Changing cultural attitudes towards female genital cutting

Vogt, Sonja ; Mohmmed Zaid, Nadia Ahmed ; Ahmed, Hilal El Fadil ; Fehr, Ernst ; Efferson, Charles

Abstract: As globalization brings people with incompatible attitudes into contact, cultural conflicts inevitably arise. Little is known about how to mitigate conflict and about how the conflicts that occur can shape the cultural evolution of the groups involved. Female genital cutting is a prominent example1, 2, 3. Governments and international agencies have promoted the abandonment of cutting for decades, but the practice remains widespread with associated health risks for millions of girls and women4, 5. In their efforts to end cutting, international agents have often adopted the view that cutting is locally pervasive and entrenched1. This implies the need to introduce values and expectations from outside the local culture. Members of the target society may view such interventions as unwelcome intrusions1, 2, 3, 6, 7, 8, 9, and campaigns promoting abandonment have sometimes led to backlash1, 7, 8, 10, 11 as they struggle to reconcile cultural tolerance with the conviction that cutting violates universal human rights1, 9. Cutting, however, is not necessarily locally pervasive and entrenched1, 3, 12. We designed experiments on cultural change that exploited the existence of conflicting attitudes within cutting societies. We produced four entertaining movies that served as experimental treatments in two experiments in Sudan, and we developed an implicit association test to unobtrusively measure attitudes about cutting. The movies depart from the view that cutting is locally pervasive by dramatizing members of an extended family as they confront each other with divergent views about whether the family should continue cutting. The movies significantly improved attitudes towards girls who remain uncut, with one in particular having a relatively persistent effect. These results show that using entertainment to dramatize locally discordant views can provide a basis for applied cultural evolution without accentuating intercultural divisions.

DOI: https://doi.org/10.1038/nature20100

LETTER

doi:10.1038/nature20100

Changing cultural attitudes towards female genital cutting

Sonja Vogt[1]*, Nadia Ahmed Mohmmed Zaid[2], Hilal El Fadil Ahmed[3], Ernst Fehr[1]§ & Charles Efferson[1]*§

As globalization brings people with incompatible attitudes into contact, cultural conflicts inevitably arise. Little is known about how to mitigate conflict and about how the conflicts that occur can shape the cultural evolution of the groups involved. Female genital cutting is a prominent example[1–3]. Governments and international agencies have promoted the abandonment of cutting for decades, but the practice remains widespread with associated health risks for millions of girls and women[4,5]. In their efforts to end cutting, international agents have often adopted the view that cutting is locally pervasive and entrenched[1]. This implies the need to introduce values and expectations from outside the local culture. Members of the target society may view such interventions as unwelcome intrusions[1–3,6–9], and campaigns promoting abandonment have sometimes led to backlash[1,7,8,10,11] as they struggle to reconcile cultural tolerance with the conviction that cutting violates universal human rights[1,9]. Cutting, however, is not necessarily locally pervasive and entrenched[1,3,12]. We designed experiments on cultural change that exploited the existence of conflicting attitudes within cutting societies. We produced four entertaining movies that served as experimental treatments in two experiments in Sudan, and we developed an implicit association test to unobtrusively measure attitudes about cutting. The movies depart from the view that cutting is locally pervasive by dramatizing members of an extended family as they confront each other with divergent views about whether the family should continue cutting. The movies significantly improved attitudes towards girls who remain uncut, with one in particular having a relatively persistent effect. These results show that using entertainment to dramatize locally discordant views can provide a basis for applied cultural evolution without accentuating intercultural divisions.

Ethnographic work emphasizes that cutting is subject to constant negotiation and renegotiation within the culture in question[2,3,7,8]. If true, ideas of abandonment should already be present within a cutting community. Recent analyses of large data sets have confirmed this view by showing that attitudes associated with cutting vary greatly at the household and individual levels[12–14]. Importantly, individual variation in attitudes within households[3,13] is effectively as local as possible, and programmes that take this kind of local heterogeneity as a starting point for abandonment could provide a basis for change[3]. Put differently, the international movement to reduce violence against women and girls could be effective to the extent that it is translated into local terms[2]. Locally heterogeneous views about cutting offer an opportunity to do this by casting the debate about cutting versus abandonment into a local vernacular[2,3].

To see whether this would work, we produced four telenovela-style movies that portray locally disparate views on cutting. As explained below, the movies vary systematically in terms of content. Some movies examine the link between cutting and individual values, whereas others examine the effects of cutting on the marriage prospects of young women[15]. The movies were shown to people in two fully randomized and controlled experiments. The variation in content between movies allowed us to determine whether certain arguments related to cutting were more compelling than others by systematically decomposing and manipulating the content of interest across several treatments.

In addition, we sought to test whether entertainment focused on local heterogeneity could change attitudes about cutting. The exploration of gender bias through entertainment dates back at least to a Peruvian telenovela in 1969 (ref. 16) and has included a small number of documentaries and radio dramas that have addressed cutting[17,18]. Only recently, however, have social scientists developed methods to identify causal effects, as opposed to mere correlations, associated with entertainment. Existing causal studies based on entertainment have not addressed female genital cutting[19–22], and more broadly our study is one of very few to include causal inferences about programmes of any kind related to cutting[23].

In accord with the focus on local heterogeneity, the movies in our experiments take a non-judgemental approach to cutting. They do not consistently present arguments in favour of abandonment, and they do not associate negative characters with the support of cutting. Rather, the movies dramatize how difficult a decision about cutting can be for parents who want the best for their daughters in a society where cutting is common, but attitudes and practices vary. The focus on local heterogeneity also works around a challenge that typically arises when using entertainment to change attitudes. Attitudinal change through entertainment involves a basic trade-off. It must present people with characters and situations that are identifiable but still provide examples of new ideas[16]. As previous research has not revealed much about how to negotiate this trade-off[16], our movies work around this by allowing characters within an extended family to disagree about cutting. Importantly, the members of the family do not choose cutting or abandonment independently of each other. They do, however, debate the issue. This reflects the individual variation in attitudes known to exist[13,14] and situates the cultural tension between cutting and abandonment as locally as possible[2,3].

We produced four movies that served as experimental treatments (Supplementary Information Section 2). The movies differ in terms of the arguments the fictional characters articulate when expressing their views on cutting (see Methods). A number of motives for cutting have been proposed, and they broadly fall into two categories[15]. First, people may have opinions for or against cutting because people hold individual values related to health, purity and perceived religious obligations. For opinions of this sort, cutting has an intrinsic value, positive or negative, rooted in the individual's view of the practice. Second, people may have opinions for or against cutting because they hold certain beliefs about how cutting will affect the future marriage prospects of their daughters. In this sense, cutting has an extrinsic value widely assumed to depend on whether families expect cut wives for their sons. We varied whether the characters in the movies express opinions in these two categories.

