact and Soccer in Post-ISIS Iraq

Salma Mousa^{*}

September 13, 2019

Abstract

Can intergroup contact build social cohesion after war? I answer this question by randomly assigning Iraqi Christians displaced by ISIS either to an all-Christian soccer team or to a team mixed with Muslims. I find persistent changes to behaviors: Christians assigned to mixed teams are 12 percentage points more likely to attend a social event open to Muslims, 54 percentage points more likely to train with Muslims six months after the intervention ends, and 25 percentage points more likely to vote for a Muslim player to receive a sportsmanship award. These results seem to be driven by changing norms around social contact as well as a positive experience, with top-performing teams being more likely to patronize a restaurant in Muslim-dominated Mosul. The impact on personal beliefs, however, was mixed. These findings point to the potential for positive and cooperative contact across social lines to build tolerant behaviors after conflict — even if underlying prejudice remains unchanged.

Key Words: Conflict, intergroup contact, migration, Middle East

^{*}Department of Political Science, Stanford University; smousa@stanford.edu. I thank Rabie Zakaria for project management throughout, and Kellsey Beal for project management during the pilot. I also thank Ala' Alrababa'h, Lisa Blaydes, Simon Ejdemyr, Jens Hainmueller, David Laitin, Joshua Kalla, Jonathan Mummolo, Christiana Parreira, Jeremy Weinstein, and seminar participants at NYUAD's WESSI workshop, the Stanford Law School, the Immigration Policy Lab, and the Empirical Studies of Conflict project for comments on earlier versions of this paper. I gratefully acknowledge the support of the United States Institute of Peace, the Stanford Center for International Conflict and Negotiation, the Stanford Center for Stanford Center on Global Poverty and Development and the Freeman Spogli Institute for the study pilot, and the Abdul Latif Jameel Poverty Action Lab (J-Pal), Innovations for Poverty Action (IPA), and the Program on Governance and Local Development (GLD) for their generous support of the scaled-up study.

On June 10 2014, the Islamic State of Iraq and Syria (ISIS) captured the Iraqi city of Mosul after only six days of fighting. ISIS' offensive in northern Iraq culminated in a genocide against Yazidis, Christians, Shi'a, and other minorities, displacing 100,000 Christians to the safety of Iraqi Kurdistan overnight ([Al-Ameen, Shaida 2014](#)). Many Christians believe their Sunni Arab neighbors were complicit in these raids. These suspicions have discouraged Christians from returning to liberated areas, fueled support for self-defense militias, and heightened the potential for reprisal killings and future conflict ([Center for the Prevention of Genocide 2016](#)). Shi'ites, meanwhile, have been moving into Christian enclaves, leading Iraq's Christians to fear the dilution of their culture and identity. Christian-Muslim relations in northern Iraq continue to be marked by mutual distrust and de facto segregation.

How can social cohesion be rebuilt in the wake of violent conflict? Countries recovering from war often backslide into violence and instability despite heavy international investment in state-building and peacekeeping programs. International intervention has largely failed in achieving sustainable peace ([Samuels 2005](#)). Policymakers have turned to promoting dialogue and building local institutions to improve intergroup relations. Truth and reconciliation forums have been found to encourage civic participation and intergroup cooperation, but at the cost of re-triggering war traumas ([Cilliers, Dube and Siddiqi 2016](#)). Community-driven development (CDD) programs can also increase intergroup cooperation in some settings ([Fearon, Humphreys and Weinstein 2009](#)) but not others ([Humphreys, de la Sierra and Van der Windt 2019](#); [Casey, Glennerster and Miguel 2012](#); [Humphreys, De la Sierra and Van der Windt 2013](#)), suggesting that effects are contextual ([Beath, Christia and Enikolopov 2012](#); [Fearon, Humphreys and Weinstein 2015](#)). Educational campaigns yield similarly mixed results when it comes to ethnic tolerance ([Finkel, Horowitz and Rojo-Mendoza 2012](#); [Blattman, Hartman and Blair 2014](#); [Staub et al. 2005](#)).

Another promising approach to prejudice reduction is intergroup contact. The 'contact hypothesis' proposes that interpersonal contact across group lines can reduce prejudice if it is positive, cooperative, endorsed by communal authorities, and places participants on equal footing ([Allport, Clark and Pettigrew 1954](#)). Such contact reduced prejudice toward roommates from ethnic outgroups in South Africa ([Burns, Corno and La Ferrara 2015](#)) and in the United States ([Carrell, Hoekstra and West 2015](#)), and toward poor students ([Rao et al. 2013](#)), neighbors ([Barnhardt 2009](#)), and other caste groups in India ([Lowe 2017](#)). On the other hand, wordless physical exposure was found to worsen existing prejudice toward outgroups in Afghanistan ([Condra and Linardi 2019](#)), refugees in Greece ([Hangartner et al. 2017](#)) and toward ethnic ([Enos 2014](#)) and socioeconomic ([Sands 2017](#)) minorities in the United States. In addition, intergroup competition increased violence ([Jha 2013](#)) and weakened otherwise

positive returns to social contact in India (Lowe 2017). These results suggest that positive and cooperative contact (conditions I label as *meaningful*) might hold the potential to rebuild tolerance — at least in times of peace.

It remains unclear, however, whether contact should have similarly positive effects in conflict settings. Fewer than 3% of the 515 contact studies reviewed in Pettigrew and Tropp’s 2006 meta-analysis involved groups in conflict. These same societies arguably have the most to gain from successful contact interventions, given the potential for conflict to spread to new regions via the experiences carried by displaced people (Ditlmann and Samii 2016; Salehyan and Gleditsch 2006). The evidence we do have suggests that ethnic prejudice is hard to dislodge relative to other types of prejudice (Paluck, Green and Green 2017). Methodological constraints also limit our knowledge of intergroup contact: outcomes are typically measured using short-term surveys or lab-in-the-field games that can struggle with external validity (Paluck, Green and Green 2017). It remains to be seen whether tolerance observed within an intervention can spill over into real-world behaviors, which are all the more critical in settings with an elevated risk of violence. If subjects return to their homogenous bubbles — exposed to the same social and structural barriers to tolerance as they were before the intervention — then “our current policies for encouraging contact may not be enough, even for those people the policies actually reach” (Enos 2014, p. 249).

I propose that meaningful contact based on a shared interest, where ethnic identity and politics take a backseat, can breed tolerance between hostile groups. Over time, contact with outgroup peers within such interventions can normalize contact outside of it, with outgroup strangers. This is especially true when contact is endorsed by communal authorities, and when subjects have a positive experience. I test these ideas using a field experiment among Iraqis displaced by ISIS. I randomly assign amateur soccer players to an all-Christian team, or to a team mixed with Muslims, for a two-month league. Despite the atrocities suffered during war, I show that contact can improve everyday behaviors — even if social and political beliefs remain stubborn to change.

I find that Christians with Muslim teammates are more likely to attend a mixed social event four months after the intervention ends (12 pp., $p < 0.19$), sign up for a mixed soccer team (14 pp., $p < 0.09$), and vote for a Muslim player (not on their team) to receive a sportsmanship award (25 pp., $p < 0.03$). The treatment is self-sustaining at the half-year mark: treated Christians are 54 percentage points more likely to train with Muslims six months later ($p < 0.01$). These gains did not come at the expense of backlash effects among all-Christian teams, as shown by match-level data on yellow and red cards. However, effects on attitudes were mixed. Treated Christians are more likely to believe in coexistence (0.46 SD, $p < 0.08$) but less likely to hold positive views about Muslims (0.45 SD,

$p < 0.01$). I hypothesize that standard measures of prejudice tend capture loaded policy positions in post-conflict settings, and may spark cognitive dissonance for those who have befriended members of the outgroup. I argue that mixed effects on attitudes are an acceptable cost for improving behaviors in post-conflict societies, where human safety is the first order concern.

This study makes three main contributions. First, this experiment is among the few causal tests of intergroup contact in a post-conflict setting. The results suggest that carefully designed intergroup contact can be a viable strategy for social reconstruction after ethnic conflict. Second, I expand the range of outcomes typically used in contact studies. Of the small subset of contact experiments that measure behaviors, most tend to capture tolerance toward individuals encountered in the experiment (e.g., [Lowe \(2017\)](#); [Burns, Corno and La Ferrara \(2015\)](#); [Rao et al. \(2013\)](#)). To test whether such behaviors generalize to outgroup strangers, I construct behavioral outcomes, like patronizing restaurants in noncoethnic neighborhoods, that reflect the costlier outcome of comfort around the outgroup writ large. By measuring behaviors outside of the intervention, as well as spillover effects among local residents, this study shows that intergroup contact has the potential to chip away at structural factors, like social segregation and communal norms, that keep communities divided.

Third, I expand our understanding of prejudice reduction by proposing that conflict may render behaviors easier to shift than attitudes. Social psychology theories suggest that behaviors precede and even produce attitudes, which develop more slowly ([Bem 1972](#); [Laird and Bresler 1992](#)). I argue that this attitude-behavior gap is especially pronounced in the shadow of war. Adjusting deeply rooted beliefs about an outgroup that one is conditioned to fear and distrust is more emotionally and cognitively burdensome than improving behaviors toward members of that group. Prejudice reduction interventions in Rwanda ([Paluck 2009](#)) and Nigeria ([Scacco and Warren 2018](#)) find boosts to behaviors but not attitudes. I discover a similar pattern here, challenging us to rethink core theories about how personal and group dynamics evolve. The results suggest that meaningful intergroup contact can build tolerant behaviors within and beyond the intervention — even if underlying beliefs remain sticky. Behavioral change, even if modest, may thus be a more feasible and worthwhile goal than uprooting latent prejudice.

The rest of the paper is structured as follows. In Section 1, I draw on the social contact literature to generate empirical hypotheses. Section 2 provides context on the ISIS genocide and Christian-Muslim relations in northern Iraq. Section 3 lays out the empirical strategy. Section 4 outlines the main results, and Section 5 speaks to causal

pathways that shed light on these results while ruling out backlash effects among the control group. Finally, in Sections 6 and 7, I interpret the “attitude-behavior” gap present in the results and speak to their generalizability.

1 Can Contact Build Tolerance After Conflict?

Allport’s intergroup contact hypothesis (1954) proposes that contact across group lines can reduce prejudice, build friendships, and improve intergroup relations overall. Because tolerant individuals are more likely to select into contact, scholars have turned to experiments to isolate the causal effects of contact. Positive and cooperative contact (what I label *meaningful contact*) successfully reduced socioeconomic, ethnic, and caste-based prejudice in the United States, South Africa, and India, respectively (Burns, Corno and La Ferrara 2015; Rao et al. 2013; Lowe 2017), in line with meta-analytic evidence demonstrating that contact generally reduces prejudice (Paluck, Green and Green 2017; Pettigrew and Tropp 2006).

Should we expect contact to be similarly effective in conflict settings? Observational studies of contact rarely involve groups in conflict (Pettigrew and Tropp 2006), while no experimental studies of ethnic prejudice among adults over the age of 25 were found in Paluck, Green and Green’s 2017 meta-analysis. On the one hand, interventions aimed at reducing ethnic or racial prejudice generate “substantially weaker effects” relative to prejudice toward the elderly or the disabled, for instance, suggesting that the cleavages common to war are particularly unpliable (Paluck, Green and Green 2017). More broadly, ethnic violence solidifies group identities, ethnic prejudices, and anxieties around being physically proximate to the outgroup (Scacco and Warren 2018; Fearon and Laitin 2000; Beber, Roessler and Scacco 2014). Groups in conflict are also more likely to be residentially segregated, making intergroup interactions less likely in the first place, and possibly worsening prejudice (Enos and Gidron 2016; Kunovich and Hodson 2002).

On the other hand, there are many reasons to believe that contact can reduce prejudice after conflict. Intense prejudice, together with few opportunities for antagonistic groups to meet, imply that such communities have a lot to gain when they *do* interact. This is because prejudice and social segregation feed into one another (Kasara 2014; Bhavnani et al. 2014; Condra and Linardi 2019). The more that groups stay apart, the more that encounters across group lines become stigmatized. Prejudice then lingers unchecked, making these encounters less likely and ultimately deepening both segregation and prejudice. I argue that meaningful interactions can break this cycle. A string of such interactions can normalize interacting with the outgroup, both as an idea and in practice. It becomes

easier to make the leap from feeling comfortable with the specific outgroup individuals one encounters, to feeling comfortable navigating interactions with outgroup strangers. One begins to feel more at ease dealing with the outgroup in different settings.

I hypothesize that interventions can break the cycle of social segregation and prejudice if four scope conditions are met. First, given the difficulty of bringing hostile groups together in the first place, the endeavor must be clearly in the interest of both groups. It should be the case that members of both groups would participate in the program, absent prejudice. Second, the program should be based on a shared skill or interest, rather than the express purpose of discussing the conflict. Only the most tolerant individuals typically self-select into intergroup dialogue initiatives ([Pettigrew 1998](#)), which also run the risk of re-triggering war traumas ([Cilliers, Dube and Siddiqi 2016](#)). The disadvantage of this approach, however, is that deep-seated grievances toward the outgroup may persist without being directly addressed. Third, it should be reasonably safe to interact with the outgroup. Individuals should fear neither retaliation from ingroup members nor the escalation of disagreements with outgroup members. Fourth, the outgroup must be geographically accessible. Even if neighborhoods are de facto segregated, the transaction costs of interacting should not be prohibitive.

These conditions tackle the first order problem of bringing segregated groups together, ideally in a setting relatively untainted by the weight of conflict. If these conditions are met, all but the most prejudiced individuals can theoretically be incentivized to engage in meaningful contact. Soccer leagues in the formerly ISIS-occupied Iraq satisfy these conditions: amateur athletes from both groups place a high value on participation, the intervention is based on a shared interest, intergroup relations are stable enough that participation is unlikely to trigger violence, and Muslims and Christians live in proximate neighborhoods. The fulfillment of these conditions may help explain the positive findings from the few experimental studies of contact that we do have from post-conflict settings: meaningful contact increased cooperation between classmates in Bosnia-Herzegovina ([Alexander and Christia 2011](#)), reduced discrimination in mixed classrooms in Nigeria ([Scacco and Warren 2018](#)), and improved perceptions of the outgroup among Jewish patients being treated by Arab doctors in Israel [Weiss \(2019\)](#).

Once segregated groups are brought together for a joint endeavor, the conditions considered key for activating the contact hypothesis are: 1) a positive experience, 2) a common goal, 3) cooperation to achieve that goal, 4) equal power status within the intervention, and 5) the endorsement of communal authorities, customs, or laws. I map these conditions onto the Iraqi context, and the design features of this experiment in particular, to further draw

testable predictions on the efficacy of contact in this case. While I cannot experimentally distinguish between these conditions, Section 5 demonstrates the particular importance of the endorsement of communal authorities in encouraging uptake, and of a positive experience in amplifying treatment effects.

First, participants will likely vary in the degree to which they have a positive experience. Participants are still healing from the trauma of displacement, and few competitive sports teams would enjoy absorbing strangers for a high-stakes tournament even in the best of circumstances. Both of these factors undermine the positivity of the experience, and negative contact is known to disproportionately affect tolerance relative to positive contact (Graf, Paolini and Rubin 2014; Paolini, Harwood and Rubin 2010). Nevertheless, team performance can create a positive experience. Because team success is unrelated to treatment status (demonstrated in Table 6), I expect that treatment effects are amplified for players on top-performing teams. A successful season proves that cooperating with the outgroup can be productive and enjoyable, while opening the door to new friendships. A poor season, however, may heighten intergroup tensions and prompt members of both groups to trade insults, argue with one another, and act aggressively.