[1]Department of Economics, University of Zurich, 8006 Zurich, Switzerland. [2]Omdurman, Sudan. [3]Khartoum, Sudan.
*These authors contributed equally to this work.
§These authors jointly supervised this work.

We did so in all combinations, which gave us a 'values' movie, a 'marriageability' movie, a 'combined' movie that addresses both values and marriageability, and a 'control' movie.

All movies are about the same extended family in a rural area in Sudan. The extended family includes multiple couples and their children, a venerable grandfather and other relatives and spouses. Family members live together in a large compound. The movies share a main plot that is not about cutting. The main plot involves a mix of love, intrigue, deception, forgiveness and engaging personalities. The main plot is always the same and consists of nearly identical scenes across all of the four movies. The control includes only the main plot, whereas the three remaining movies include subplots related to female genital cutting. For the subplots, several couples in the extended family have daughters approaching the age of cutting, and family members debate whether they should continue or abandon the practice. The three subplots consist of parallel scenes and dialogues in which characters have discussions or disputes about cutting. Regardless of treatment, the same characters are always included in a scene, and the scene is always shown in the same position in the narrative structure of the movies. Only the dialogues vary depending on the treatment. All movies are approximately 90 min long, and subplots about cutting constitute approximately 30% of each of the three movies about cutting (approximately 27 min).

The movies do not simply display conflict and leave disputes unresolved. All three movies about cutting provide examples of how to negotiate conflicting views within a family. The movies reveal that husbands and wives discuss cutting together. This is an important innovation in Sudan, where people can generally have difficulty discussing cutting openly when men and women are both present. Moreover, the three movies about cutting end with the family presenting the idea of abandoning the practice to the ageing grandfather, the senior member of the extended family. This scene is the culmination of one key source of conflict in the movies. Specifically, conflicts occur repeatedly because some characters feel that family members must approach the grandfather to discuss cutting, whereas others feel that doing so would cause him unnecessary stress. When the family finally approaches the grandfather, he approves of the proposal to abandon cutting. He speculates that his dead wife would also approve, and he draws an analogy between cutting and facial scarring, a practice once common in Sudan but now in rapid decline. The grandfather's reaction confirms the family's decision to stop cutting, and in this way the movies provide a model for how to include young and old generations in debates about cutting. In keeping with our general strategy, the final scenes with the grandfather are parallel across the three movies about cutting; only the details of the discussions vary.

We used the four movies as treatments in two experiments. For the first experiment (Supplementary Information Section 3), we randomly assigned individuals within communities to the four treatments and collected attitudinal data using an implicit association test (Supplementary Information Section 1) immediately after participants watched the movies. By minimizing socially desirable responses, implicit association tests have been used to successfully measure attitudes about sensitive topics in which respondents may not be prepared to explicitly reveal their attitudes[24]. For the second experiment (Supplementary Information Section 4), we combined 122 communities into 88 groups of communities by assigning communities within three kilometres of each other to the same group. We randomly assigned these 88 groups to the four treatments. In each community, we collected baseline attitudinal data using the implicit association test, and several weeks later we screened the appropriate movie a single time. We collected further attitudinal data approximately one week (5–24 days, mean 6.46 days, with 6 and 7 days accounting for more than 90% of observations) after the movie screening.

The two experiments complemented each other methodologically. By randomizing at the individual level, we conducted the first experiment in a given community in about three hours, which maximized experimental control. All participants watched the assigned movies, and we were thus able to treat everyone we intended to treat. Moreover, apart from one participant who did not complete the implicit association test, we have data for everyone who watched a movie. Finally, we were able to determine whether the movies had an immediate effect on viewer attitudes.

Randomizing within communities, however, was not well-suited to identifying effects over a longer time period. Had we returned to measure attitudes after one week, for example, participants assigned to different treatments within communities would have presumably discussed the movies they saw, and the treatments would have contaminated each other as a result. In the second experiment, we addressed this problem by randomly assigning entire groups of communities in a given geographical area to a single treatment.

However, because the second experiment involved three activities over several weeks, not all recruited participants were able to see a movie. Movie attendance was very high but not complete (80.2–87.1%; Supplementary Information Section 4 and Supplementary Table 4). This meant that we had to distinguish between effects associated with intending to treat participants and effects associated with participants who saw their assigned movies (Supplementary Information Section 4). Furthermore, participation in the two data collection events was high, but again not complete (88.4–93.9%; Supplementary Information Section 4 and Supplementary Table 4). Therefore, we used both inverse probability weighting[25] and multiple imputation[26] to ensure that results were robust, even with missing observations (Supplementary Information Section 4).

Our implicit association test measured attitudes about cut versus uncut girls (Supplementary Information Section 1). The test generated D scores for individual participants, where $-2 \leq D \leq 2$. A negative score, $-2 \leq D < 0$, indicates that the subject had relatively negative associations with uncut girls, whereas a positive score, $0 < D \leq 2$, indicates that the subject had relatively positive associations with uncut girls. A score of $D = 0$ indicates that the subject had no implicit associations. Implicit association tests have high predictive validity when used to address socially sensitive topics with subjects who have little previous experience with such measures[27]. Implicit attitudinal measures have also been shown to reduce or eliminate social desirability bias when measuring attitudes related to female genital cutting[28]. We validated our test by showing that average implicit attitudes by community were significantly correlated with cutting rates by community across 45 communities in Gezira (Extended Data Fig. 1, Pearson's correlation, $\rho = -0.423$, one-sided $P = 0.0008$, based on a two-dimensional weighted bootstrap explained in Supplementary Information Section 1).

After the implicit association test, each participant completed a short questionnaire in which we collected data on gender, age, endogamy, whether the participant had sons or daughters and whether the participant and the participant's family had a history of nomadism. In the area in Gezira where the research was performed, two common forms of production were farming and herding. One hypothesized motive for cutting and infibulation is to reduce fears about paternity uncertainty[6]. If this motive is important, we predicted that high mobility would have made such fears more important among herders than among farmers, and families with a history of nomadism would have had relatively negative attitudes about uncut girls. With the exception of age and sex, all data were collected under anonymous laboratory conditions (Supplementary Information Section 1), which can be effective in reducing social desirability biases when researching culturally sensitive issues[29].