Second, for mixed teams, this experiment satisfies the condition that groups should cooperate to achieve a common goal. All-Christian teams, however, encounter Muslims as opponents within the league. Previous work shows that groups will remain neutral (Lowe 2017) or gravitate toward hostility rather than friendship if contact is competitive (Amir 1976; Jha 2013). At the same time, another research agenda predicts the opposite effects of the competitive contact theory. Scholars have shown that secondary (Cameron et al. 2006), vicarious (Simonivitz, Kezdi and Kardos 2017), and even imagined (Crisp and Turner 2009) contact with outgroup members can reduce prejudice. If indirect contact theories hold, then all-Christian teams should become more tolerant over time by virtue of sharing a league with Muslims. By analyzing before-and-after changes among control teams, this study can distinguish between these competing hypotheses on the effects of indirect and competitive contact.

Third, the equal power status criterion seems moderately satisfied. Muslim and Christian teammates are equal in their status as players, though Muslims are always a numeric minority at either 25% or 28.6% of the squad. Team coaches, captains, and the bulk of spectators are also Christian, but another high-status figure, referees, are both Muslim and Christian.¹ Where equality matters most, however — on the field — all players are subject to the fundamentally equalizing effect of sports. Aside from the captainship, there is no official hierarchy between

¹ Christian residents also attended more games, at a median of 18 matches relative to Muslim residents' 15.

players, and all share the burden of contributing to the team effort. The idea of team sports as equalizers have rendered sports a popular backdrop for studies of prejudice (Kidd 2013; Dittmann and Samii 2016; Lowe 2017).

Turning to the final condition of supportive norms, local Christian partners worked to offset the stigma around intergroup contact at the onset of the study. At baseline, most Christian participants reported that their friends and colleagues were exclusively fellow Christians (53.7%), and just 6.6% characterized their social circle as ‘very’ diverse. Playing soccer with Muslims was also seen as particularly unacceptable. Only two of the study’s 51 teams were training with Muslims at baseline, some teams threatened to walk out if Muslims were included in the leagues, and coaches agreed that Muslims should remain a minority on each team should they be included (described in Section 5). Further, in the weeks before the study launch, the Syriac Catholic church cut rent subsidies equalling around \$400 per month to displaced Christian families (who have an average income of \$500 – \$1,000 per month), and the Kurdish government announced that displaced children in Ankawa would have to enroll in public schools located 80 miles away in their hometown of Qaraqosh.² The political climate surrounding the leagues thus made the displacement experience more salient, straining the patience for interacting with Muslims.

Despite these pressures, the leagues were endorsed by an NGO operated by the Syriac Catholic church. As the center of Christian social, spiritual, and even economic life, the church’s promotion of the mixed leagues helped to counteract the norm against interacting with Muslims, at least for the purpose of a soccer league.³ As I outline in the scope conditions above, an endeavor that is clearly beneficial to one’s ingroup, such as a sports association, trade union, or rotating credit association, can weaken taboos against contact. Church officials viewed soccer as a productive use of time for the community’s young men, 36% of whom (in the study, $n = 433$) were unemployed or would otherwise spend their time idling in *shisha* cafés. Overall, approval from communal authorities seemed to have shaped social norms such that participating in the soccer leagues was deemed minimally acceptable.

In sum, this experiment satisfies the conditions I outline for interventions build tolerance in post-conflict settings, and many of the traditional conditions considered important for contact to unlock tolerance in particular. Whether any effects can extend to behaviors outside the intervention, however, remains to be seen. This research design is

²The influx of IDPs in Ankawa increased rents by more than five-fold. Of the $n = 1,163$ displaced Christians Ankawa surveyed by the author between March and August 2017, 87% lived in private accommodations subject to these rent increases.

³The operational partner for the first wave of the study was the Humanitarian Nineveh Relief Organization, an NGO operated by the Syriac Catholic church. Although HNRO was unable to support the second wave of the study, the organization continued to be associated with the project by reputation.

well-suited to capture not only the impacts of contact on real-world behaviors after ethnic violence, but also the importance of cooperative vs. competitive contact, as well as a positive experience, in shaping results.

2 Context: ISIS Genocide, Displacement, and Muslim-Christian Relations

Iraq has been home to an indigenous Christian community for almost two millennia. Most Iraqi Christians consider themselves ethnic Assyrians, speak both Eastern Aramaic and Arabic, and subscribe to either to the Syriac Orthodox, Syriac Catholic, or Chaldean Catholic churches. Iraq's Christians are thus ethnically, religiously, and linguistically distinct from their compatriots. These distinctions, along with Assyrians' periodic agitations for autonomy, led to the targeting of Christians in northern Iraq first under the Ottomans (in conjunction with the Armenian genocide) in 1914, and later at the hands of the Iraqi army and their Kurdish allies, newly emboldened by the British withdrawal from Iraq, during the Simele massacre of 1933. Often caught between the Arabization policies of Iraq and the Kurdification policies of Iraqi Kurdistan, Ninewa's Christians have traditionally had a tense relationship with their Muslim neighbors relative to Christians in Baghdad and other southern cities.

Despite attempts to Arabize Christian culture, Saddam Hussein largely tolerated Christians and even allowed them to play a prominent role in public life. Under the Ba'ath Party, which itself was founded by a Christian, the Deputy Prime Minister, Minister for Transport, Minister for Science and Technology, and Minister for Industry all counted Christians among their ranks. A close connection with the Ba'ath regime, however, made Christians prime targets for retribution when Hussein was toppled in 2003. Christians subsequently poured out of Iraq. The war itself, in addition to the rise of Islamist fundamentalism, shrunk one of the oldest continuous Christian communities in the world from around 1.5 million in 2003 (7% of the population) to as low as 300,000 in 2013 ([Basu 2016](#)).

The fall of Mosul in June 2014 ushered in a new period of violence against Christians. With unfettered road access to hundreds of Yazidi, Christian, and Shi'a towns, ISIS enslaved and slaughtered thousands in what was formerly Iraq's most diverse province. Those they expelled — Muslim and Christian, Arab and Kurd — joined the three million Iraqis displaced by heavy fighting between the Iraqi Security Forces and armed groups from December 2013 to April 2017 ([UNHCR 2017](#)). The city of Qaraqosh typifies this exodus. An ancient hub of Assyrian and Babylonian civilization, Qaraqosh was emptied of its Christians for the first time in its history on August 6, 2014, when its residents were given hours to flee 50 miles away to Erbil, often on foot after reaching the Kurdish checkpoints. Most stayed in humanitarian camps, church accommodation, or private residences in the

Christian suburb of Ankawa.⁴ A Qaraqoshi mother-of-three interviewed by the author “brought enough clothes for my kids and myself for a couple of nights.”⁵ Qaraqoshis would remain in Erbil for two and a half years.

Internally displaced people (IDPs) began to trickle back to their hometowns after Ninewa was liberated in the Battle of Mosul in October 2016. The major Christian cities in the area saw a return of around half of their previous population as of September 2018. The other half remain scattered around Iraq, with some seeking asylum abroad in Jordan, Lebanon, North America, and Europe. Those that did return found their town all but destroyed. ISIS commanders and combatants had been living and looting in Qaraqosh with abandon. What ISIS did not loot, they torched on their way out of the city. Surviving homes, businesses, and churches were ransacked, peppered with bullet holes from firefights, and covered in hateful graffiti.⁶ Of the 476 Qaraqoshi Christians in the study sample, 46% had their homes looted, 36% had their homes destroyed, and 16% of property owners had deeds taken by ISIS (a major source of land disputes). A small minority of respondents report family members killed or missing (4%) and three individuals each report arrest, sexual abuse, or torture.

Christians feel a deep betrayal toward Sunnis (around 45% of Muslims in the study), whom they view as ISIS collaborators. That some Sunnis ‘chose’ to live under ISIS rule rather than flee fuels this perception. Interviews with Qaraqoshi Christians reveal a view that Sunnis who remained in the town could have done more to keep Christian valuables and homes safe during displacement. One of the first Christians to return Qaraqosh told Public Radio International: “When I bump into [my Muslim neighbors] now, they turn their faces and walk away... They know what they did. They know they’re guilty. I don’t even say hello to them” (Hall 2017). Sensing this hostility, only 23.2% of Muslims in the study claim to feel comfortable in Christian areas. This power dynamic also manifests in Muslims reporting more tolerant attitudes than Christians: 68.2% of Muslims would not mind a Christian as a neighbor and 89.7% would consider selling land to Christians, whereas only 23.7% of Christians would sell land to Muslims, and a mere 21.1% are ready to accept a Muslim neighbor.

In contrast to Sunnis, Shi’ites (55% of Muslims in the study) were displaced and victimized by ISIS much like Christians. Most belong to the 100,000-strong Shabak ethnic minority and speak Shabaki, Kurdish, and Arabic. A common displacement experience has done little to ease historic tensions between Christians and Shi’ites in

⁴Of the $n = 1,163$ displaced Christians in the Erbil area surveyed by the author between March and August 2017, around 13% lived in camps.

⁵Author interview, 33 year-old female homemaker, September 2018. Qaraqosh, Iraq.

⁶ISIS fighters spray-painted the doors of Christian homes with phrases such as “get out, Jews” and “the Islamic State will return.”

Qaraqosh. Encouraged by favorable land policies, residents of small Shabak villages have gravitated toward the urban center of Qaraqosh, tilting demographics away from a Christian supermajority.⁷ Aside from some intermingling in schools, Muslims are largely seen as outsiders to Qaraqosh. Qaraqoshis view Muslims as diluting the culture and identity of Iraq's last Christian strongholds. One interviewee, a 56 year-old schoolteacher, lamented this perceived encroachment: "They have the entire country. Why can't they let us have Qaraqosh?... I really would not be comfortable if one of them [a Muslim family] moved next door. I would feel uncomfortable on my own block. Their traditions and their habits are different."⁸ The activity of Shabak militias in the area has further stoked resentment.⁹ These recent frictions have made social cohesion all the more challenging after return.

3 Empirical Strategy

Experimental Design

Despite the differences between Christians and Muslims in the Ninewa plains, amateur soccer is popular among both groups. As described by [Putnam, Leonardi and Nanetti \(1994\)](#), civic associations like amateur sports clubs are engines for social capital; their 'cross-cutting' nature is crucial for building social trust and cooperativeness between citizens. Team sports also fulfill many of the conditions seen as key for contact to reduce prejudice, notably cooperation to achieve a common goal and an equal power status. It is thus unsurprising that actors from government, the private sector, and civil society, not to mention flagship initiatives by the United Nations, have focused on sports as a tool for community development since the 1920s ([Kidd 2013](#)). Researchers have, in parallel, documented a positive correlation between team sports and pro-social, civic, and cooperative outcomes ([Ditlmann and Samii 2016](#); [Lowe 2017](#)).

Intergroup sports also seem to fit the bill according to policymaker recommendations for integrating communities devastated by ISIS. A call to action by the International Organization for Migration (IOM) stresses the promise of "interventions that are specifically oriented around social cohesion... entail[ing] sustained, meaningful interpersonal contact," while the the International Republican Institute (IRI) proposes "positive, energetic, community

⁷ Saddam Hussein's Arabization policy in the 1980s encouraged Shabak to register as Arabs in exchange for land in the Qaraqosh area. Shabak have continued to purchase land and homes in Christian cities. Votes of recent Shabak arrivals also influenced the 2005 and 2009 elections, weakening the Christian bloc. Shabak comprised roughly one-tenth of Qaraqosh's population before the ISIS occupation and one-fifth after it.

⁸ Author interview, 56-year old female schoolteacher, September 2018. Qaraqosh, Iraq.

⁹ Stronger than their Christian counterparts, it is alleged that Shabak militia looted and damaged Qaraqosh in the immediate wake of ISIS' ouster, possibly to deter Christian return.

events... centered on nonpolitical issues to facilitate engagement between [Iraqi] communities in lower-pressure environments” (International Organization for Migration 2019; Zupruk, Whelan and Brouch 2018). Further affirming the promise of soccer as a backdrop for social cohesion, the same IRI report quotes a displaced Christian man in Ankawa as stating: “I am integrated in the community here. I have friends that I go out with to play football with... I feel more comfortable here than before.”

Leveraging the social potential of team sports and the universal popularity of soccer in Iraq, the experiment comprised four soccer leagues spread across two sites. Two leagues took place in Ankawa (a suburb of Erbil) and two in the city of Qaraqosh. Like much of life in Ankawa and Qaraqosh, amateur sports teams are largely segregated by religion. Research staff randomly recruited a total of 51 Christian teams out of a possible 60 teams across both sites. Captains were told that a local Christian-led NGO was working with a U.S.-based university to set up a ten-week soccer league for displaced people and returnees in the area.¹⁰ Participants were then told of two conditions for participating. First, each team would be allocated an additional three players who may or may not be Christian, bringing their team total from nine to twelve men.¹¹ Second, all players agree to complete a brief survey on the displacement experience and their views on Iraqi society before and after the league.

Treated teams thus received additional Muslim players while control teams received fellow Christians. I conduct a block randomization based on baseline responses to an item capturing empathy toward Muslims.¹² All contacted teams accepted despite initially protesting the inclusion of added players, especially non-Christians. This selection process resulted in an average participant who is 24 years old, unmarried, unemployed, with a high-school degree in hand and with a household income of around \$500 per month. He has few to no Muslim friends, believes Muslims are cursed, and would not consider selling land to Muslims, although he believes that Iraqis should treat each other as Iraqis first. Further descriptive statistics can be found in Table 10.

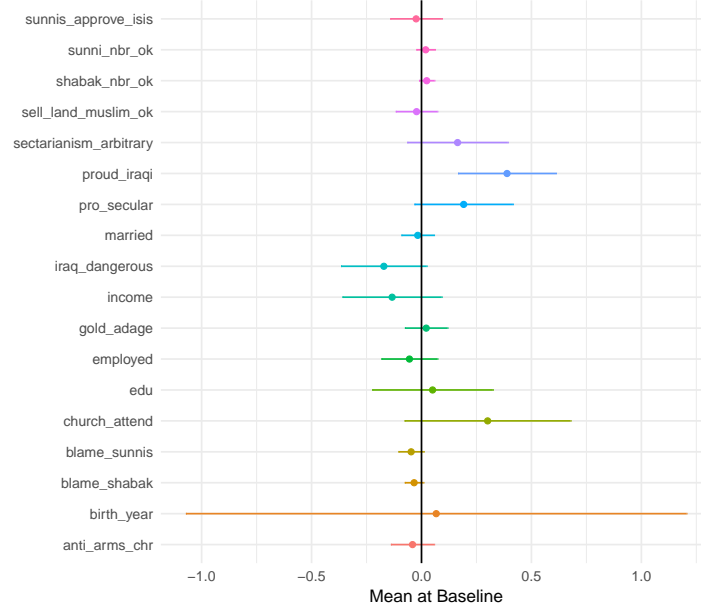
Of course, even players on all-Christian teams are exposed to Muslims by virtue of competing in the same league. A counterfactual of no exposure to Muslims at all (i.e., playing on an all-Christian team, in an all-Christian league) is needed to parse out the effect of competing in a mixed league. To address this, I create a ‘quasi-pure control’ group. This is the only league without Muslim players, allowing me to track outcomes among players that did not encounter Muslims at all. Assignment to the quasi-pure control league is non-random, however: teams were

¹⁰The Humanitarian Nineveh Relief Organization, an Iraqi NGO affiliated with the Syriac Catholic Church and serving IDPs, was the operational partner for the Ankawa-based pilot league.

¹¹In the pilot, the teams consisted of ten players, and received an additional four.