For the first experiment, none of the control variables were significantly related to attitudes (Extended Data Table 1). The three movies about cutting all produced large, robust and significant increases in positive attitudes about uncut girls. The three experimental treatments improved attitudes towards uncut girls with effects of 55–64% of a standard deviation in D scores (Extended Data Table 1; $P \leq 0.001$ or $0.001 < P \leq 0.01$). Only 35.6% of participants in the control treatment

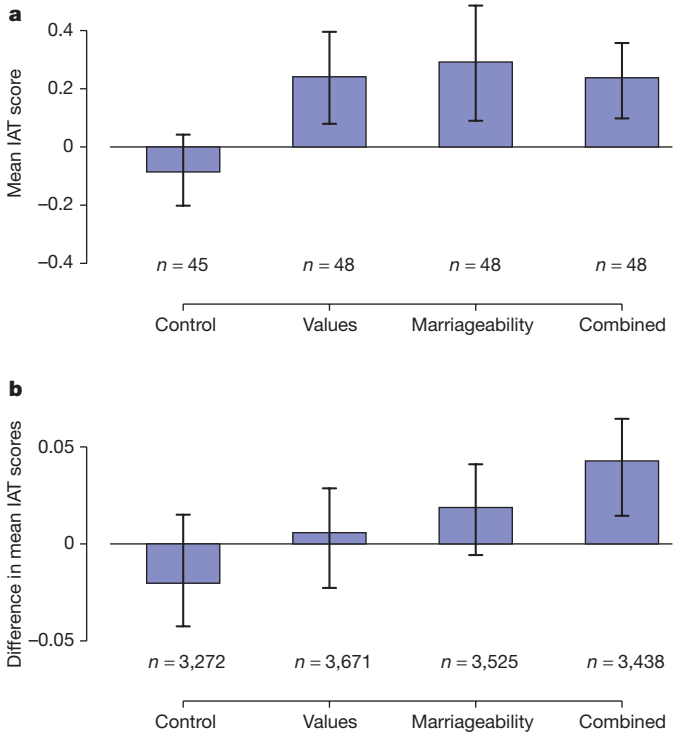


Figure 1 | Treatment effects for unnormalized implicit association scores. Larger scores indicate relatively positive attitudes about uncut girls. **a**, Experiment 1, mean implicit association test (IAT) scores as measured immediately after movies (Extended Data Table 1). Error bars show 95% bootstrapped confidence intervals clustered on 40 movie groups, where each group consisted of three to five subjects who watched a movie together. Clustering accounts for dependence due to social interactions during the movies. Data are from 189 randomly sampled participants randomly assigned to treatments. The *n* values indicate the number of observed scores by treatment, where we have one score per participant. **b**, Experiment 2, mean IAT scores from the follow-up data collection, approximately one week after the movies, minus mean scores from the baseline data collection. Error bars show 95% bootstrapped confidence intervals clustered on 88 groups of communities randomly assigned to treatments. Data are from 7,729 participants randomly sampled from 122 communities. These communities were combined, based on geographic proximity, into the 88 groups assigned to treatments. The *n* values indicate the number of observed scores by treatment, where we have either one score (baseline or follow-up) or two scores (baseline and follow-up) per participant (Supplementary Information Section 4). Difference-in-difference estimations (Supplementary Information Section 4) consistently produced significant effects for the combined treatment of 10.1–11.3% of a standard deviation in *D* scores (Extended Data Tables 2, 3, 6, 7).

had positive associations with uncut girls ($D > 0$), whereas 66.7–72.9% of participants in the experimental treatments had positive associations. The effects of the three experimental treatments were indistinguishable from each other (Fig. 1a and Extended Data Table 1). Arguments related to individual values, marriageability or a combination of the two all led to similar improvements in attitudes towards uncut girls. This finding suggests that both individual values and marriageability concerns influence attitudes about cutting and attitudinal change.

The results of the second experiment confirmed the combined importance of individual values and marriageability. We found a robust and significant improvement in attitudes about uncut girls, but only for the combined movie. The combined movie improved implicit association scores with effects of approximately 10–11% of a standard deviation in *D* scores (Extended Data Table 2 with $0.001 < P \leq 0.01$ and Extended Data Table 3 with $0.001 < P \leq 0.01$ or $0.01 < P \leq 0.05$). Notably, the effects of the combined movie were heterogeneous (Extended Data Tables 4–7). In a follow-up analysis, we ranked participants according to their baseline implicit association scores and divided participants at the median. The effect of the combined treatment was largely due to changes among those participants with baseline scores above the median, namely participants who held less negative attitudes about uncut girls at the start of the experiment (Extended Data Tables 6, 7 with $0.01 < P \leq 0.05$).

These results remained consistent when estimating both the intention to treat (Extended Data Tables 2 and 6), which did not account for which participants actually saw their assigned movies, and the effects of the treatments on the movie-goers (Extended Data Tables 3, 7 and Supplementary Information Section 4). Moreover, regardless of how we handled missing observations, effect sizes and significance levels were stable (Extended Data Tables 2–7).

Apart from the effect associated with the combined treatment, we also found highly significant positive effects associated with female participants ($P \leq 0.001$) and with increasing age ($0.001 < P \leq 0.01$), as well as significant positive effects associated with participants having daughters ($0.01 < P \leq 0.05$) (Extended Data Table 2). This indicates that women were more positive about uncut girls than men, older participants were more positive than younger participants, and participants with daughters were more positive than those without. We additionally found significant negative effects associated with nomadism (Extended Data Table 2). The negative effect associated with participants identifying themselves as nomads was about 21% of a standard deviation in *D* scores ($P \leq 0.001$), and the negative effect associated with having ancestors who were nomads was about 15% of a standard deviation in *D* scores ($P \leq 0.001$). These results suggest that a history of paternity concerns[6] could contribute to support for cutting among participants with a nomadic lifestyle.

Our results show that movies emphasizing local heterogeneity in terms of both values and marriageability can change attitudes to favour the abandonment of cutting. Three important implications follow. First, programmes that take local heterogeneity as a starting point for promoting abandonment[2,3] offer a promising avenue for cultural change when a tension exists between human rights as an international directive and a local culture. Our experiments show that dramatizing discordant views on cutting within a family can improve attitudes about uncut girls. Such an approach could be applicable beyond Gezira, Sudan, because empirical evidence has revealed that local heterogeneity in attitudes related to cutting is widespread[3,12,13]. However, to the extent that some areas are characterized by locally homogeneous support for cutting (Supplementary Information Section 5), a focus on local heterogeneity might be less effective in these areas than what we observed.

Second, addressing individual values and marriageability concerns together produced attitudinal changes that lasted longer than those associated with addressing values and marriageability separately. The link between cutting and marriageability has been an influential hypothesis[6], and our results confirm that marriageability can be critical. However, our results indicate that exposing variation in individual values is also critical for changing cultural attitudes.

Finally, our results suggest that changing attitudes through entertainment could contribute to the abandonment of cutting. The demand for entertainment is enormous and seemingly ubiquitous[22]. Messaging related to cutting can thus be incorporated, for example, in television shows that are conceived from the beginning as for-profit ventures. In support of this idea, recent studies revealed that commercial television has improved the status of women in India[19] and Brazil[20]. Consequently, even if movies or television shows are expensive to produce and distribute, the demand for entertainment might cover much of the cost. Furthermore, because the demand for entertainment is so widespread, messages embedded in popular entertainment could potentially reach a broad cross-section of the population. This is in contrast, for example, to non-fiction documentaries, which mainly reach those already amenable to the message[22].