¹²The item asks respondents to rate how much they have in common with Sunni Arabs, on a four-point scale with no neutral option.

Figure 1: Balance Between Treatment and Control Group ($n = 433$)



Mean differences when subtracting the control group data at t_1 from the treatment group data at t_1 , with 90% confidence intervals.

Balance was achieved on the demographic covariates used in the analyses and on the 14 survey outcomes taken as inputs for the attitudinal indices.

recruited based on being located in Ankawa, about 50 miles from Qaraqosh. Driving time between both locations allows Ankawites to attend the occasional social event in Qaraqosh, but makes assignment to a Qaraqosh-based league — with its regular training sessions and games — too burdensome. Christians from Ankawa are slightly more urbanized (less religious, more likely to be employed) than their coethnics in Qaraqosh, but are otherwise indistinguishable when it comes to baseline attitudes toward Muslims (Figure 8). Given that Christians in Ankawa and Qaraqosh are part of the same ethnoreligious community, but bearing in mind that assignment to this league was non-random, I classify this group as a quasi-pure control. I exclude data from this league from the main analyses, but present results from the full sample in the appendix (see Table 11).

Incentives created with the team captains, such as professional referees and uniforms, reserved fields, and trophies awarded to the top three teams, led to near-perfect compliance and committed participation throughout the ten weeks (Figure 2).¹³ The vast majority (91.8%) of contacted participants were retained until the end of the study. The remaining participants dropped out before treatment assignments were made, or because of injury. These participants were replaced using the same procedure for recruiting added players, described below. Including

¹³Only 2% of matches played were forfeited due to absenteeism.

Table 1: League Descriptions

| League Type | Site | Duration | Christian Pop. | N Teams | Sample (Christian) |
|-------------------------|----------|------------------------|----------------------|---------|--------------------|
| 1. Experimental (Pilot) | Ankawa | Mar. 2017 – May 2017 | Displaced Qaraqoshis | 14 | 196 (168) |
| 2. Quasi-Pure Control* | Ankawa | Sept. 2018 – Nov. 2018 | Ankawites | 9 | 117 (117) |
| 3. Experimental | Qaraqosh | Sept. 2018 – Nov. 2018 | Qaraqoshi Returnees | 14 | 182 (161) |
| 4. Experimental | Qaraqosh | Sept. 2018 – Nov. 2018 | Qaraqoshi Returnees | 14 | 182 (161) |

*The quasi-pure control group (League 2) is excluded from the main analyses, but leveraged for exploratory analyses in Section 6.

coaches, the 51 teams yield a sample of $n = 677$. Just under half of these ($n = 280$) are treated, while the remainder are in the control condition ($n = 280$) or in the quasi-pure control ($n = 117$), used only for exploratory analyses in Section 6. To integrate the new players, encourage buy-in, and build team identity, each team received new uniforms and attended a one-day team orientation.¹⁴ Guidelines on substitutions in addition to a high-intensity format that requires regular rotations between players ensured roughly equal playing time between all team members.¹⁵ Each team played at least 13 games over eight weeks, with semi-finalists playing for an additional two weeks.¹⁶

Figure 2: Photo of a typical league game, March 27, 2017

Matches were attended by hundreds of spectators, further enhancing players' stake in cooperation and motivating participation.

The added players were drawn from rosters of amateur teams not chosen for the study. Recruiting added players from similar teams helps to ensure that added players do not systematically differ on skill or motivation. Research

¹⁴Orientations for treated and control teams were held separately to avoid contact spillovers between the two conditions.

¹⁵See Section 7.2 for more information on the league structure.

¹⁶Teams in the quasi-pure control league played at least eight games rather than 13, as there were nine teams in this league.

assistants also ranked each added player’s skill on a ten-point scale. The difference between the average rating of added Christian (6.03) and Muslim (6.45) players is neither substantively meaningful nor statistically significant ($p < 0.27$). Neither are their differences in goals scored per game (0.131 vs. 0.135, $p < 0.71$, see Table 6). Randomization therefore occurs on three levels: (1) which teams are selected to join the study, (2) treatment assignment, and (3) the added players assigned to each team, conditional on treatment condition.¹⁷

Estimation

For the main analysis, I estimate the average treatment effect (ATE) on a range of behavioral outcomes and attitudinal indices (Table 2). All estimates are ordinary least squares (OLS) where the treatment indicator represents assignment to a mixed soccer team, which occurs at the team level. I take several demographic items as covariates to increase precision, in addition the outcome variable measured at baseline (t_1) where possible. The demographic covariates are: age, education, marital status, church attendance, income, whether the respondent is an added player or an original team member, the randomization block, and abuse suffered at the hands of ISIS, asked last to avoid priming or ordering effects. Following Lin et al. (2013), I demean each covariate and interact it with the treatment indicator. This model has the benefit of increasing precision if covariates are predictive of outcomes, and yielding a consistent ATE estimator, without hurting asymptotic precision even if the model is incorrect.¹⁸

To address concerns that the effective sample size may be too small to trust asymptotic results, I conduct block-bootstrapped analyses to generate standard errors clustered at the team level. The number of clusters is either $n = 28$ or $n = 42$ depending on whether the outcome was measured during the pilot (Table 1).¹⁹ The exploratory analyses that include the quasi-pure control league have $n = 51$ clusters. As a robustness check, I also show that the results hold with a permutation test (Table 17). The items in these indices, along with t_1 covariates, and heterogeneous treatment effect analyses, were pre-registered.²⁰ Finally, missing data for covariates are imputed using multiple imputation by chained equations. Only 1.9% of rows contain a missing value requiring imputation.

¹⁷Players in the Qaraqosh-based leagues also experienced an fourth level of randomization: which league their team is assigned to. Two leagues took place concurrently at this site (see Section 7.2 in the Appendix for more information on the leagues’ set-up).

¹⁸Interacting each covariate with the treatment indicator following Lin et al. (2013) is a deviation from the pre-analysis plan, but one which improves precision while keeping point estimates roughly the same.

¹⁹Attending the social event, signing up for a mixed team next season, and training with Muslims are measured among the full sample of 42 teams while the remaining outcomes are measured only after the pilot league (Table 1).

²⁰EGAP registration #20170603AA (pilot) and AEA registration # AEARCTR-0003540 (scale-up).

Outcome Measures

I measure attitudes and everyday behaviors up to six months after the intervention ends. I co-design behavioral outcomes with local research staff, themselves Christians displaced by ISIS, and prototyped to ensure construct validity with participants via focus groups.²¹ As a result, these outcomes are tailored to capture tolerance in an unimposing and locally relevant manner. I also assess how exposure to the intervention correlates with attitudes among local residents using panel surveys collected at two points in time: one week before the final game, and three months later. By proxying for exposure to the soccer leagues through ties of family and friendship, distance to the field, and the number of games attended, this descriptive analysis illuminates possible spillover effects.

The first set of behavioral outcomes reflects tolerance on the field (i.e., within the intervention). The endline (t_2) survey asks players if they agree to register for a mixed team next season, or whether they prefer teammates from their own group. The research staff also contact players six months after the league's end to record whether they regularly train with Muslims.²² This measure tells us whether the intervention persisted beyond its official conclusion. If newly mixed teams continue to train together — when the formal incentives to do so are gone — then the intervention was habit-forming and, at least by the half-year mark, self-sustaining. Finally, participants vote for an added player to receive a sportsmanship prize. This player cannot be on the respondent's team. Voting for a fellow Christian to receive this prize, which is unrelated to skill, signifies ingroup bias relative to voting for a Muslim player.²³

The second set of behavioral outcomes reflects tolerance off the field (i.e., outside of the intervention). First, all players are invited to attend a neighborhood social event. For one of the leagues, this event was a dinner around sunset time on the 8th day of Ramadan, three weeks after the league.²⁴ It was common knowledge that this timing coincides with the fast-breaking meal (*iftar*) for Muslims. For most if not all, this dinner was the first instance of intermingling since displacement three years prior. In the other three leagues, the social event consisted of dinner, traditional dancing, and games four months after the league ended.²⁵ Players were encouraged to bring

²¹The same Iraqi research assistants record all outcomes to avoid pro-social behavior encouraged by the presence of foreign observers (Cilliers, Dube and Siddiqi 2015).

²²Players are asked who is on their current team roster rather than explicitly asking if they train with Muslims, to allay social desirability bias.

²³Following the measure of ingroup favoritism Lowe (Lowe 2017), the prompt stresses that the award is based on sportsman-like conduct rather than skill, and that the winner will receive a trophy at the end-of-league event, a coveted reward.

²⁴League 1 in Ankawa, described in Table 1.

²⁵This event was delayed due to flooding and scheduling conflicts with an annual tribal league, in which most Christian players were enrolled. However, this delay biases against finding treatment effects.

Table 2: Primary Behavioral and Attitudinal Outcomes

Behavioral Outcomes

1. Attend mixed dinner event (five months post-intervention; three weeks for pilot)
 2. Train with Muslims at least once a week (six months post-intervention)
 3. Patronize restaurant in Mosul (one to four months post-intervention)
 4. Donate \$ survey compensation to Church vs. neutral NGO (two weeks to five months post-intervention)
 5. Vote for Muslim player to receive sportsmanship prize (two weeks to five months post-intervention)
 6. Register for mixed team in the future (two weeks to five months post-intervention)
-

Attitudinal Indices

1. Coexistence
 - 1.1 Believe that Iraq would be a better society if Iraqis treated each other as Iraqis first
 - 1.2 Believe that dividing Iraq into ethnic and religious groups is arbitrary
 - 1.3 Believe that life these days is unpredictable and dangerous
 2. Intergroup Attitudes
 - 2.1 Sharing at least “a little” in common with Muslims
 - 2.2 Proud or very proud to be Iraqi
 - 2.3 Comfortable going to noncoethnic areas
 - 2.4 Some or most of friends are non-Christians
 - 2.5 Reject claim that most Sunni Arabs approve of ISIS
 3. Muslims as Neighbors
 - 3.1 Comfortable with Shi’ite Shabak as Neighbor
 - 3.2 Comfortable with Sunni Arab as Neighbor
 - 3.3 Comfortable with Sunni Shabak as Neighbor
 - 3.4 Willing to sell land to non-Christians
 4. Blaming Muslims
 - 4.1 Believe that Shi’ite Shabak are responsible for Christian suffering
 - 4.2 Believe that Sunni Arabs are responsible for Christian suffering
-

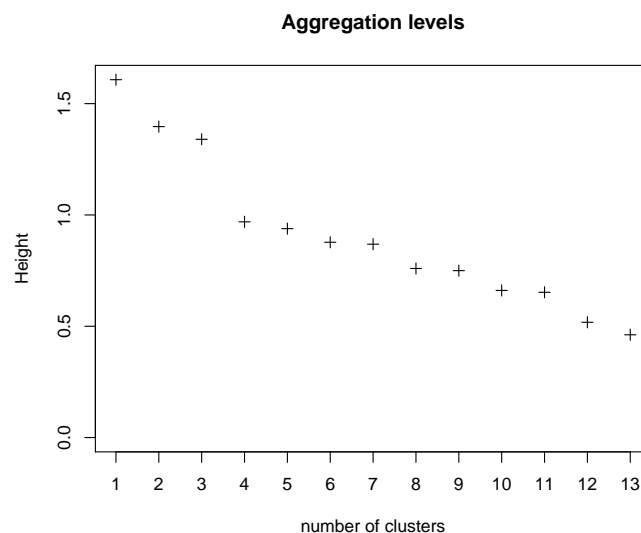
their families and friends, meaning that Christians were confronted with the possibility of socializing not only with Muslim players but with the latter’s family members and friends as well. The outcome of interest is whether a player attended this mixed social event, and conditional on attendance, whether he brought his female family members.²⁶ This measure is an especially high bar for social trust, as Christian players expressed concern that

²⁶This outcome was recorded by the research staff from direct observation at the event.

bringing their wives to the event would invite the unwelcome gaze of other men. Bringing one's wife or sister meant that one trusted that other men in the environment would behave respectfully.

Fourth, I instituted a voucher system to track whether treated players are more likely to patronize businesses in Muslim neighborhoods. This outcome speaks to comfort in diverse public spaces, a crucial step toward coexistence in segregated settings. All players received a voucher for two restaurants: (1) an \$8 voucher for restaurant in Muslim-majority Mosul around 35 minutes away by car, and (2) a \$5 voucher for a local, Christian-owned restaurant in Qaraqosh. The difference in voucher values compensated for the logistical and psychological costs associated with visiting these locales. Author interviews revealed that many Christian Qaraqoshis had not visited Mosul since the ISIS occupation, despite it being a proximate, major city where many had attended university or worked previously.²⁷ Interviewees perceived Mosul as a hotbed of Islamist activity and sympathy, in addition to voicing concerns over the city's security situation. Each voucher was stamped with the player's unique ID and was valid for up to four months after the intervention ended. The restaurant managers stored the vouchers such that I could track the amount spent and observe which vouchers were presented together. Visiting the Mosul restaurant reflects comfort with Muslim strangers, whereas visiting the local Qaraqosh restaurant should only be affected by the treatment if contact increased sociability in general.

Figure 3: Scree Plot of Survey Item Clusters



²⁷ Author interviews with four Christian residents aged 35 – 56, September 3 — 11, 2018, Qaraqosh, Iraq.

The final set of outcomes consists of attitudes toward intergroup relations measured at t_1 and t_2 . I combine similar survey items into an index to reduce measurement error. I do this by running an unsupervised hierarchical clustering algorithm on the t_1 data collected for a sample of Christian respondents ($n = 198$) with a dummy treatment vector. Using a data-driven method to identify latent clusters in the survey data removes subjectivity from the process of selecting items to form an index. The fourteen survey items of interest are then collapsed into four indices.²⁸ As a final step, I conduct a factor analysis on the four clusters to create scores that will serve as t_2 outcomes. The resulting indices — now dependent variables — cover the prospects for coexistence, intergroup attitudes, comfort with Muslims as neighbors, and blaming Muslims for Christian suffering (Table 2).²⁹

Two survey questions aim to measure ingroup bias in particular. The first asks respondents to choose a name from a list of fictional profiles (e.g., ‘Mohammed from Baghdad’) that reflects a person they would trust to receive a cash transfer on their behalf.³⁰ Only a fifth of the Iraqi population have bank accounts, with most relying on networks of cash transfer agencies to wire money domestically (Mousa 2013). These agencies are often locally owned and operated, and while accessible to anyone with a national ID card, are prone to coethnic favoritism. The outcome of interest here is whether one trusts a Muslim to manage a financial transaction. The second ingroup bias measure is the final question on the survey, informing respondents that \$1 will be donated on their behalf to a charity or NGO of their choosing. Donating to an explicitly Christian organization (most commonly, one’s local church) is a manifestation of high ingroup bias, while donating to an organization that benefits both Muslims and Christians (e.g., a cancer ward or orphanage) represents low ingroup bias.

Does exposure to the league prompt spillover effects among local residents? I proxy for league exposure in three ways: living within walking distance of the leagues’ soccer fields, having a friend or relative compete, and attending at least one match (all respondents are invited to the final game). The research team surveyed 152 randomly selected Christian and Muslim households in Qaraqosh one week before the leagues’ end, and again three months later. The outcomes of interest are the following three items: 1) *“To what extent do you agree with this statement: it is arbitrary to divide Iraqis into ethnic and sectarian identities?”*; 2) *“Would you prefer community activities like the soccer league to be mixed, or limit participation to those from your group?”*; and 3) *Do you think the league had a positive influence on your community?”* I expect that exposure to the leagues correlates with

²⁸The “elbow” in the scree plot produced by this method (Figure 4) plateaus sharply at four clusters, indicating that four primary indices should capture most of the variation in outcomes.