Notably, the distinguishing characteristics of entertainment could offset the limitations our results revealed in terms of attitudinal change. Our first experiment showed large immediate improvements

in attitudes about uncut girls after approximately 27 min of material related to cutting. Our second experiment, however, showed smaller effects after one week. Sustained exposure to appropriate messaging could thus be essential for long-term change. If so, covering costs through the demand for entertainment is especially attractive. Our second experiment also revealed that participants who entered the experiment with the most negative attitudes about uncut girls did not change their attitudes. Again, sustained exposure might lead to change among more resistant segments of the population, but these segments must first be reached. The appeal of entertainment is again attractive because it limits the risk that the message never reaches those most in favour of cutting[22]. By positioning the debate about cutting locally, we found that entertaining movies addressing both individual values and marriageability can lead people to have more positive attitudes about uncut girls and provide a means of gently provoking change from within a cutting culture.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 4 December 2015; accepted 16 September 2016.
Published online 12 October 2016.

1. Shell-Duncan, B. & Hernlund, Y. in *Female "Circumcision" in Africa: Culture, Controversy, and Change* (eds Shell-Duncan, B. & Hernlund, Y.) 1–40 (Boulder, 2000).
2. Merry, S. E. *Human Rights and Gender Violence: Translating International Law Into Local Justice* (Univ. Chicago Press, 2006).
3. Hernlund, Y. & Shell-Duncan, B. Contingency, context, and change: negotiating female genital cutting in the Gambia and Senegal. *Afr. Today* **53,** 43–57 (2007).
4. Bishai, D. *et al.* Estimating the obstetric costs of female genital mutilation in six African countries. *Bull. World Health Organ.* **88,** 281–288 (2010).
5. World Health Organization. Female genital mutilation. http://www.who.int/topics/female_genital_mutilation/en/ (2014, accessed 16 September 2015).
6. Mackie, G. in *Female "Circumcision" in Africa: Culture, Controversy, and Change* (eds Shell-Duncan, B. & Hernlund, Y.) 253–283 (Boulder, 2000).
7. Gruenbaum, E. *The Female Circumcision Controversy: An Anthropological Perspective* (Univ. Pennsylvania Press, 2001).
8. Boddy, J. *Civilizing Women: British Crusades in Colonial Sudan* (Princeton Univ. Press, 2007).
9. Shell-Duncan, B. From health to human rights: female genital cutting and the politics of intervention. *Am. Anthropol.***110,** 225–236 (2008).
10. Thomas, L. in *Female "Circumcision" in Africa: Culture, Controversy, and Change* (eds Shell-Duncan, B. & Hernlund, Y.) 129–150 (Boulder, 2000).
11. Camilotti, G. Interventions to stop female genital cutting and the evolution of the custom: evidence on age at cutting in Senegal. *J. Afr. Econ.* http://dx.doi.org/10.1093/jae/ejv013 (2015).
12. Shell-Duncan, B. & Hernlund, Y. Are there "stages of change" in the practice of female genital cutting?: Qualitative research findings from Senegal and The Gambia. *Afr. J. Reprod. Health* **10,** 57–71 (2006).
13. Bellemare, M. F., Novak, L. & Steinmetz, T. L. All in the family: explaining the persistence of female genital cutting in West Africa. *J. Dev. Econ.* **116,** 252–265 (2015).
14. Efferson, C., Vogt, S., Elhadi, A., Ahmed, H. E. F. & Fehr, E. Female genital cutting is not a social coordination norm. *Science* **349,** 1446–1447 (2015).
15. Shell-Duncan, B., Wander, K., Hernlund, Y. & Moreau, A. Dynamics of change in the practice of female genital cutting in Senegambia: testing predictions of social convention theory. *Soc. Sci. Med.* **73,** 1275–1283 (2011).
16. La Ferrara, E. Mass media and social change: can we use television to fight poverty? SSRN: CEPR Discussion Paper No. *DP10954* (2015).
17. Greiner, K., Singhal, A. & Hurlburt, S. With an antenna we can stop the practice of female genital cutting: a participatory assessment of *Ashreat Al Amal*, an entertainment-education radio soap opera in Sudan. *Investigación & Desarrollo* **15,** 226–259 (2007).
18. The Cruel Cut. *A documentary for ITV4.* http://leylahussein.com/the-cruel-cut/ (2014, accessed 25 May 2016).
19. Jensen, R. & Oster, E. The power of TV: cable television and women's status in India. *Q. J. Econ.* **124,** 1057–1094 (2009).
20. La Ferrara, E., Chong, A. & Duryea, S. Soap operas and fertility: evidence from Brazil. *Am. Econ. J. Appl. Econ.* **4,** 1–31 (2012).
21. Kearney, M. S. & Levine, P. B. Media influences on social outcomes: the impact of MTV's 16 and pregnant on teen childbearing. *Am. Econ. Rev.* **105,** 3597–3632 (2015).
22. DellaVigna, S. & La Ferrara, E. in *Handbook of Media Economics* 1st edn (eds Anderson, S., Waldfogel, J. & Stromberg, D.) Ch. **19** (Elsevier, 2015).
23. Berg, R. C. & Denison, E. Effectiveness of interventions designed to prevent female genital mutilation/cutting: a systematic review. *Studies in Family Planning* **43,** 135–146 (2012).
24. Nosek, B. A., Greenwald, A. G. & Banaji, M. R. in *Automatic Processes in Social Thinking and Behavior* (ed. Bargh, J. A.) 265–292 (Psychology Press, 2007).
25. Robins, J. M., Rotnitzky, A. & Zhao, L. P. Estimation of regression coefficients when some regressors are not always observed. *J. Am. Stat. Assoc.* **89,** 846–866 (1994).
26. Schafer, J. L. *Analysis of Incomplete Multivariate Data* (Chapman & Hall, 1997).
27. Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L. & Banaji, M. R. Understanding and using the implicit association test: III. Meta-analysis of predictive validity. *J. Pers. Soc. Psychol.* **97,** 17–41 (2009).
28. de Cao, E. & Lutz, C. Measuring attitudes regarding female genital mutilation through a list experiment. *CSAE Working Paper WPS/2015-20* (Centre for the Study of African Economies, 2015).
29. Chauchard, S. Using mp3 players in surveys: the impact of a low-tech self-administration mode on reporting of sensitive attitudes. *Public Opin. Q.* **77,** 220–231 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the Gezira State government, the Sudan Nation Council for Child Welfare, the Gezira State Council for Child Welfare, and local authorities in Umalgoura and East Gezira for supporting the study. We acknowledge the efforts of the Gezira State Council for Child Welfare, our many facilitators and participants, and the national field office of UNICEF in Khartoum. We thank O. A. A. Abdalla, W. O. B. Alalfi, and A. Al Shibli for their contributions to the materials for the IAT, W. O. B. Alalfi for his screenplays and M. Banaji, P. Biroli, A. Elhadi and R. Winkelmann for valuable feedback. This study was funded primarily by the Swiss National Committee of UNICEF. It is also part of a European Research Council grant on the Foundations of Economic Preferences. Funding agencies played no role in the design of the study, data collection, data analysis and interpretation, or the writing and submission of the paper.