²⁹Items 1.1 and 1.2 are taken from the 2004 survey of Iraqis by Moaddel, Tessler and Inglehart (2008).

³⁰The other profiles are ‘Ali from Erbil’ (a Muslim), ‘Behnam from Erbil’ (a Christian) and ‘George from Beirut’ (a Christian).

tolerant responses on these items. While descriptive, this analysis provides exploratory evidence on the potential for contact to change social norms beyond direct participants.

4 Results

Table 3 summarizes the main results. Looking first at tolerance on the field, treated players are 14 percentage points more likely to report that they “would not mind” being assigned to a mixed team next season, 54 percentage points more likely to train with Muslims six months after the intervention ends, and 25 percentage points more likely to vote for a Muslim player (not on their team) to receive a sportsmanship prize. Moving to tolerance off the field, treated players are 12 percentage points more likely to attend a mixed social event and 8 percentage points more likely (although not statistically significantly so) to patronize a Muslim-owned restaurant in Mosul up to three months after the intervention ends ($p < 0.33$ when excluding the quasi-control league and $p < 0.09$ when including this league; see Table 11). Conditional on attending the social event, treated players brought their wives at almost identical rates as control players (27.3% vs. 27.4%, $p < 0.99$). Treatment effects are strongest among the most successful teams, operationalized by reaching the semi-final. Interaction effects here range in magnitude from 33% to 48% (Table 5). No similar subgroup effects were found when analyzing baseline contact or empathy (Table 7). The placebo outcome of visiting the local Christian-owned restaurant in Qaraqosh saw no treatment effects (20.9% in control vs. 24.2% in the treatment, $p < 0.39$). Applying the [Benjamini and Hochberg \(1995\)](#) multiple comparisons correction, all statistically significant outcomes remain so at the $\alpha = 0.10$ level.

Personal beliefs proved harder to improve. Treated players became less comfortable with Muslims as neighbors (0.45 SDs, $p < 0.01$), which would undermine the already-precarious Christian presence in northern Iraq. Views on other salient issues, like blaming Muslim civilians for Christian suffering, remained unchanged. The ingroup bias measures — donating one’s survey compensation to a cross-cutting organization, and trusting a Muslim with a financial transaction — were likewise stagnant. On the other hand, I do observe positive treatment effects for less salient beliefs. The coexistence index saw moderately positive treatment effects of 0.46 SDs ($p < 0.08$). This index combines items stating that ethnic and religious divisions are arbitrary, that Iraq would be a better society if citizens treated one another as Iraqis first, and that life in Iraq is generally stable and peaceful. Relative to the other indices, the coexistence index most captures abstract attitudes rather than concrete policy positions.

Table 3: Main Results

| | Control | Treated | p-value |
|----------------------------|---------|---------|---------------|
| Behavioral Outcomes | | | |
| Event Attendance | 26.0% | 38.4 % | 0.197 [0.220] |
| Mixed Team Sign-Up | 55.4% | 69.5% | 0.087 [0.094] |
| Train w/ Muslims | 7.72% | 61.7% | 0.000 [0.000] |
| Vote Muslim | 25.2% | 50.2% | 0.033 [0.045] |
| Visit Mosul | 23.0% | 31.4% | 0.347 [0.327] |
| Donate Neutral NGO | 63.5% | 87.6% | 0.597 [0.629] |
| Attitudinal Indices | | | |
| Coexistence | -0.10 | 0.13 | 0.081 [0.172] |
| Intergroup Attitudes | -0.12 | -0.18 | 0.951 [0.768] |
| Muslim Neighbor | -0.01 | -0.30 | 0.011 [0.052] |
| Muslim Blame | -0.24 | -0.12 | 0.734 [0.412] |

Estimates based on an OLS model with controls for randomization block, age, education, income, church attendance, ISIS abuse, respondent type (added player, core player, or coach), and the outcome measured at t_1 where available (attitudinal outcomes, donation outcome, and training outcome). All covariates are interacted with the treatment indicator. Standard errors are clustered at the team level. Block bootstrapped standard errors drawn from 1,000 samples are presented in brackets. Sample sizes are: $n = 318$ for going to Mosul (37 clusters), $n = 433$ for event attendance and training with Muslims (37 clusters), and $n = 265$ for the remaining outcomes (37 clusters). All variables are coded in a tolerant direction.

I also find descriptive evidence of spillover effects. Moving from watching one game to watching thirteen (equivalent to the first phase of the league; Table 1) is associated with a 66.7% increase in the likelihood of endorsing community programs that cross group lines at the three-month mark, where game attendance itself is strongly predicted by having a family or friend playing in a league (Table 15). The three channels through which residents are exposed to the leagues are associated with tolerant attitudes: living within walking distance of a league field, having a family or friend competing, and attending the final game upon the research staff's invitation are associated with a 19% to 32% boost to the belief that league had a positive effect on the local community (Table 4). Three months later, game attendance and attending the final game upon the league staff's invitation — which

included lengthy celebrations and a closing ceremony reiterating the league's 'fair play' messaging — persist in their correlation with positive attitudes.

5 Exploring Causal Pathways

The results show that intergroup contact improved tolerant behaviors on and (to a more modest extent) off the field, accompanied by mixed attitudinal shifts. What drives these effects? I explore three mechanisms that loom large in the contact literature: changing social norms, increased information about the outgroup, and increased empathy. To these mechanisms, I add a fourth: a positive experience, here defined as team success.

This section draws on exploratory analyses to show that contact on soccer teams normalized further contact in everyday life. A string of largely positive interactions with Muslim teammates seemed to make Christians more comfortable interacting with Muslims in less moderated environments, including the forging of new friendships. These pro-contact norms permeated the local community and correlate with league exposure. I also show that team success boosted treatment effects, affirming the importance of a positive experience in activating the contact hypothesis. In contrast, I find comparatively little evidence for contact working through information or empathy pathways.

Normalizing Contact: Social Norms and Spillover Effects

Playing with Muslim teammates increased the willingness to interact with Muslims within the realm of soccer, and eventually, outside of it. There appears to be a positive feedback loop between increased comfort interacting with the outgroup (i.e., lowered anxiety) and strengthened social norms that endorse these interactions. Perhaps the strongest piece of evidence in support of reduced anxiety and strengthened norms was the gradual acceptance of the idea that soccer teams can include non-Christians. The dominant norm in both Ankawa and Qaraqosh is that teams are segregated by religion.³¹ At baseline, only two of the 51 teams in the study (3.9%) included non-Christian players. Players were hesitant to absorb unfamiliar teammates, especially non-Christians. One coach walked out of an early information session and threatened to pull his team if Muslims were included. Another team, formed by members of an Assyrian nationalist movement, made similar threats. Some players protested

³¹For instance, the largest league in Qaraqosh is an all-Christian one based on tribal and kinship affiliations (thus excluding outsiders).

to the research staff that Muslims would “ruin the league” if they came to “our field” in the Christian enclave of Ankawa, even though the field in question is Muslim-owned and open to all communities.³²

Christian interviewees underscored that “rarely would you see Muslims playing with a Christian team,” and that “honestly, before, each community had their own team,” noting that “this is the first time that Christians are playing with our Muslim brothers” and that “this league was the thing that brought us together.”³³ Five months after the league ended, 16 teams (30.1%) were training regularly with Muslims, 14 of which were treated. A Christian player on one such team highlighted his Muslim teammates’ commitment to join the team for an upcoming tournament, “proving that there is some sort of bond between us.” Some participants proposed to league staff that they invite all-Muslim teams from the area to participate in the future, suggesting that anxieties toward the outgroup had been tempered. One treated team in particular went out of their way to accommodate the continued presence of Muslims on their squad. When Muslim players confessed that taxi rides from their neighborhood to the field were getting unmanageably expensive, their Christian teammates agreed to share the \$35 weekly cost despite a mean household income of \$500 to \$1,000 per month.³⁴

An increased openness toward Muslims in the leagues can also signify a reduction in ingroup bias. As shown in Table 3, players on mixed teams were double as likely to vote for a Muslim player (not on their team) to receive a sportsmanship prize (25.2% control participants vs. 50.2% in treatment) and around 14 percentage points more likely to indicate that they “would not mind” playing with noncoethnic teammates next season (55.4% vs. 69.5%). Relatedly, research staff noted that ‘blind’ attachments to fellow Christian players began to dampen. One mixed team elected a Muslim player as their captain, entrusting him to represent the team to (Christian) league staff and manage important team affairs, such as pleading the case of a teammate due to miss an upcoming game after accumulating two yellow cards. Moreover, some intra-team disputes were resolved in favor of Muslim team members, much to the surprise of local staff. These data points suggest that social distance may also have decreased as a result of the treatment.

Treated players seemed more comfortable around their Muslim teammates off the pitch as well. For instance, some Christian players invited their Muslim teammates to watch the UEFA Champion’s League final at a local

³²Field notes, Kellsey Beal, March 2017. Erbil, Iraq.

³³Interview conducted by Marie-Helene Carleton and Micah Garen, with a Christian league player in his early 30’s, April 2019. Qaraqosh, Iraq.

³⁴Field notes, Kellsey Beal, April 2017. Erbil, Iraq.

restaurant and shisha café during the league’s final days, with no prompting from the research staff.³⁵ Christian-owned cafés and social clubs in Ankawa enforce a strict sectarian door policy: national ID cards are checked to ensure that only Christians (or foreigners) are granted entry. These Christian players would almost certainly have had to negotiate with the café management and security staff to allow their Muslim guests in. Evolving attitudes toward the social event also illustrated a newfound comfort around Muslims. Some Christian players expressed concern over the prospect of Muslim players bringing unfamiliar family members to this event, voicing a hesitation to bring their wives lest they be leered at by the newcomers. At the time of the event six months later, noticing the absence of their Muslim teammates at the start of dinner, several Christian attendees phoned them to encourage them to attend. These anecdotes reflect small but important challenges to the segregated status quo in Ankawa and Qaraqosh. As one Qaraqoshi staff member remarked, “even if it’s a little bit, it changed something.”³⁶

Five months after the league ended, one popular Christian player remarked: “The important thing is the result, but there is another important thing, that all of us [from different communities] are together on one team. When the game is over we hug, kiss, congratulate each other even when we lose... that was really a huge thing. We see each other in the neighborhood, call each other, invite each other for a glass of tea or coffee at home or in a café ... we’re still in touch.”³⁷ The emergence of intergroup friendships — one of the contact hypothesis’s key predictions — was also noted by Muslim players: “in a soccer game, there isn’t this idea of which community you’re from... you’re playing, you’re competing, and friendships will grow.” Muslim players, for their part, demonstrated weakly positive shifts along most of the survey outcomes from t_1 to t_2 .³⁸ Confirming the emergence of cross-group friendships, treated players were about 8 percentage points more likely to report that their friends were “mainly mixed” or “mainly non-Christians” relative to the control group ($p < 0.18$, Figure 9).

Norm shifts are likely generated by a dynamic interaction between bottom-up and top-down forces (MacGinty 2010). Intergroup contact in this study was endorsed by influential figures, like coaches and local leaders, in ways that likely diffused norms among players. During the inaugural weeks of the league, Christians spoke Neo-Assyrian Aramaic (unintelligible to the Arabic-speaking Muslim players) and did not introduce themselves during

³⁵Field notes, Kellsey Beal and Rabie Zakaria, June 2017. Erbil, Iraq.

³⁶Interview, Rabie Zakaria, March 2019. Qaraqosh, Iraq.

³⁷Interview conducted by Marie-Helene Carleton and Micah Garen, with a Christian league player in his early 30’s, April 2019. Qaraqosh, Iraq.

³⁸Figure 7 looks at survey items making up the trust indices that were posed to both Muslims and Christians. Of the eight eligible items, six demonstrated shifts in a pro-tolerant direction when comparing t_1 and t_2 surveys in difference-in-means t-tests. A sample size of 55 renders these differences statistically insignificant although perhaps indicative, especially given that social desirability bias likely encouraged Muslims to over-report tolerant responses at t_1 .

orientation sessions.³⁹ The coaches made a point to speak in Arabic and their players followed suit, pointing to the influence of communal authority figures in moderating intergroup contact as predicted by contact scholars (Allport, Clark and Pettigrew 1954). A local Christian NGO, known to be operated by the Syriac Catholic Church, also lent operational and symbolic support to the first wave of the study, and advertised this support on their Facebook page and through the league's promotional materials. Local research staff described the NGO's stamp of approval as key in encouraging participation.

Tolerant norms appear to have permeated the town of Qaraqosh at large. The leagues gained substantial local traction, with hundreds of spectators lining the field to watch big games (Figure 2). Measures were taken to welcome local residents, including the provision of bleachers, concession stands, and activities for children. A Facebook group announcing schedules, reporting scores, and live streaming matches attracted over 2,600 members.⁴⁰ A survey of a random sample of 154 Qaraqoshis revealed that 81% had attended at least one game (with a median of 17 games attended) despite only 52.8% of respondents having relatives or friends competing. Open-ended responses reveal that residents viewed the leagues as an "example of coexistence" and "strengthening" of ties between different communities, lauding the fact that the "tournament is bringing us all together." Manipulation checks confirmed that virtually all residents were aware that Muslims were participating in the league.⁴¹

The half of Qaraqoshi residents (47.2%) without family or friends competing were still exposed to the leagues by virtue of three factors: the city's compact size (73.2% lived within walking distance of a league field), the placement of league fields next to major churches or town centers, and the tight-knit nature of community life in Qaraqosh. These factors likely accelerated the diffusion of pro-contact norms. One to four weeks after the final match, 96.2% of residents sampled "strongly agreed" that the league had a positive influence on their community, with 88.6% affirming the arbitrariness of ethnic and religious boundaries. These responses are mediated by the intensity of exposure (Table 4). Support for these propositions dropped, however, when the same respondents were surveyed three months after the league ended (78.6% and 80.9% respectively).

³⁹Field notes, Kellsey Beal, April 2017. Erbil, Iraq.

⁴⁰At the time of this writing in April 2019, the page has a 5/5 rating, over 9,000 photos (many uploaded by fans), and some videos viewed over 3,000 times.

⁴¹All 27 spectators surveyed during the pilot correctly identified that the league included Muslim players. Religion is easily identified in this context, both because of the close-knit nature of the Christian community (making outsiders quick to spot), and because player names were printed on the back of each uniform.

Table 4: Spillover Effects on Local Residents

| | <i>Attitudinal Outcomes</i> | | | | |
|-------------------------|-----------------------------|-----------------------|-----------------------|-----------------------|--------------------------------|
| | Pro-Leagues (t_1) | Pro-Secular (t_1) | Pro-Leagues (t_2) | Pro-Secular (t_2) | Pro-Mixed Activities (t_2) |
| | (1) | (2) | (3) | (4) | (5) |
| Walking Distance | 0.192* (0.112) | 0.067 (0.178) | -0.165 (0.215) | -0.037 (0.173) | -0.067 (0.167) |
| Family or Friend | 0.324*** (0.117) | 0.351* (0.186) | 0.322 (0.256) | 0.134 (0.207) | -0.250 (0.199) |
| Games Watched | 0.0002 (0.003) | -0.0004 (0.005) | -0.009 (0.007) | -0.004 (0.005) | 0.015*** (0.005) |
| Attended Final | | | 1.075*** (0.225) | 0.486*** (0.181) | -0.236 (0.175) |
| Constant | 3.318*** (0.131) | 3.182*** (0.209) | 2.514*** (0.476) | 2.919*** (0.383) | 1.078*** (0.370) |
| Observations | 121 | 121 | 102 | 102 | 102 |
| Adjusted R ² | 0.100 | 0.012 | 0.164 | 0.065 | 0.047 |

Note:

*p<0.1; **p<0.05; ***p<0.01

Models 1 and 3 analyze agreement that “the soccer leagues had a positive effect on our community” (4-point scale). Models 2 and 4 analyze agreement with the following statement: “Dividing the Iraqi people into Sunnis, Shi’is, and Kurds is artificial and contrary to the reality of the Iraqi society” (4-point scale). The t_1 survey was administered from the league’s final week (November 10, 2018) through to December 17, 2018, and the t_2 survey was administered from March 1, 2019 to March 9, 2019 (three months after the league concluded). Model 5 analyzes responses to the following statement: ‘Would you prefer community activities like the soccer league to be mixed, or limit participation to those from your group?’ (3-point scale). All models control for age and gender.