Author Contributions S.V., E.F. and C.E. initiated the study. S.V. and C.E. designed the experiments with input from E.F. S.V. and C.E. contributed to the writing and production of the movies and developed the implicit association test for measuring attitudes. N.A.M.Z. pre-tested the movies. S.V., N.A.M.Z., H.E.F.A. and C.E. recruited participants and liaised with government and community officials. S.V., N.A.M.Z., H.E.F.A. and C.E. planned and conducted the experiments. C.E. analysed the data with input from S.V. and E.F. S.V., E.F. and C.E. interpreted the results and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.V. (sonja.vogt@econ.uzh.ch), E.F. (ernst.fehr@econ.uzh.ch) or C.E. (charles.efferson@econ.uzh.ch).

Reviewer Information *Nature* thanks N. Christakis, B. Shell-Duncan and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

No statistical methods were used to predetermine sample size. The experiments were randomized as described in the Supplementary Information Sections 3, 4. Data collectors were not blinded to treatment allocation in experiment 1, but they were blinded to treatment variation (i.e. variation in content across the movies). Data collectors were blinded to treatment allocation in experiment 2. Data collectors were blinded to outcome assessment (i.e. response times) in all cases.

In the values movie (Supplementary Information Section 2), characters present arguments based on their individual evaluations of cutting. For some, health is paramount. They want to avoid a decision that could lead to health problems for their daughters. Others worry that Islam requires cutting. Others fear that if they do not cut, their daughters will not grow up to be feminine, morally upright women. These different views share the feature that they are all based on individual values that do not explicitly depend on whether other families cut[15].

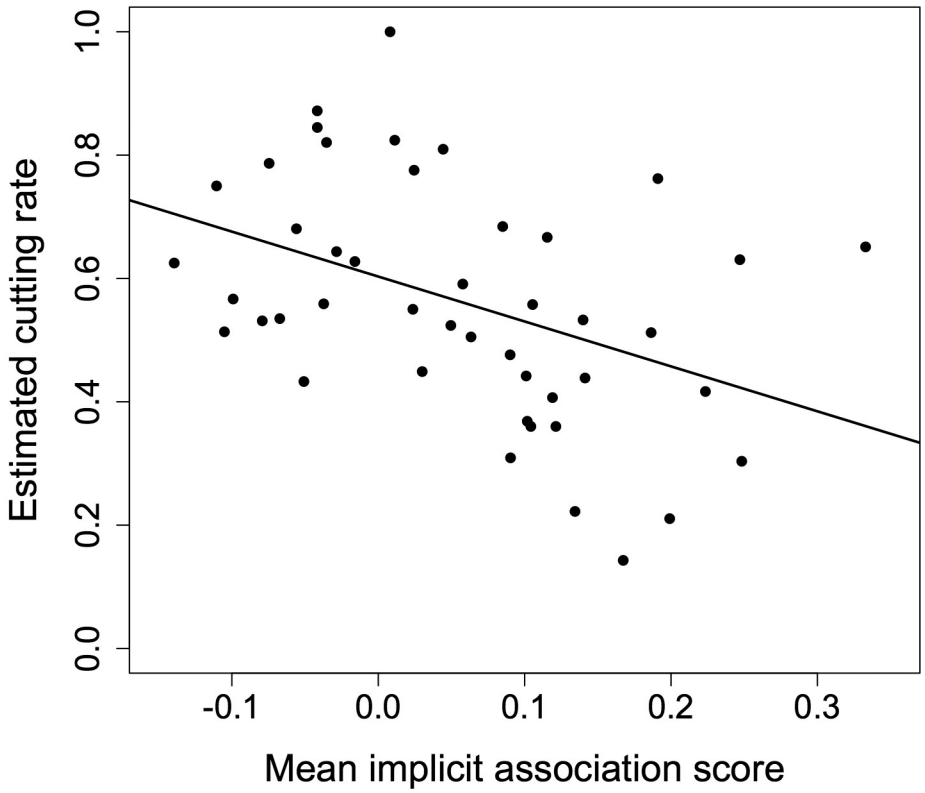
The values movie begins by depicting a set of competing fears and concerns. The characters then examine and eventually alleviate the fears and concerns that favour cutting. For example, in one scene, the mothers discuss a young woman, Samia, who is highly regarded for her good behaviour. The discussion revolves around the question of whether Samia is well-behaved because she is cut or because of how her mother had raised her. After considerable debate, the women conclude that Samia is well-behaved because of how her mother had raised her. Additionally, both the mothers and fathers discuss whether Islam requires cutting. Again after considerable debate, two key points lead them to conclude that it does not. First, the oldest brother in the family lives in Saudi Arabia, and he has revealed that in Saudi Arabia, the birthplace of Islam, families do not cut. Moreover, in another scene, the mothers watch a Sheikh on television as he explains that Islam does not require cutting. To record this scene, we recruited a well-known Sheikh in Sudan who is an advocate for abandoning female genital cutting, and he played himself in the movie.

The marriageability movie (Supplementary Information Section 2) focuses on the question of how cutting would affect the future marriage prospects of the daughters in the family. Specifically, everyone in the family agrees they must consider how cutting would influence whether their daughters grow up to find good husbands. Family members disagree, however, about the details. Some argue that times are changing and that cutting will soon disappear. If the family waits to abandon cutting, they will be viewed as backward, and their daughters will suffer for it. Others concede that some families may have stopped cutting, but they argue that change, if it comes at all, will take generations. Thus, they argue, the best way to guarantee a good future for their daughters right now is to cut them.

The dispute is resolved as family members gather more information. They learn that other families in the area have publicly abandoned cutting, and a record of this is available in the local community office. The family members also discuss the possibility that, if they abandon cutting, the decision might hasten the decision of other families to also stop because these other families are also trying to anticipate the trend. In addition, the family considers that, even if they abandon cutting before most other families, their daughters could still marry within the extended family. Extended families are a common source of marriage partners in Sudan, and distant relatives within extended families often marry. Accordingly, the family recognizes that, if they stop cutting together, the extended family itself would form an interim marriage pool for their daughters until cutting is abandoned more widely. Finally, in a crucial scene the mothers watch a public declaration of abandonment on television. In this scene, several families on television publicly declare that they have abandoned cutting, and by doing so all at once they collectively ensure good marriage prospects for their uncut daughters in the future.