The predictive power of attending the final game, upon the league staff's invitation, is longer lasting: attendance correlates with the view that the league was a positive influence on the community, and the idea that sectarianism is arbitrary, at the three-month mark (Table 4). The final was marked by fanfare, messaging around sportsmanship and community spirit, and residents witnessing the positive impacts of diversity first-hand. Along the same lines, each additional game watched is associated with a 1.5% increase in support for future programs that cut across group lines, such as women's volleyball tournaments or gardening clubs (Table 4). All in all, exposure to the intervention therefore correlates with support for intergroup contact within soccer leagues and within other programs in the city, as well as a broader rejection of sectarianism. I cannot precisely measure social networks, but the high degree of interconnectedness between Qaraqoshis likely facilitated these spillover effects.⁴²

Team Success

A positive experience is thought to be crucial for contact to build tolerance (Graf, Paolini and Rubin 2014; Paolini, Harwood and Rubin 2010; Pettigrew and Tropp 2006). To rule out the concern that team success is post-treatment — in other words, that teams perform better as a result of being diverse — I show balance across treatment and control on three measures of team performance: the skill rating of added players, goals scored by added players, and total points accumulated by the team at the end of the league (Table 6).

Subgroup analyses run the risk of over-stating effect sizes due to the sample size (Gelman and Carlin 2014). With this constraint in mind, playing on a successful team — operationalized by advancing to the knock-out semifinal stage — improved attitudes rather than behaviors (Table 5). Interacting team success with treatment status, however, produces statistically significant interaction terms for three of the six behaviors (Table 5), including patronizing a restaurant in Muslim-dominated Mosul. These results imply that success can encourage positive attitudes, but the addition of contact is needed to boost tolerant behaviors. Anecdotally, participants seemed more willing to absorb players that had a concrete impact on the scoreline. When a Muslim striker scored several goals in the same match, one participant remarked that he would have preferred “a Muslim player who plays like *that*” instead of the less prominent Christian players that his team received. By the same token, Christian players seemed more quick to criticize the underperformance of Muslims as opposed to fellow Christians, including for minor transgressions like arriving a few minutes late to training. Successful teams also played together for an additional two weeks, potentially suggesting added returns to bundling dosage with a positive experience.

⁴²Qaraqoshis are so interconnected that most respondents had connections to both treated and control teams.

Table 5: Heterogenous Effects by Team Success

| | | <i>Outcome</i> | | | | | | | | |
|-------------------------|---------------------|--------------------|---------------------|---------------------|---------------------|---------------------|-------------------|---------------------|---------------------|-------------------|
| | Attend | Train | Mosul | Donate | Vote Muslim | Mixed Team | Coexistence | Intergroup | Muslim Neighbor | Muslim Blame |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Treated | -0.003 (0.091) | 0.326** (0.143) | -0.063 (0.087) | -0.059 (0.054) | 0.214** (0.097) | 0.025 (0.072) | 0.261* (0.152) | -0.021 (0.118) | -0.256** (0.112) | 0.016 (0.139) |
| Success | -0.042 (0.127) | 0.167 (0.261) | -0.016 (0.140) | -0.202 (0.166) | 0.295*** (0.103) | -0.043 (0.103) | -0.167 (0.105) | 0.862*** (0.241) | 1.216* (0.617) | -0.191 (0.121) |
| Treated:Success | 0.479*** (0.180) | 0.359 (0.289) | 0.331* (0.216) | 0.194 (0.172) | 0.042 (0.247) | 0.414*** (0.105) | -0.112 (0.191) | -0.213 (0.192) | -0.239 (0.368) | 0.157 (0.236) |
| Constant | 0.311* (0.164) | 0.200 (0.123) | 0.392*** (0.131) | 0.705*** (0.177) | 0.225 (0.194) | 0.356* (0.194) | -0.210 (0.214) | -0.121 (0.259) | -0.076 (0.447) | -0.115 (0.373) |
| Clustered S.E. | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Observations | 423 | 423 | 231 | 221 | 183 | 231 | 231 | 231 | 231 | 231 |
| Adjusted R ² | 0.127 | 0.339 | 0.105 | -0.008 | 0.070 | 0.041 | 0.062 | 0.124 | 0.181 | 0.172 |

Note:

*p<0.1; **p<0.05; ***p<0.01

Models are the same as those used for the main analyses (Table 3) with the addition of an interaction term to reflect team success, operationalized using a binary indicator for advancing to the semi-finals, and its lower order terms. Models 1 — 6 analyze behavioral outcomes, while Models 7 — 10 analyze attitudinal indices. The behavioral outcomes, from left to right, are: attending the social event, training with Muslims six months after the intervention, donating to a mixed NGO, voting for a Muslim player to receive a sportsmanship prize, and registering for a mixed team next season.

Table 6: Balance Table: Team Performance

| | Control | Treatment | p-value | <i>n</i> |
|---------------------|---------|-----------|---------|----------|
| Added Player Rating | 6.03 | 6.45 | 0.27 | 95 |
| Added Player Goals | 0.131 | 0.135 | 0.71 | 99 |
| Team Total Points | 16.79 | 20.50 | 0.32 | 28 |

Absorbing talented players in and of itself, however, does not seem to systematically shape treatment effects. When operationalizing team performance by a team receiving added players in the 25th skill percentile, neither this measure of performance nor its interaction with the treatment indicator yield statistically significant results (Table 14). How does team success relate to the talent of added players? Only one of the seven teams categorized as receiving the most talented players managed to reach the semifinal, underscoring that individual merit does not map perfectly onto team performance. The highest-scoring Muslim player in the league, for instance, happened to be assigned to bottom-three team. Having talented teammates is therefore not enough to build tolerance — the team must be successful too.

Other Mechanisms: Empathy and Information

In addition to changing norms, scholars posit that contact can build tolerance by increasing empathy and information about the outgroup. Contact is thought to induce empathy by highlighting similarities, thereby reducing perceived social distance between the ingroup and the outgroup. When analyzing the two survey items that directly measure empathy, however — having a “a little” or “a lot” in common with the two Muslim outgroups encountered in the study — I find null results when the Muslims in question are Sunni Arabs, and negative treatment effects for Shi’ite Shabak (Figure 9). Similarly, breaking down results by baseline empathy toward Muslims does not produce significant interaction effects, although I run into small sample constraints (Table 7). Previous work also stresses that contact can fill information gaps, often measured by a rejection of inaccurate stereotypes. I record two survey outcomes that capture stereotypes about Muslims: that they are cursed, and that most Sunnis supported ISIS. Neither outcome was affected by the treatment (Figure 9). This may be a disadvantage of apolitical contact: underlying grievances with the outgroup are not explicitly confronted.

Table 7: Heterogenous Effects by Baseline Contact and Empathy

| | <i>Dependent Variables</i> | | | | | |
|-------------------------|----------------------------|-----------------|------------------|-----------------|-----------------|-----------------|
| | Mix Team | Vote Muslim | Train Muslim | Mix Team | Vote Muslim | Train Muslim |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Constant | 0.20 (0.30) | -0.15 (0.50) | 0.52* (0.27) | -0.33 (0.40) | 0.66* (0.39) | -0.20 (0.44) |
| Contact : Treat | 0.01 (0.13) | 0.22 (0.17) | -0.03 (0.045) | - - | - - | - - |
| Empathy : Treat | - - | - - | - - | -0.03 (0.09) | 0.05 (0.09) | 0.01 (0.02) |
| Team Clustered S.E. | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Observations | 223 | 223 | 223 | 223 | 223 | 223 |
| R ² | 0.18 | 0.15 | 0.81 | 0.16 | 0.20 | 0.81 |
| Adjusted R ² | 0.07 | 0.04 | 0.78 | 0.04 | 0.06 | 0.78 |

Note:

*p<0.1; **p<0.05; ***p<0.01

Models are the same as those used for the main analyses (Table 3), with the addition of an interaction term either reflecting baseline contact with Muslims or baseline empathy, as well as the lower order term. Outcomes are: registering for a mixed team (models 1 and 4), voting for a Muslim player to receive a sportsmanship prize (Models 2 and 5), and training with Muslim players (Models 3 and 6). These outcomes had the most robust treatment effects in the main analyses.

Previous work shows that information must be new to affect outcomes (Dunning et al. 2017). In this case, contact provided neither new information on commonalities between groups, nor new facts about Muslims. Because Muslims and Christians have lived side by side in the Ninewa plains for centuries, the barrier to tolerance may not necessarily be a lack of knowledge about the outgroup. The most obvious commonality that this intervention would highlight, a shared love of soccer, is old news for Christian players — it is common knowledge that the region’s Muslims have their own teams and tournaments. Several coaches also proposed inviting Muslim teams to compete in future editions of the league, confirming an awareness of parallel Muslim teams. If anything, intergroup contact made Christians less likely to perceive Shi’ite Muslims as similar to them. Even if social

contact highlights differences, which post-conflict communities are keenly aware of, it can still improve tolerant behaviors. Fortunately, it seems that hostile groups do not need to view each other as similar in order to coexist.

Containing Backlash Effects

Table 8: Referee Identity and Infractions

| | <i>Dependent Variables:</i> | | |
|-------------------------|-----------------------------|-------------------|-----------------|
| | Total Cards (1) | Yellow (2) | Red (3) |
| Intercept | 4.46*** (1.05) | 4.01*** (0.83) | 0.45 (0.34) |
| Mixed | 0.47 (0.72) | 0.46 (0.57) | 0.01 (0.23) |
| Both Treated | 0.29 (0.86) | 0.17 (0.68) | 0.12 (0.28) |
| Muslim Ref | -3.74* (2.03) | -3.26** (1.60) | -0.48 (0.65) |
| Team Clustered S.E. | ✓ | ✓ | ✓ |
| Goal Controls | ✓ | ✓ | ✓ |
| Referee F.E. | ✓ | ✓ | ✓ |
| Observations | 98 | 98 | 98 |
| Adjusted R ² | 0.01 | 0.09 | -0.14 |

Note: *p<0.1; **p<0.05; ***p<0.01

Introducing intergroup contact in the aftermath of war can risk misunderstandings spiraling into violence, or competitive contact breeding resentment (Lowe 2017; Jha 2013). Paluck, Green and Green's 2017 review warns that naturalistic studies are likely to involve some amount of negative contact experiences, such as misunderstandings or outright conflict that could negatively affect outcomes. Proxying for aggression using yellow and red cards, I do not find evidence of increased hostility among those on all-Christian teams.⁴¹ Tables 10 and 17 demonstrate that

Table 9: Mean Infractions by Match-Up Treatment Status ($n = 99$)

| | Same Status | Different Status | p-value |
|--------------|-------------|------------------|---------|
| Total Cards | 3.00 | 2.92 | 0.89 |
| Yellow Cards | 2.68 | 2.56 | 0.80 |
| Red Cards | 0.32 | 0.36 | 0.82 |

the prevalence of cards does not differ across match types: matches that bring together all-Christian teams with mixed teams are not more hostile than matches between two treated or two control teams. Control participants also do not seem to become more prejudiced over time (Figure 6). Together, these results on backlash effects suggest that competitive contact does not worsen prejudice, but does not alleviate it either.⁴³

Experimental evidence reveals that both Muslim and Christian players contained potential conflict by censoring aggressive speech and actions toward outgroup members. The 6% of games (randomly) officiated by a Muslim referee saw an average of 1.80 cards given out, as compared with an average of 3.03 cards for games overseen by Christian officials ($p > 0.10$, Table 8). Anecdotally, the usually expressive Christian players did not dispute any of the Muslim referees' decisions, with one commenting that they wanted to hire a Muslim referee viewed as "very fair" for future leagues. Players regularly challenged opponents and referees from their own group but seemed to deliberately avoid confronting those from the outgroup.

This behavior accords with ingroup censoring: a cooperative, ingroup policing strategy that prevents disagreements from escalating (Fearon and Laitin 1996; Dittmann and Samii 2016; Cikara and Paluck 2013) and is associated with improved outgroup regard (Avenanti, Sirigu and Aglioti 2010). It is rational for the outnumbered Muslim players to avoid conflict. The restraint exercised by Christian players, on the other hand, points to an awareness that personal disputes could become dangerous given the tense political context. Along these lines, some intra-team disputes were also observed to be resolved in favor of Muslim teammates, perhaps rooted in the same desire to avoid escalation. It is important to keep in mind, however, that the leagues were a highly regulated environment overall. The presence of referees and local security forces at each game may have played an important role in containing backlash effects.

⁴³This analysis will be updated to include match-level data from over 500 games by October 2019.

6 Discussion: The Attitude-Behavior Gap

To sum up, contact normalized interactions with Muslims, reflected in heightened comfort around Muslim teammates, friends, and strangers, and in changing norms around the acceptability of contact among local residents. Team success was also decisive in converting contact into tolerance, with added returns to reaching the league's knock-out phase across all behavioral outcomes. A puzzle remains, however: why do we observe mixed changes to attitudes, and tolerant shifts across all behavioral outcomes except going to Mosul and donating to the Church?

Social psychologists have long noted that behaviors precede and often shape attitudes, which are generally slower to develop (Bem 1972; Laird and Bresler 1992). 'Self-perception theory' asserts that people develop attitudes by observing their own behavior and inferring what attitudes must have caused it. Scholars have found this to be true of prejudice (Ito et al. 2006; Guadagno et al. 2010) and pro-sociality (Wilson and Musick 1999). Regarding the null results on two of the four attitudinal indices, it may be the case that I observe intermediate results, and that attitudes are in the process of becoming more tolerant in the long run. Either way, the literature on self-perception suggests that finding behaviors to have changed before attitudes is unsurprising. The negative treatment effect on the intergroup attitudes index, by contrast, may be a product of cognitive dissonance (0.45 SD, p -value < 0.01, Table 3). Christians on mixed teams became the friendliest toward Muslims, but have been conditioned by conflict to resent, fear, and distrust Muslims as a group. Being asked to report positive views toward Muslims may have made subjects uncomfortable in ways that induced negative responses. An empathy-building intervention in Israel sparked a backlash in attitudes for the same reason (Gubler, Halperin and Hirschberger 2015).

The attitude-behavior gap seems particularly plausible in conflict settings. The psycho-social baggage of war makes processing attitudinal changes more cognitively taxing than usual, while changes to behavior become relatively easier. I argue that it is unreasonable to expect attitudes to change in high-stakes environments like societies recovering from conflict. Reducing prejudice is notoriously difficult when issue salience is high (Broockman, Kalla and Sekhon 2017), when the type of prejudice is ethnic or religious (Paluck, Green and Green 2017), and in the wake of violence (Fearon and Laitin 2000; Beber, Roessler and Scacco 2014) — all of which are true in post-ISIS Iraq. Moreover, our conceptions of 'old-fashioned prejudice' become complicated in war zones, where heuristics about the outgroup can be a matter of life or death, the traumas of violence have yet to heal, and complicated geopolitical alliances continue to drive wedges between neighbors.

Standard survey items on prejudice take on a sharpened meaning after war, becoming akin to policy positions. A commonly used item on comfort with different outgroups as neighbors, or another on willingness to sell land to outgroups, may be read as support for increasing the number of (often poorer) Muslim residents in Qaraqosh. In a country where Christians have all but disappeared, and in areas with relatively high housing prices, it would be difficult to label a Christian as prejudiced for responding negatively to these questions. A similar logic explains the lack of movement on the survey donation outcome. Christians rely extensively on the church for material and spiritual assistance: In a 2017 survey of Christian IDPs ($n = 1,163$), I find that nearly half received emergency aid from the church, and a third attend church once a week or more.⁴⁴ Two of the study's four leagues were adjacent to prominent churches and six teams (11.5%) named themselves after specific churches.⁴⁵ The centrality of the church to Christian identity, daily life, and overall wellbeing raises the possibility that the survey donation outcome is nearly orthogonal to intergroup tolerance.⁴⁶ It again seems that conflict hardens group identities, encourages the interpretation of attitudinal questions as high-stakes policy positions, or both.

Conversely, I find that the treatment did push one of the four indices in a tolerant direction — the coexistence index (Table 2). This index reflects an abstract optimism that warring groups can live together in peace, rather than requiring victims of war to renounce loaded social and political opinions. Breaking this index into its component parts, the treatment had no effect on the view that ethnic and religious boundaries are “artificial and contrary to the reality of Iraqi society” — because these group boundaries are powerful in Iraq — but did strengthen the belief that Iraq would be a better place if everyone treated each other “like Iraqis first.” Like [Paluck in Rwanda \(2009\)](#) and [Scacco and Warren in Nigeria \(2018\)](#), I find that prejudiced behaviors and norms may be more malleable in post-conflict societies than self-reported attitudes, and agree with [Greenwald and Pettigrew \(2014\)](#) that “the connection of prejudicial attitude to discriminatory behavior is not something to be assumed but, rather, something that requires empirical demonstration.”

A final question persists. Why did the treatment fail to alter the willingness to visit Mosul? Comparing results with and without the quasi-control group provides some answers. When pooling data from the quasi-control leagues, which had no Muslims at all, the ATE increases from 8 to 12 percentage points (a change from $p < 0.34$ to $p < 0.091$; Tables 3 and 11). Importantly, this is the outcome most reflective of comfort around Muslim

⁴⁴Almost a quarter (22.1%) claim that they became more religious after displacement.

⁴⁵Other popular team names often include references to local, regional identities, ethnic identities.

⁴⁶See [Makari \(2007\)](#) for a deeper discussion of the role of churches in organizing social, political, and economic life among Arab Christians.

strangers. One explanation is as follows. Encountering Muslim at a distance within a soccer league is analogous to encountering Muslims at a distance in another public space endorsed by the league's administration. For players on all-Christian teams, there may be little difference between Muslim strangers they come across in the league vs. Muslim strangers they come across in a Mosul restaurant. Players on quasi-control teams, however, have no exposure to Muslim strangers. As a result, none of these players felt comfortable enough to visit Mosul, while 23% of control players and 34% of treated players did (Table 12). When it comes to comfort with outgroup strangers, contact in a mixed ecosystem (e.g. a mixed soccer league) seems to be doing most of the legwork as opposed to additional contact on a mixed team. This result reflects a broader pattern of treatment effects concentrated among on-the-field outcomes. Given the volatile context, however, even modest changes to real-world behaviors are important.

7 Conclusion

Ongoing civil wars in the Middle East and Africa, worsening sectarianism across the Arab world, and a dearth of policies aimed at reintegrating communities hit by ISIS in particular have reinvigorated the question of how to build social cohesion in the wake of violence. Recent experimental evidence highlights the potential for social contact to improve relationships between groups. Despite its potential, however, we know little about whether contact can build lasting tolerance outside of the intervention setting, and even less about its impact in conflict zones. This study provides causal evidence on both of these questions by measuring the impact of intergroup contact on attitudes and real-world behaviors among Iraqis displaced by ISIS.

I find that being randomly assigned to compete on a soccer team with Muslim teammates increased tolerant behaviors on the field, and to a more modest extent, off of it. Treated Christians are more likely to attend a social event open to Muslims, vote for a Muslim to receive a sportsmanship prize, train with Muslim players, and sign up for a mixed soccer team next season, up to six months after the intervention ends. In addition, I find boosts to abstract beliefs in coexistence, although salient social attitudes remained unchanged or in some instances, worsened. These lasting behavioral effects are robust to several study waves, robustness checks, and a comparison with a quasi-pure control group, and do not come at the cost of exacerbating prejudice among the control group, as has been the case in other studies of Muslim-Christian contact ([Scacco and Warren 2018](#)).

Turning to the question of pathways, contact primarily built tolerance by normalizing contact with Muslims, be they teammates, friends, or strangers. Two-thirds of mixed teams had integrated Muslims as core team members six months after the intervention ended, indicating that contact can be habit-forming after formal incentives to interact fall away. As [Pettigrew \(1998\)](#) writes, “repetition makes intergroup encounters comfortable and ‘right.’” The endorsement of local leaders and coaches also played an important role in bolstering these new norms, which spilled over to the close-knit residents of Ankawa and Qaraqosh in the short term. A successful team performance was likewise decisive in unlocking tolerance, with the top-performing teams being more likely to patronize a restaurant in Muslim-dominated Mosul — an especially high bar for comfort around Muslims.

Following experimental findings from other post-conflict settings ([Paluck 2009](#); [Scacco and Warren 2018](#)), the results affirm that prejudice reduction interventions can improve the treatment of outgroups even if prejudice appears to persist. Social psychologists note that attitudes are often slower to change than behaviors. I hypothesize that this is particularly true in the shadow of conflict, where standard survey items taking on new meaning as loaded policy positions and may trigger cognitive dissonance among those who have befriended members of the outgroup. I argue that in contexts where the risk of violence is high, even modest improvements to behaviors are worth the risk of unchanged or somewhat worsened attitudes toward the outgroup, which in themselves are not nearly as dangerous as behaviors ([Fiske 1989](#)).

These findings hold several policy implications. First, interventions seeking to build social cohesion after conflict should consider aiming to change everyday behaviors rather than personal beliefs. Improving daily interactions in contexts where open hostility is the norm is a worthy goal, and seems achievable regardless of underlying prejudice. Second, outcomes measuring tolerance outside of an experiment are valuable yardsticks for change. Behavioral outcomes can be customized to track the lived experience of everyday coexistence in a given context. Such behaviors are ultimately the quantity of interest, especially when the threat of violence looms large. Third, engaging the local community as peripheral participants holds the potential to accelerate norm shifts, although more experimental work on spillover effects of contact is needed.

Amateur sports associations are remarkable for their potential to activate the contact hypothesis and build social capital, as well as their ubiquity across the globe.⁴⁷ Nonetheless, I propose that other interventions can facilitate

⁴⁷ According to the 2010 — 2014 wave of the World Values Survey, about 22% of the world’s population report membership in a sports association ([Inglehart and Puranen 2014](#)), not to mention those exposed to these associations second-hand.

meaningful contact after conflict if participation is mutually beneficial, based on a shared interest, unlikely to trigger violence, and accessible to both groups in spite of social segregation. I borrow two more conditions from contact scholars. I posit that endorsement from communal authorities is needed to encourage uptake in the first place, and that a positive experience is decisive in amplifying the effects of contact. Classrooms, dormitories, and civic associations are low-hanging fruit for policymakers interested in building social cohesion in unimposing ways, and have produced positive findings in other experimental settings ([Alexander and Christia 2011](#); [Burns, Corno and La Ferrara 2015](#); [Carrell, Hoekstra and West 2015](#); [Rao et al. 2013](#); [Lowe 2017](#)). Generalizability also depends on the particularities of the study sample. As a minority group in the study sites, the Muslim communities here are routinely exposed to Christians, making them amenable to a contact intervention. Yet these effects likely represent a lower bound on the impact of contact, because the leagues mainly attracted young Christian men harboring extreme views toward the outgroup even when compared to the general population.⁴⁸

The results support the idea that cross-cutting associations — those based on a joint skill or interest rather than an immutable identity — can build social capital and forge “weak ties” between groups ([Putnam, Leonardi and Nanetti 1994](#); [Granovetter 1973](#); [Banfield 1967](#)). By extension, these results undermine the argument that some societies are bound by path dependency to remain “uncivic” and resistant to policy interventions ([North 1990](#); [Zamagni 1978](#); [Putnam, Leonardi and Nanetti 1994](#); [Sabetti 1996](#); [Fukuyama 2001](#)). The promise of injecting such settings with meaningful contact can inform the \$612 million allotted by USAID for civil society, conflict mitigation, and reconciliation activities in 2016 ([USAID 2017](#)) and the millions of dollars spent globally on programs focused on exposure to outgroups ([McKone 2015](#)). Civil society actors in Iraq recommend building social cohesion by “helping to ensure sustained, meaningful inter-personal contact” specifically through “positive, energetic community events” that build “stronger inter-ethnoreligious relations based on common experiences between and within groups” ([International Organization for Migration 2019](#); [Zupruk, Whelan and Brouch 2018](#)). Iraqi NGOs have proven willing and able to collaborate with research and policy partners on issues of diversity, and were crucial in brokering peace accords that prevented post-ISIS revenge bloodshed and facilitated the return of 380,000 displaced Iraqis — saving the U.S. an estimated \$150 million per month ([Lindborg 2017](#)).

Ultimately, external validity is best determined by replicating the results from multiple internally valid studies in other contexts. Further research is needed to investigate the impact of intergroup contact in other conflict zones,

⁴⁸Figure 10 shows that the study participants are consistently more prejudiced toward Muslims than their counterparts from the general population ($n = 1,115$). 35.1% of participants agreed that “even if a Muslim is made of gold, [he] will burn a hole in your pocket.”

between different groups, and in different intervention settings. Much work remains to answer the call put forth by [Paluck, Green and Green](#) to causally isolate which conditions are necessary, sufficient, or optimal for intergroup contact, and to isolate mechanisms ([2017](#)). Future studies should also assess how effects may differ depending on whether existing units are diversified (e.g., embedding contact into pre-existing soccer teams) vs. constructed from scratch. Nevertheless, the results indicate that designing contact to optimize for meaningful interactions — as opposed to the wordless exposure that often marks daily life in segregated areas — can rebuild communities after war, and possibly sidestep the negative effects of ethnic diversity on social cohesion documented elsewhere ([Putnam 2007](#); [Alesina and La Ferrara 2000](#); [Sands 2017](#); [Enos 2014](#); [Hangartner et al. 2017](#); [Steinmayr 2016](#)).

References

- Al-Ameen, Shaida. 2014. "Christians displaced again by Iraqi violence." *Al-Monitor* .
URL: <https://www.al-monitor.com/pulse/originals/2014/07/iraq-christians-displaced-mosul-isis-crisis.html>
- Alesina, Alberto and Eliana La Ferrara. 2000. "Participation in heterogeneous communities." *The quarterly journal of economics* 115(3):847–904.
- Alexander, Marcus and Fotini Christia. 2011. "Context modularity of human altruism." *Science* 334(6061):1392–1394.
- Allport, Gordon Willard, Kenneth Clark and Thomas Pettigrew. 1954. *The nature of prejudice*. Addison-Wesley Reading, MA.
- Amir, Yehuda. 1976. The role of intergroup contact in change of prejudice and ethnic relations. In *Towards the elimination of racism*. Elsevier pp. 245–308.
- Avenanti, Alessio, Angela Sirigu and Salvatore M Aglioti. 2010. "Racial bias reduces empathic sensorimotor resonance with other-race pain." *Current Biology* 20(11):1018–1022.
- Banfield, Edward C. 1967. *The moral basis of a backward society*. Free Press.
- Barnhardt, Sharon. 2009. "Near and Dear? Evaluating the Impact of Neighbor Diversity on Inter-Religious Attitudes." *Unpublished working paper* .
- Basu, Moni. 2016. "In biblical lands of Iraq, Christianity in peril after ISIS." *CNN* .
- Beath, Andrew, Fotini Christia and Ruben Enikolopov. 2012. *Winning hearts and minds through development? Evidence from a field experiment in Afghanistan*. The World Bank.
- Beber, Bernd, Philip Roessler and Alexandra Scacco. 2014. "Intergroup violence and political attitudes: evidence from a dividing Sudan." *The Journal of Politics* 76(3):649–665.
- Bem, Daryl J. 1972. Self-perception theory. In *Advances in experimental social psychology*. Vol. 6 Elsevier pp. 1–62.
- Benjamini, Yoav and Yosef Hochberg. 1995. "Controlling the false discovery rate: a practical and powerful approach to multiple testing." *Journal of the Royal statistical society: series B (Methodological)* 57(1):289–300.
- Bhavnani, Ravi, Karsten Donnay, Dan Miodownik, Maayan Mor and Dirk Helbing. 2014. "Group segregation and urban violence." *American Journal of Political Science* 58(1):226–245.
- Blattman, Christopher, Alexandra C Hartman and Robert A Blair. 2014. "How to promote order and property rights under weak rule of law? An experiment in changing dispute resolution behavior through community education." *American Political Science Review* 108(1):100–120.

- Broockman, David E, Joshua L Kalla and Jasjeet S Sekhon. 2017. "The Design of Field Experiments With Survey Outcomes: A Framework for Selecting More Efficient, Robust, and Ethical Designs." *SSRN*.
- Burns, Justine, Lucia Corno and Eliana La Ferrara. 2015. Interaction, prejudice and performance. Evidence from South Africa. Technical report Working paper.
- Cameron, Lindsey, Adam Rutland, Rupert Brown and Rebecca Douch. 2006. "Changing children's intergroup attitudes toward refugees: Testing different models of extended contact." *Child development* 77(5):1208–1219.
- Carrell, Scott E, Mark Hoekstra and James E West. 2015. The impact of intergroup contact on racial attitudes and revealed preferences. Technical report National Bureau of Economic Research.
- Casey, Katherine, Rachel Glennerster and Edward Miguel. 2012. "Reshaping institutions: Evidence on aid impacts using a preanalysis plan." *The Quarterly Journal of Economics* 127(4):1755–1812.
- Center for the Prevention of Genocide, Simon Skjodt. 2016. "'Our Generation is Gone': The Islamic State's Targeting of Iraqi Minorities in Ninewa [Bearing Witness Report]."
- Cikara, Mina and Elizabeth Levy Paluck. 2013. "When going along gets you nowhere and the upside of conflict behaviors." *Social and Personality Psychology Compass* 7(8):559–571.
- Cilliers, Jacobus, Oeindrila Dube and Bilal Siddiqi. 2015. "The white-man effect: How foreigner presence affects behavior in experiments." *Journal of Economic Behavior & Organization* 118:397–414.
- Cilliers, Jacobus, Oeindrila Dube and Bilal Siddiqi. 2016. "Reconciling after civil conflict increases social capital but decreases individual well-being." *Science* 352(6287):787–794.
- Condra, Luke N and Sera Linardi. 2019. "Casual Contact and Ethnic Bias: Experimental Evidence from Afghanistan." *The Journal of Politics* 81(3):000–000.
- Crisp, Richard J and Rhiannon N Turner. 2009. "Can imagined interactions produce positive perceptions?: Reducing prejudice through simulated social contact." *American Psychologist* 64(4):231.
- Ditlmann, Ruth K and Cyrus Samii. 2016. "Can intergroup contact affect ingroup dynamics? Insights from a field study with Jewish and Arab-Palestinian youth in Israel." *Peace and Conflict: Journal of Peace Psychology* 22(4):380.
- Dunning, T, G Grossman, M Humphreys, S Hyde and C McIntosh. 2017. *Information and Accountability: A New Method for Cumulative Learning*. Cambridge University Press.
- Enos, Ryan D. 2014. "Causal effect of intergroup contact on exclusionary attitudes." *Proceedings of the National Academy of Sciences* 111(10):3699–3704.
- Enos, Ryan D and Noam Gidron. 2016. "Intergroup behavioral strategies as contextually determined: Experimental evidence from Israel." *The Journal of Politics* 78(3):851–867.