The combined movie (Supplementary Information Section 2) integrates scenes and dialogues from both the values and marriageability movies. In this sense it combines both heterogeneous values related to cutting and heterogeneous beliefs about how cutting could affect future marriage prospects. The control movie consists only of the main plot (Supplementary Information Section 2). All movies are approximately 90 min long. Subplots about cutting constitute approximately 30% of the relevant movies (approximately 27 min). We extended several scenes in the control movie to make the total length 90 min. To do so, we did not change the storyline or narrative structure of the main plot in any way.

We conducted both experiments in the state of Gezira, Sudan (Supplementary Information Sections 3, 4). We used different communities for the two studies. Adult participants came from randomly selected households with a roughly equal mix of men and women in each community. For the second experiment, we collected data from recruited participants, but each movie screening was open to the entire community. The studies were approved by the National Council for Child Welfare in Khartoum, the Gezira State Council for Child Welfare in Wad Madani, the Gezira State Ministries of Health and Education, all relevant authorities in all 127 communities, and the Human Subjects Committee of the Faculty of Economics, Business Administration and Information Technology at the University of Zurich. Participants provided verbal consent after being briefed about the study.



Extended Data Figure 1 | Correlation between attitudes and cutting practices at the community level. The relationship between estimated cutting rates and average scores for each of 45 communities in the state of Gezira, Sudan[14], from the implicit association test (Pearson's correlation, $\rho = -0.423$, one-sided $P = 0.0008$, based on a two-dimensional weighted bootstrap explained in the Supplementary Information Section 1). The unweighted least squares line is shown as a reference.

Extended Data Table 1 | Experiment 1, regression results for the implicit association test

Parameter	Estimate (Std. error)	Estimate (Std. error)
Intercept	-0.442***	-0.762***
	(0.100)	(0.223)
Age		0.002
		(0.006)
Male		-0.083
		(0.141)
Nomad		-0.212
		(0.161)
Ancestors nomads		0.036
		(0.148)
Spouse born same community		0.081
		(0.174)
Participant has daughter(s)		0.127
		(0.222)
Participant has son(s)		0.167
		(0.188)
Values movie	0.553***	0.577***
	(0.161)	(0.172)
Marriageability movie	0.639**	0.653***
	(0.194)	(0.191)
Combined movie	0.548***	0.619***
	(0.147)	(0.141)

*** $p \leq 0.001$ ** $p \in (0.001, 0.01]$ * $p \in (0.1, 0.05]$

The response variable consisted of implicit association scores normalized to have a mean of 0 and a standard deviation of 1. Standard errors were calculated by clustering at the level of movie group, a procedure that accounts for any dependence due to social interactions while watching the movies. Treatment coefficients showed large, significant improvements in participant attitudes towards uncut girls. The three experimental treatments were indistinguishable from each other (model without controls, Wald tests: values = marriageability, $F = 0.169$, $P = 0.682$; values = combined, $F = 0.001$, $P = 0.973$; marriageability = combined, $F = 0.212$, $P = 0.646$). Because two participants did not complete the anonymous questionnaire after the implicit association test, the model with controls was fit to 187 observations, whereas the model without controls was fit to all 189 observations.

Extended Data Table 2 | Experiment 2, intention-to-treat effects estimated by difference-in-difference

	Complete cases		Inverse prob. weighting		Multiple imputation	
	Estimate (Std. error)	Estimate (Std. error)	Estimate (Std. error)	Estimate (Std. error)	Estimate (Std. error)	Estimate (Std. error)
Intercept	-0.0544 (0.0492)	-0.101 (0.0703)	-0.0536 (0.0491)	-0.0984 (0.0696)	-0.0556 (0.0517)	-0.0993 (0.0700)
Woman		0.0938*** (0.0253)		0.0933*** (0.0253)		0.0936*** (0.0241)
Age		0.00219** (0.000818)		0.00218** (0.000816)		0.00215** (0.000790)
Spouse born same community		-0.0323 (0.0218)		-0.0340 (0.0217)		-0.0319 (0.0217)
Nomad		-0.213*** (0.0231)		-0.211*** (0.0231)		-0.212*** (0.0229)
Ancestors nomads		-0.150*** (0.0235)		-0.147*** (0.0233)		-0.149*** (0.0229)
Participant has daughter(s)		0.0521* (0.0238)		0.0516* (0.0239)		0.0528* (0.0241)
Participant has son(s)		0.0515 (0.0277)		0.0526 (0.0273)		0.0505 (0.0271)
East Gezira	0.0550 (0.0367)	0.0212 (0.0318)	0.0548 (0.0367)	0.0205 (0.0316)	0.0538 (0.0381)	0.0196 (0.0335)
High cutting group of communities	0.0164 (0.0396)	0.0144 (0.0340)	0.0159 (0.0394)	0.0126 (0.0335)	0.0192 (0.0405)	0.0172 (0.0353)
Large population group of communities	0.0604 (0.0407)	0.0407 (0.0342)	0.0596 (0.0405)	0.0402 (0.0338)	0.0592 (0.0415)	0.0395 (0.0353)
Follow-up	-0.0348 (0.0265)	-0.0463 (0.0268)	-0.0341 (0.0265)	-0.0464 (0.0267)	-0.0377 (0.0281)	-0.0495 (0.0278)
Values movie	-0.0905 (0.0623)	-0.0610 (0.0542)	-0.0899 (0.0622)	-0.0628 (0.0540)	-0.0890 (0.0636)	-0.0611 (0.0556)
Marriageability movie	-0.0550 (0.0477)	-0.0500 (0.0413)	-0.0550 (0.0477)	-0.0508 (0.0412)	-0.0526 (0.0498)	-0.0486 (0.0433)
Combined movie	-0.0201 (0.0532)	-0.00951 (0.0465)	-0.0202 (0.0533)	-0.00923 (0.0464)	-0.0160 (0.0545)	-0.00796 (0.0478)
Follow-up × Values movie	0.0450 (0.0369)	0.0406 (0.0371)	0.0447 (0.0370)	0.0423 (0.0370)	0.0445 (0.0389)	0.0411 (0.0388)
Follow-up × Marriageability movie	0.0675* (0.0338)	0.0611 (0.0339)	0.0667 (0.0340)	0.0610 (0.0339)	0.0667 (0.0366)	0.0607 (0.0361)
Follow-up × Combined movie	0.111** (0.0345)	0.113** (0.0352)	0.111** (0.0346)	0.111** (0.0350)	0.106** (0.0382)	0.110** (0.0376)

*** $p \leq 0.001$ ** $p \in (0.001, 0.01]$ * $p \in (0.1, 0.05]$

The response variable consisted of implicit association scores normalized to have a mean of 0 and a standard deviation of 1. Interactions between the follow-up data collection and the three experimental treatments estimated the difference-in-difference effects relative to the control treatment (for example, follow-up × combined movie; Supplementary Information Section 4). These effects are highlighted in blue. To ensure results robust to missing data, we analysed the unweighted complete cases, the inverse probability weighting of complete cases and multiple imputations. The complete cases included 13,906 observations from 7,729 participants. Under multiple imputations, we imputed 1,552 observations for each of 200 imputations. For treatment assignment, 122 communities were combined into 88 groups according to geographic proximity. Robust standard errors were calculated by clustering on these 88 groups.