- Fearon, James D and David D Laitin. 1996. "Explaining interethnic cooperation." *American political science review* 90(4):715–735.
- Fearon, James D and David D Laitin. 2000. "Violence and the social construction of ethnic identity." *International organization* 54(4):845–877.
- Fearon, James D, Macartan Humphreys and Jeremy M Weinstein. 2009. "Can development aid contribute to social cohesion after civil war? Evidence from a field experiment in post-conflict Liberia." *American Economic Review* 99(2):287–91.
- Fearon, James D, Macartan Humphreys and Jeremy M Weinstein. 2015. "How does development assistance affect collective action capacity? Results from a field experiment in post-conflict Liberia." *American Political Science Review* 109(3):450–469.
- Finkel, Steven E, Jeremy Horowitz and Reynaldo T Rojo-Mendoza. 2012. "Civic education and democratic backsliding in the wake of Kenya's post-2007 election violence." *The Journal of politics* 74(1):52–65.
- Fiske, Susan T. 1989. "Examining the role of intent: Toward understanding its role in stereotyping and prejudice." *Unintended thought* 253:283.
- Fukuyama, Francis. 2001. "Social capital, civil society and development." *Third world quarterly* 22(1):7–20.
- Gelman, Andrew and John Carlin. 2014. "Beyond power calculations: Assessing type S (sign) and type M (magnitude) errors." *Perspectives on Psychological Science* 9(6):641–651.
- Graf, Sylvie, Stefania Paolini and Mark Rubin. 2014. "Negative intergroup contact is more influential, but positive intergroup contact is more common: Assessing contact prominence and contact prevalence in five Central European countries." *European Journal of Social Psychology* 44(6):536–547.
- Granovetter, Mark S. 1973. "The strength of weak ties." *American journal of sociology* 78(6):1360–1380.
- Greenwald, Anthony G and Thomas F Pettigrew. 2014. "With malice toward none and charity for some: Ingroup favoritism enables discrimination." *American Psychologist* 69(7):669.
- Guadagno, Rosanna E, Adam Lankford, Nicole L Muscanell, Bradley M Okdie and Debra M McCallum. 2010. "Social influence in the online recruitment of terrorists and terrorist sympathizers: Implications for social psychology research." *Revue internationale de psychologie sociale* 23(1):25–56.
- Gubler, Joshua Ronald, Eran Halperin and Gilad Hirschberger. 2015. "Humanizing the outgroup in contexts of protracted intergroup conflict." *Journal of Experimental Political Science* 2(1):36–46.
- Hall, Richard. 2017. "Iraqi Christians are slowly returning to their homes, wary of their neighbors." *Public Radio International* .
- Hangartner, Dominik, Elias Dinas, Moritz Marbach, Konstantinos Matakos and Dimitrios Xeferis. 2017. "Does Exposure to the Refugee Crisis Make Natives More Hostile?" *American Political Science Review* pp. 1–14.

- Humphreys, Macartan, Raul Sanchez De la Sierra and Peter Van der Windt. 2013. "Fishing, commitment, and communication: A proposal for comprehensive nonbinding research registration." *Political Analysis* 21(1):1–20.
- Humphreys, Macartan, Raul Sanchez de la Sierra and Peter Van der Windt. 2019. "Exporting democratic practices: Evidence from a village governance intervention in Eastern Congo." *Journal of Development Economics* .
- Inglehart, R., C. Haerpfer A. Moreno C. Welzel K. Kizilova J. Diez-Medrano M. Lagos P. Norris E. Ponarin and B. Puranen. 2014. "World Values Survey: Round Six - Country-Pooled Datafile." *Madrid: JD Systems Institute* .
- International Organization for Migration. 2019. Taking Stock of Refugee Resettlement: Policy objectives, practical tradeoffs, and the evidence base. Technical report International Organization for Migration.
- Ito, Tiffany A, Krystal W Chiao, Patricia G Devine, Tyler S Lorig and John T Cacioppo. 2006. "The influence of facial feedback on race bias." *Psychological Science* 17(3):256–261.
- Jha, Saumitra. 2013. "Trade, institutions, and ethnic tolerance: Evidence from South Asia." *American political Science review* 107(4):806–832.
- Kasara, Kimuli. 2014. "Does Local Ethnic Segregation Lead to Violence?: Evidence from Kenya." *Evidence from Kenya (February 25, 2014)* .
- Kidd, Bruce. 2013. A new social movement: Sport for development and peace. In *Sport and Foreign Policy in a Globalizing World*. Routledge pp. 36–46.
- Kunovich, Robert M and Randy Hodson. 2002. "Ethnic diversity, segregation, and inequality: A structural model of ethnic prejudice in Bosnia and Croatia." *The Sociological Quarterly* 43(2):185–212.
- Laird, James D and Charles Bresler. 1992. "The process of emotional experience: A self-perception theory." *Sage Publications, Inc* .
- Lin, Winston et al. 2013. "Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique." *The Annals of Applied Statistics* 7(1):295–318.
- Lindborg, Nancy. 2017. "Stabilize Iraq After ISIS, Try a Method That Works." *United States Institute of Peace* .
- Lowe, Matt. 2017. "Types of Contact: A Field Experiment on Collaborative and Adversarial Caste Integration." *MIT Working Paper* .
- MacGinty, Roger. 2010. "Hybrid peace: The interaction between top-down and bottom-up peace." *Security dialogue* 41(4):391–412.
- Makari, Peter E. 2007. *Conflict & cooperation: Christian-Muslim relations in contemporary Egypt*. Syracuse University Press.
- McKone, Kelly. 2015. *Reconciliation in practice*. United States Institute of Peace.

- Moaddel, Mansoor, Mark Tessler and Ronald Inglehart. 2008. "Foreign occupation and national pride: The case of Iraq." *Public Opinion Quarterly* 72(4):677–705.
- Mousa, Falah. 2013. "Overview of Banking System in Iraq." *Arab Business Review* .
 URL: <http://www.arabbusinessreview.com/en/article/overview-banking-system-iraq>
- North, Douglass C. 1990. *Institutions, institutional change and economic performance*. Cambridge University Press.
- 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, Seth Green and Donald P Green. 2017. "The Contact Hypothesis Revisited." *SSRN* .
- Paolini, Stefania, Jake Harwood and Mark Rubin. 2010. "Negative intergroup contact makes group memberships salient: Explaining why intergroup conflict endures." *Personality and Social Psychology Bulletin* 36(12):1723–1738.
- Pettigrew, Thomas F. 1998. "Intergroup contact theory." *Annual review of psychology* 49(1):65–85.
- Pettigrew, Thomas F and Linda R Tropp. 2006. "A meta-analytic test of intergroup contact theory." *Journal of personality and social psychology* 90(5):751.
- Putnam, Robert D. 2007. "E pluribus unum: Diversity and community in the twenty-first century the 2006 Johan Skytte Prize Lecture." *Scandinavian political studies* 30(2):137–174.
- Putnam, Robert D, Robert Leonardi and Raffaella Y Nanetti. 1994. *Making democracy work: Civic traditions in modern Italy*. Princeton university press.
- Rao, Gautam et al. 2013. "Familiarity does not breed contempt: Diversity, discrimination and generosity in Delhi schools." *Job Market Paper* .
- Sabetti, Filippo. 1996. "Path dependency and civic culture: some lessons from Italy about interpreting social experiments." *Politics & Society* 24(1):19–44.
- Salehyan, Idean and Kristian Skrede Gleditsch. 2006. "Refugees and the spread of civil war." *International Organization* 60(2):335–366.
- Samuels, Kirsti. 2005. "Post-conflict peace-building and constitution-making." *Chi. J. Int'l L.* 6:663.
- Sands, Melissa L. 2017. "Exposure to inequality affects support for redistribution." *Proceedings of the National Academy of Sciences* p. 201615010.
- Scacco, Alexandra and Shana S Warren. 2018. "Can social contact reduce prejudice and discrimination? Evidence from a field experiment in Nigeria." *American Political Science Review* pp. 1–24.
- Simonivitz, Gabor, Gabor Kezdi and Peter Kardos. 2017. "Seeing the World Through the Other's Eye: An Online Intervention Reducing Ethnic Prejudice." *American Political Science Review* pp. 1–8.

- Staub, Ervin, Laurie Anne Pearlman, Alexandra Gubin and Athanase Hagengimana. 2005. "Healing, reconciliation, forgiving and the prevention of violence after genocide or mass killing: An intervention and its experimental evaluation in Rwanda." *Journal of social and clinical psychology* 24(3):297–334.
- Steinmayr, Andreas. 2016. "Exposure to refugees and voting for the far-right:(unexpected) results from Austria.".
- UNHCR. 2017. "Iraq emergency." *UNHCR* .
- USAID. 2017. "Shared Progress, Shared Future: Agency Financial Report, Fiscal Year 2016." *United States Agency for International Development* .
- Weiss, Chagai M. 2019. "Can contact cure prejudice? A Natural Experiment in Israeli Medical Clinics." *University of Wisconsin Working Paper* .
- Wilson, John and Marc Musick. 1999. "The effects of volunteering on the volunteer." *Law and Contemp. Probs.* 62:141.
- Zamagni, Vera Negri. 1978. *Industrializzazione e squilibri regionali in Italia: bilancio dell'età giolittiana*. Mulino.
- Zupruk, Brian, Ashleigh Whelan and Emily Brouch. 2018. Social and Political Perspectives of Iraqi IDPs from Ninewa and their Host Communities. Technical report International Republican Institute.

Appendix

7.1 Descriptive Statistics

Table 10: Sample Descriptive Statistics and Balance ($n = 433$)

| | Control ($n = 259$) | Treated ($n = 174$) | p-value* |
|-------------------------------------|-----------------------|-----------------------|----------|
| Demographics | | | |
| Age | 24.06 | 23.99 | 0.91 |
| Married | 0.15 | 0.18 | 0.52 |
| Employed | 0.31 | 0.30 | 0.38 |
| High School | 0.80 | 0.79 | 0.80 |
| Income \leq \$500 | 0.53 | 0.48 | 0.35 |
| “Good” Health | 0.57 | 0.50 | 0.26 |
| ISIS Abuse | 0.32 | 0.33 | 0.88 |
| Church \geq 1x/week | 0.09 | 0.07 | 0.37 |
| Had Muslim Friends | 0.23 | 0.21 | 0.81 |
| Attitudes (t_1) | | | |
| Muslims Cursed | 0.61 | 0.67 | 0.43 |
| Anti-Sectarian | 0.72 | 0.69 | 0.49 |
| Sell Land to Muslims OK | 0.25 | 0.23 | 0.58 |
| Sunnis Approve of ISIS | 0.41 | 0.42 | 0.30 |

*P-value testing the null hypothesis of no difference between the groups. All variables are binary, except age. ISIS abuse defined as property theft, physical or sexual abuse, torture, arrest, or the kidnappings of family members. “Had Sunni Friends” refers to respondents who report having at least a few Sunni Arab friends before displacement. The “Muslims Cursed” variable refers to agreement with the prejudiced folk saying that “even if a Muslim is a piece of gold, [he] will burn a hole in your pocket.” Anti-sectarian refers to agreeing or strongly agreeing with the saying that “Iraq will be a better society if people treat one another as Iraqis, rather than Christians, Shi’is, Sunnis, or Kurds.”

7.2 League Set-Up

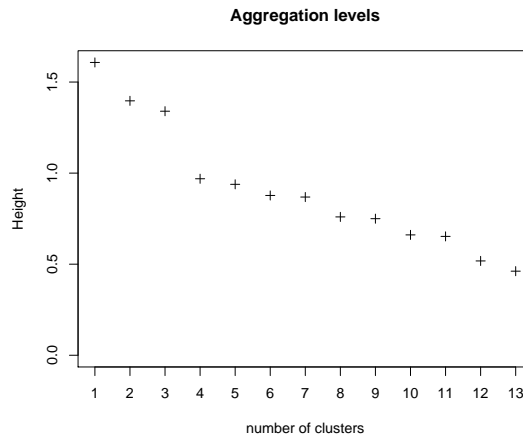
The teams play 6-a-side, 60-minute matches on a half-pitch, the common format among the amateur teams. The intensity of this set-up means that players must alternate regularly. Coaches are also instructed to ensure that players rotate regularly. Taken together, all team members should play for roughly the same number of minutes each game. An additional constraint is that each player must sit out at least two games over the course of the league to accommodate the 6-a-side or 7-a-side format, depending on the field capacity.

The games proceeded in two broad phases. Phase One is a classic round-robin set-up. Each team plays 13 matches over 8 weeks to reach a total of about 26 hours on the pitch per player when warm-up and cool-down time are taken into account. Phase Two entails knock-out semi-finals played by the four highest-scoring teams (three points for a win, one for a draw, and zero for a loss). The number of additional hours played by teams that qualify for the semi-finals ranges from two to four.

7.3 Generating Attitudinal Indices

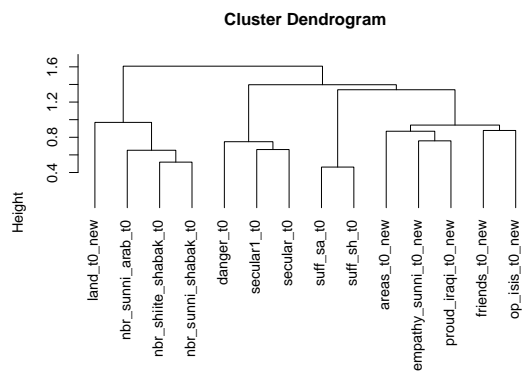
I use an unsupervised hierarchical model to detect latent clusters among the survey items. This data-driven method of identifying latent dimensions generates the scree plot below, indicating that there are three clear latent variables present in the data. The items in these indices align closely with theoretic expectations and can be labeled as trust toward Muslims, belief in coexistence, and the salience of Iraqi identity (Table 2).

Figure 4: Scree Plot of Survey Item Clusters



The scree plot generated by the clustering algorithm suggests there are four clearly delineated latent dimensions in the survey data, which I take as the key attitudinal indices. The individual items that make up these clusters can be seen in Figure 5 below:

Figure 5: Dendrogram (Tree) Diagram of Attitudinal Clusters



Each node in the above diagram represents the joining of two or more clusters; the locations of the nodes on the vertical (or horizontal) axis represent the distances at which the respective clusters are joined.

7.4 Over-Time Changes among Control Group and Muslim Participants

Figure 6: Shifts in Tolerant Attitudes among Control Group (t_1 to t_2) ($n = 189$)

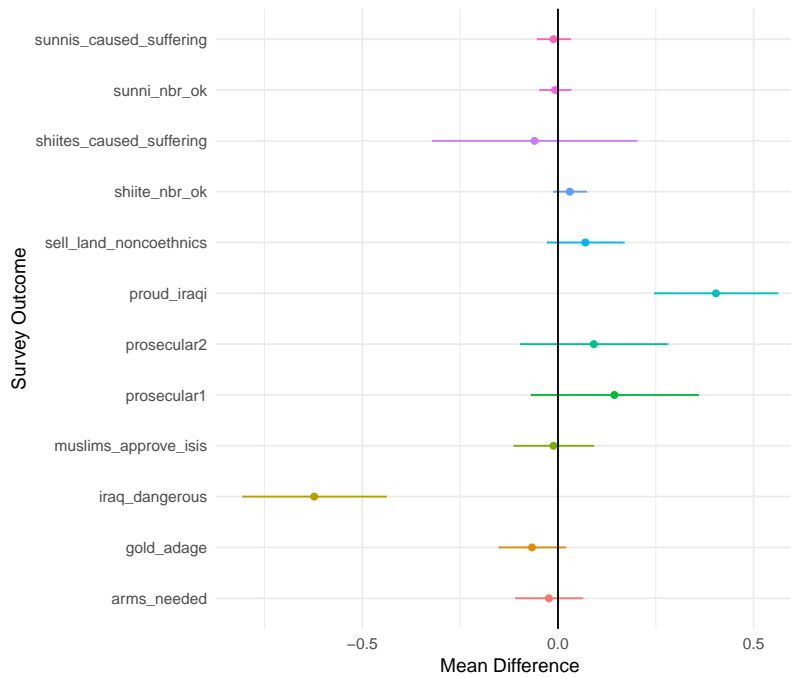
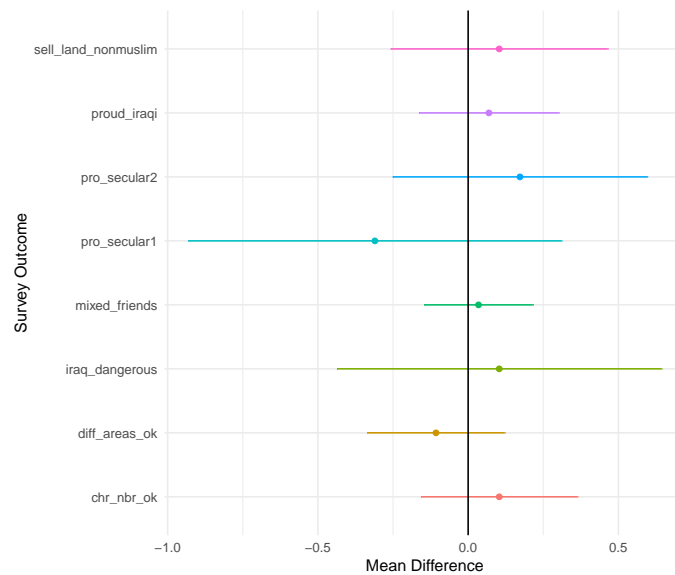


Figure 7: Shifts in Tolerant Attitudes among Among Muslims from t_1 to t_2 ($n = 55$)



7.5 Additional Analyses and Robustness Checks

Figure 8 shows that Christians in the Ankawa and Qaraqosh study sites are roughly comparable on a range of attitudinal and demographic variables, allowing for a meaningful interpretation of the quasi-pure control league.

Figure 8: Baseline Balance: Ankawa vs. Qaraqosh Leagues ($n = 265$)

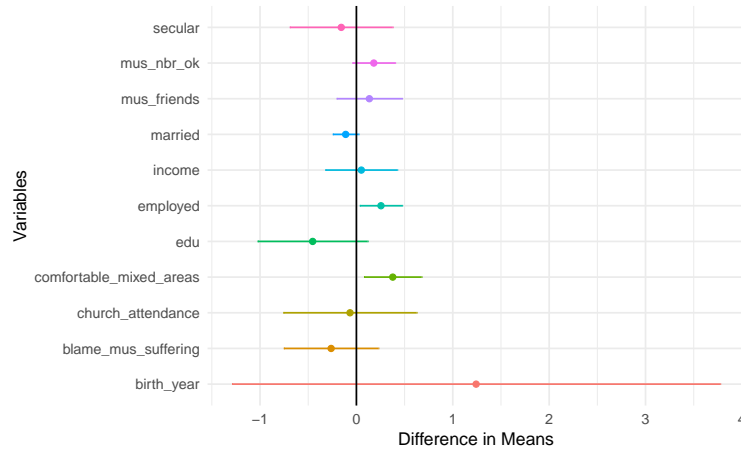


Figure 9 below breaks down the survey indices into their component parts, showing the ATE for individual survey items.

Figure 9: Treatment Effects: Individual Survey Items

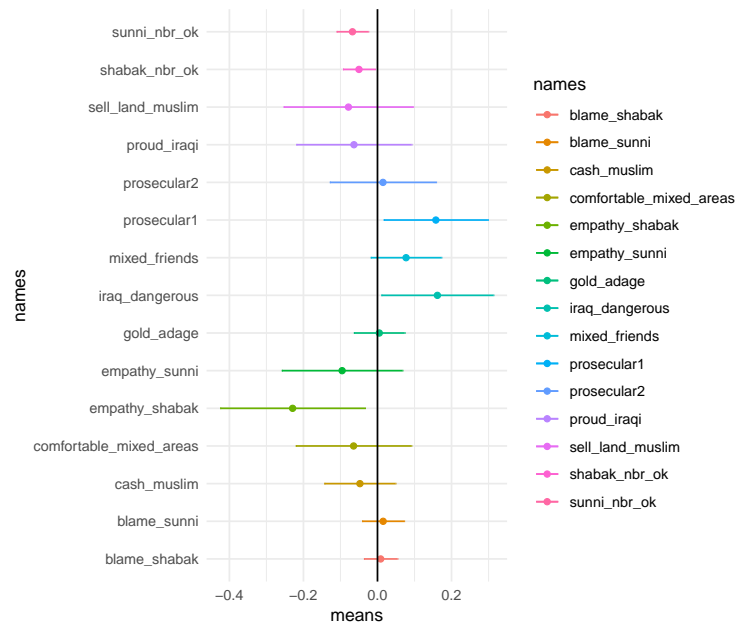


Table 11 below displays the main results for the full sample, pooled with the quasi-pure control league.

Table 11: Main Results (Full Sample)

| | Control | Treated | p-value |
|----------------------------|---------|---------|---------------|
| Behavioral Outcomes | | | |
| Event Attendance | 29.0% | 45.0 % | 0.078 [0.086] |
| Mixed Team Sign-Up | 49.9% | 57.0% | 0.070 [0.076] |
| Train w/ Muslims | 7.55% | 61.3% | 0.000 [0.002] |
| Vote Muslim | 42.5% | 63.5% | 0.011 [0.032] |
| Visit Mosul | 12.7% | 24.2% | 0.091 [0.120] |
| Donate Neutral NGO | 60.0% | 66.8% | 0.814 [0.346] |
| Attitudinal Indices | | | |
| Coexistence | -0.27 | -0.04 | 0.049 [0.153] |
| Intergroup Attitudes | 0.34 | 0.21 | 0.844 [0.367] |
| Muslim Neighbor | 0.45 | 0.15 | 0.021 [0.040] |
| Muslim Blame | 0.04 | 0.14 | 0.434 [0.795] |

Estimates based on an OLS model with controls for randomization block, age, education, income, church attendance, ISIS abuse, respondent type (added player, core player, or coach), and the outcome measured at t_1 where available (attitudinal outcomes, donation outcome, and training outcome). All covariates are interacted with the treatment indicator. Standard errors are clustered at the team level. Block bootstrapped standard errors drawn from 1,000 samples are presented in brackets. Sample sizes are: $n = 481$ for going to Mosul (37 clusters), $n = 433$ for event attendance and training with Muslims (51 clusters), and $n = 265$ for the remaining outcomes (37 clusters). All variables are coded in a tolerant direction.

I present the raw means for each of the key outcomes across both treatment conditions below. Raw means align closely although imperfectly with the covariate-adjusted estimates in Table 3, which is to be expected given the sample size. The three attitudinal indices are centered at a mean of zero with a standard deviation of 1. The following outcomes are calculated on a sample of $n = 242$, as they were asked only in Wave Two: donating to a neutral NGO, mixed team sign-up, voting for a Muslim player to receive a sportsmanship award, and the attitudinal indices. The going to Mosul outcome is calculated on a sample of $n = 318$, while event attendance has a sample of $n = 410$. To maintain consistency with the main results presented throughout the paper (Table 3), the quasi-control league is excluded from this table.

Table 12: Raw Means

| | t_1 Control | t_1 Treated | t_2 Control | t_2 Treated |
|-----------------------------|------------------|-----------------|-----------------|-----------------|
| Behavioral Outcomes | | | | |
| Event Attendance | — | — | 35.6% (2.91) | 44.2% (3.77) |
| Mixed Team Sign-Up | — | — | 49.3% (4.27) | 60.6% (4.82) |
| Train w/ Muslims | 10.87% (2.66) | 0.00% (0.00) | 18.2% (0.00) | 62.1% (3.69) |
| Vote Muslim | — | — | 30.7% (4.94) | 53.7% (5.14) |
| Go to Mosul | — | — | 23.0% (1.36) | 31.4% (3.23) |
| Donate Neutral NGO | 65.6 % (4.27) | 68.4% (4.05) | 76.0% (4.38) | 72.5% (4.54) |
| Attitudinal Outcomes | | | | |
| Belief in Coexistence | 0.02 (0.09) | −0.09 (0.09) | −0.01 (0.11) | 0.14 (0.10) |
| Intergroup Attitudes | 0.10 (0.08) | 0.04 (0.08) | −0.16 (0.08) | −0.16 (0.09) |
| Muslim Neighbor | 0.05 (0.09) | 0.13 (0.10) | −0.16 (0.05) | −0.18 (0.07) |
| Muslim Blame | 0.08 (0.07) | −0.04 (0.08) | −0.15 (0.11) | 0.08 (0.11) |

Table 13: Heterogenous Treatment Effects by Team Success: Attitudinal Outcomes

| | <i>Attitudinal Index</i> | | | |
|--------------------------------|--------------------------|----------------------|------------------|--------------|
| | Coexistence | Intergroup Attitudes | Muslim Neighbors | Muslim Blame |
| | (1) | (2) | (3) | (4) |
| Treated | 0.248* | −0.001 | −0.256** | 0.016 |
| | (0.142) | (0.117) | (0.129) | (0.139) |
| Success | −0.083 | 0.917*** | 1.258** | −0.218* |
| | (0.096) | (0.249) | (0.600) | (0.122) |
| Treated:Success | −0.015 | −0.353 | −0.249 | 0.089 |
| | (0.193) | (0.244) | (0.422) | (0.229) |
| Constant | −0.658 | −0.320 | −0.277 | 0.151 |
| | (0.527) | (0.490) | (0.582) | (0.795) |
| Observations | 231 | 231 | 231 | 231 |
| Adjusted R ² | 0.047 | 0.109 | 0.182 | 0.180 |
| Residual Std. Error (df = 204) | 1.012 | 0.954 | 0.934 | 0.905 |

Note:

*p<0.1; **p<0.05; ***p<0.01

Models are the same as those used for the main analyses (Table 3) with the addition of an interaction term to reflect team success (a binary indicator for advancing to the semi-finals) and its lower order terms.

Table 14: Heterogenous Effects by Added Player Rating

| | <i>Behavioral Outcomes</i> | | | | | |
|-------------------------|----------------------------|---------------------|---------------------|-------------------|-------------------|--------------------|
| | Attend | Train | Mosul | Donate | Vote | Mixed Team |
| Treated | -0.102 (0.068) | 0.666*** (0.121) | -0.069 (0.083) | 0.030 (0.062) | 0.205* (0.114) | 0.165** (0.077) |
| Success | -0.126 (0.087) | 0.214 (0.249) | 0.233 (0.153) | 0.322 (0.206) | 0.178 (0.125) | 0.198 (0.135) |
| Treated:Success | 0.097 (0.110) | -0.039 (0.277) | 0.162 (0.183) | -0.220 (0.194) | 0.122 (0.157) | -0.231 (0.145) |
| Constant | 0.443 (0.327) | 0.153 (0.189) | 0.664*** (0.201) | 0.680* (0.362) | 0.371 (0.286) | 0.072 (0.364) |
| Observations | 231 | 231 | 231 | 200 | 183 | 231 |
| R ² | 0.266 | 0.808 | 0.236 | 0.114 | 0.189 | 0.120 |
| Adjusted R ² | 0.177 | 0.784 | 0.139 | -0.020 | 0.065 | 0.013 |

Note:

*p<0.1; **p<0.05; ***p<0.01

Models are the same as those used for the main analyses (Table 3) with the addition of an interaction term to reflect team success (a binary indicator for receiving an added player in the 25th skill percentile, as rated by research assistants) and its lower order terms. Outcomes, from left to right, are: attending the social event, training with Muslims six months after the intervention, donating to a mixed NGO, voting for a Muslim player to receive a sportsmanship prize, and registering for a mixed team next season.

Table 15: Games Watched by Local Residents ($n = 117$)

| | Estimate | Std. Error | t-value | Pr(> t) |
|------------------|----------|------------|---------|----------|
| Walking Distance | -3.80 | 3.16 | -1.20 | 0.23 |
| Family or Friend | 18.38*** | 2.93 | 6.27 | 0.00 |
| Male | 5.02* | 2.96 | 1.69 | 0.09 |
| Age | 0.19 | 0.12 | 1.60 | 0.11 |
| Christian | 6.67 | 5.67 | 1.18 | 0.24 |

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Dependent variable is the number of games a respondent reports attending. Residual standard error = 15.18 on 117 degrees of freedom. Multiple R-squared = 0.2774. Adjusted R-squared = 0.2466. F-statistic = 8.985 on 5 and 117 DF, p-value: 0.00.

Table 16: Opposition Team Identity and Match-Level Aggression

| | <i>Dependent Variable</i> | | |
|-------------------------|---------------------------|------------------|-----------------|
| | Total Cards | Yellow | Red |
| | (1) | (2) | (3) |
| Intercept | 2.45 (1.52) | 2.51** (1.25) | −0.06 (0.46) |
| Mixed | 0.19 (0.56) | 0.15 (0.56) | 0.04 (0.23) |
| Both Treated | 0.07 (0.69) | −0.08 (0.69) | 0.15 (0.28) |
| Referee F.E. & Goals | ✓ | ✓ | ✓ |
| Observations | 98 | 98 | 98 |
| R ² | 0.13 | 0.15 | 0.06 |
| Adjusted R ² | 0.02 | 0.04 | −0.06 |

Note:

*p<0.1; **p<0.05; ***p<0.01

Table 17: Permutation Test

| | Treat Coef. | p-value |
|-----------------------------|-------------|---------|
| Behavioral Outcomes | | |
| Event Attendance | 0.26 | 0.009 |
| Train w/ Muslims | 0.68 | 0.000 |
| Vote Muslim | 0.32 | 0.072 |
| Go to Mosul | 0.05 | 0.000 |
| Donate Neutral NGO | −0.00 | 0.537 |
| Attitudinal Outcomes | | |
| Intergroup Attitudes | −0.10 | 0.000 |
| Coexistence | 0.22 | 0.030 |
| Muslim Neighbors | −0.22 | 0.154 |
| Muslim Blame | 0.06 | 0.010 |

Permutation test based on 5,000 resamples and clustered at the team level. Estimated using the same OLS model as in the main analysis (Table 3).

7.6 Generalizability

Figure 10 compares the League 1 sample with the general population from which it is drawn using a representative survey of displaced Christians in the Erbil area. Despite being around 11% more likely to believe that it is important to teach tolerance to children, Christian players have lower baseline levels of happiness, perceptions of safety, and trust toward Muslims, arguably making it harder to find positive treatment effects among this sample.

Figure 10: Comparison of Christian Participants ($n = 168$) and Displaced Christian Population ($n = 1,115$)