Extended Data Table 3 | Experiment 2, the effects of the treatments on the movie-goers

	Complete cases		Inverse prob. weighting		Multiple imputation	
	Estimate (Std. error)	Estimate (Std. error)	Estimate (Std. error)	Estimate (Std. error)	Estimate (Std. error)	Estimate (Std. error)
Values movie	0.0361 (0.0404)	0.0322 (0.0407)	0.0362 (0.0405)	0.0344 (0.0406)	0.0352 (0.0407)	0.0304 (0.0414)
Marriageability movie	0.0648 (0.0343)	0.0601 (0.0352)	0.0643 (0.0343)	0.0605 (0.0349)	0.0657 (0.0372)	0.0629 (0.0378)
Combined movie	0.106** (0.0369)	0.107** (0.0379)	0.106** (0.0368)	0.105** (0.0374)	0.102* (0.0390)	0.104* (0.0399)
Community controls	✓	✓	✓	✓	✓	✓
Individual controls		✓		✓		✓

*** $p \leq 0.001$ ** $p \in (0.001, 0.01]$ * $p \in (0.1, 0.05]$

The response variable consisted of implicit association scores normalized to have a mean of 0 and a standard deviation of 1. Estimates show the difference-in-difference effects by experimental treatment relative to the control treatment (Supplementary Information Section 4). To ensure results were robust to missing data, we analysed the unweighted complete cases, the inverse probability weighting of complete cases and multiple imputations. Data included 13,906 observations from 7,729 participants. Under multiple imputations, we imputed 1,552 observations for each of 200 imputations. For treatment assignment, 122 communities were combined into 88 groups according to geographic proximity. Robust standard errors were calculated by clustering on these 88 groups.

Extended Data Table 4 | Experiment 2, intention-to-treat effects for subjects with a baseline implicit association score at or below the median

	Complete cases		Inverse prob. weighting		Multiple imputation	
	Estimate (Std. error)	Estimate (Std. error)	Estimate (Std. error)	Estimate (Std. error)	Estimate (Std. error)	Estimate (Std. error)
Intercept	-0.856*** (0.0420)	-0.772*** (0.0646)	-0.855*** (0.0420)	-0.769*** (0.0650)	-0.856*** (0.0428)	-0.767*** (0.0654)
Woman		0.0210 (0.0198)		0.0207 (0.0199)		0.0204 (0.0196)
Age		0.000644 (0.000776)		0.000611 (0.000797)		0.000611 (0.000766)
Spouse born same community		0.00268 (0.0225)		0.00139 (0.0226)		0.00403 (0.0223)
Nomad		-0.0669** (0.0234)		-0.0652** (0.0235)		-0.0671** (0.0234)
Ancestors nomads		-0.0463 (0.0237)		-0.0469 (0.0240)		-0.0474* (0.0235)
Participant has daughter(s)		-0.0621* (0.0290)		-0.0596* (0.0292)		-0.0619* (0.0290)
Participant has son(s)		-0.0227 (0.0325)		-0.0232 (0.0329)		-0.0254 (0.0323)
East Gezira	0.0333 (0.0235)	0.0273 (0.0234)	0.0341 (0.0236)	0.0269 (0.0236)	0.0341 (0.0246)	0.0275 (0.0245)
High cutting group of communities	0.0288 (0.0253)	0.0289 (0.0243)	0.0285 (0.0252)	0.0281 (0.0242)	0.0302 (0.0263)	0.0307 (0.0255)
High population group of communities	-0.0186 (0.0281)	-0.0248 (0.0272)	-0.0195 (0.0280)	-0.0258 (0.0272)	-0.0203 (0.0289)	-0.0274 (0.0283)
Follow-up	0.402*** (0.0417)	0.401*** (0.0416)	0.402*** (0.0417)	0.399*** (0.0414)	0.399*** (0.0414)	0.399*** (0.0414)
Values movie	-0.0501 (0.0429)	-0.0391 (0.0420)	-0.0499 (0.0427)	-0.0399 (0.0419)	-0.0501 (0.0430)	-0.0391 (0.0421)
Marriageability movie	-0.00558 (0.0380)	-0.000844 (0.0377)	-0.00566 (0.0379)	-0.00158 (0.0375)	-0.00580 (0.0381)	-0.00120 (0.0377)
Combined movie	-0.0177 (0.0366)	-0.0116 (0.0364)	-0.0176 (0.0366)	-0.0112 (0.0363)	-0.0177 (0.0367)	-0.0117 (0.0364)
Follow-up × Values movie	0.0288 (0.0474)	0.0258 (0.0479)	0.0294 (0.0475)	0.0301 (0.0474)	0.0291 (0.0480)	0.0245 (0.0482)
Follow-up × Marriageability movie	0.0342 (0.0520)	0.0272 (0.0523)	0.0331 (0.0520)	0.0266 (0.0521)	0.0362 (0.0525)	0.0290 (0.0528)
Follow-up × Combined movie	0.0605 (0.0601)	0.0585 (0.0608)	0.0602 (0.0599)	0.0579 (0.0598)	0.0595 (0.0588)	0.0579 (0.0597)

*** $p \leq 0.001$ ** $p \in (0.001, 0.01]$ * $p \in (0.1, 0.05]$

The response variable consisted of implicit association scores normalized to have a mean of 0 and a standard deviation of 1. Interactions between the follow-up data collection and the three experimental treatments (for example, follow-up × combined movie; Supplementary Information Section 4) estimated the difference-in-difference effects relative to the control treatment. These effects are highlighted in blue. To ensure results were robust to missing data, we analysed the unweighted complete cases, the inverse probability weighting of complete cases and multiple imputations. Data included 6,659 observations from 3,541 participants. Under multiple imputations, we imputed 423 observations for each of 200 imputations. For treatment assignment, 122 communities were combined into 88 groups according to geographic proximity. Robust standard errors were calculated by clustering on these 88 groups.

Extended Data Table 5 | Experiment 2, the effects of the treatments on the movie-goers for subjects with a baseline implicit association score at or below the median

	Complete cases		Inverse prob. weighting		Multiple imputation	
	Estimate (Std. error)	Estimate (Std. error)	Estimate (Std. error)	Estimate (Std. error)	Estimate (Std. error)	Estimate (Std. error)
Values movie	0.0217 (0.0538)	0.0184 (0.0540)	0.0223 (0.0540)	0.0221 (0.0538)	0.0220 (0.0542)	0.0198 (0.0542)
Marriageability movie	0.0218 (0.0571)	0.0137 (0.0572)	0.0212 (0.0570)	0.0135 (0.0569)	0.0233 (0.0571)	0.0157 (0.0572)
Combined movie	0.0440 (0.0675)	0.0401 (0.0681)	0.0438 (0.0673)	0.0390 (0.0671)	0.0464 (0.0666)	0.0417 (0.0669)
Community controls	✓	✓	✓	✓	✓	✓
Individual controls		✓		✓		✓

*** $p \leq 0.001$ ** $p \in (0.001, 0.01]$ * $p \in (0.1, 0.05]$

The response variable consisted of implicit association scores normalized to have a mean of 0 and a standard deviation of 1. Estimates show the difference-in-difference effects by experimental treatment relative to the control treatment (Supplementary Information Section 4). To ensure results robust to missing data, we analysed the unweighted complete cases, the inverse probability weighting of complete cases and multiple imputations. Data included 6,659 observations from 3,541 participants. Under multiple imputations, we imputed 423 observations for each of 200 imputations. For treatment assignment, 122 communities were combined into 88 groups according to geographic proximity. Robust standard errors were calculated by clustering on these 88 groups.

Extended Data Table 6 | Experiment 2, intention-to-treat effects for subjects with a baseline implicit association score above the median

	Complete cases		Inverse prob. weighting		Multiple imputation	
	Estimate (Std. error)	Estimate (Std. error)	Estimate (Std. error)	Estimate (Std. error)	Estimate (Std. error)	Estimate (Std. error)
Intercept	0.771***	0.528***	0.771***	0.528***	0.771***	0.525***
	(0.0385)	(0.0574)	(0.0386)	(0.0568)	(0.0392)	(0.0578)
Woman		0.0694**		0.0695**		0.0697**
		(0.0248)		(0.0249)		(0.0241)
Age		0.00397***		0.00402***		0.00408***
		(0.000798)		(0.000797)		(0.000795)
Spouse born same community		-0.0332		-0.0349		-0.0333
		(0.0195)		(0.0196)		(0.0197)
Nomad		-0.108***		-0.108***		-0.110***
		(0.0224)		(0.0225)		(0.0226)
Ancestors nomads		-0.0983***		-0.0975***		-0.0970***
		(0.0232)		(0.0234)		(0.0231)
Participant has daughter(s)		0.0683**		0.0665**		0.0682**
		(0.0249)		(0.0252)		(0.0249)
Participant has son(s)		0.103***		0.103***		0.101***
		(0.0234)		(0.0232)		(0.0238)
East Gezira	0.00295	-0.0261	0.00204	-0.0270	0.00233	-0.0268
	(0.0270)	(0.0250)	(0.0272)	(0.0249)	(0.0278)	(0.0259)
High cutting group of communities	0.00744	0.00865	0.00732	0.00761	0.00911	0.0104
	(0.0300)	(0.0278)	(0.0300)	(0.0277)	(0.0304)	(0.0285)
High population group of communities	0.0932**	0.0805**	0.0942**	0.0818**	0.0922**	0.0792*
	(0.0322)	(0.0298)	(0.0322)	(0.0299)	(0.0327)	(0.0304)
Follow-up	-0.427***	-0.438***	-0.427***	-0.437***	-0.430***	-0.438***
	(0.0321)	(0.0328)	(0.0323)	(0.0332)	(0.0322)	(0.0331)
Values movie	-0.0412	-0.0270	-0.0410	-0.0266	-0.0413	-0.0271
	(0.0352)	(0.0327)	(0.0352)	(0.0325)	(0.0353)	(0.0328)
Marriageability movie	-0.0183	-0.0174	-0.0181	-0.0163	-0.0184	-0.0175
	(0.0333)	(0.0318)	(0.0333)	(0.0316)	(0.0334)	(0.0319)
Combined movie	-0.00176	0.00331	-0.00132	0.00495	-0.00204	0.00316
	(0.0383)	(0.0358)	(0.0383)	(0.0358)	(0.0383)	(0.0358)
Follow-up × Values movie	-0.00814	-0.0101	-0.00864	-0.0113	-0.00596	-0.00839
	(0.0526)	(0.0526)	(0.0529)	(0.0531)	(0.0521)	(0.0523)
Follow-up × Marriageability movie	0.0361	0.0409	0.0357	0.0402	0.0395	0.0429
	(0.0473)	(0.0478)	(0.0475)	(0.0480)	(0.0474)	(0.0479)
Follow-up × Combined movie	0.101*	0.104*	0.101*	0.102*	0.101*	0.104*
	(0.0414)	(0.0416)	(0.0415)	(0.0419)	(0.0423)	(0.0425)

*** $p \leq 0.001$ ** $p \in (0.001, 0.01]$ * $p \in (0.1, 0.05]$

The response variable consisted of implicit association scores normalized to have a mean of 0 and a standard deviation of 1. Interactions between the follow-up data collection and the three experimental treatments (for example, follow-up × combined movie; Supplementary Information Section 4) estimated the difference-in-difference effects relative to the control treatment. These effects are highlighted in blue. To ensure results robust to missing data, we analysed the unweighted complete cases, the inverse probability weighting of complete cases and multiple imputations. Data included 6,647 observations from 3,541 participants. Under multiple imputations, we imputed 435 observations for each of 200 imputations. For treatment assignment, 122 communities were combined into 88 groups according to geographic proximity. Robust standard errors were calculated by clustering on these 88 groups.

Extended Data Table 7 | Experiment 2, the effects of the treatments on the movie-goers for subjects with a baseline implicit association score above the median

	Complete cases		Inverse prob. weighting		Multiple imputation	
	Estimate (Std. error)	Estimate (Std. error)	Estimate (Std. error)	Estimate (Std. error)	Estimate (Std. error)	Estimate (Std. error)
Values movie	-0.0147 (0.0529)	-0.0166 (0.0532)	-0.0151 (0.0532)	-0.0175 (0.0540)	-0.0164 (0.0523)	-0.0163 (0.0523)
Marriageability movie	0.0481 (0.0489)	0.0524 (0.0488)	0.0477 (0.0490)	0.0524 (0.0487)	0.0517 (0.0488)	0.0557 (0.0487)
Combined movie	0.103* (0.0409)	0.105* (0.0407)	0.103* (0.0410)	0.103* (0.0409)	0.104* (0.0420)	0.105* (0.0419)
Community controls	✓	✓	✓	✓	✓	✓
Individual controls		✓		✓		✓

*** $p \leq 0.001$ ** $p \in (0.001, 0.01]$ * $p \in (0.1, 0.05]$

The response variable consisted of implicit association scores normalized to have a mean of 0 and a standard deviation of 1. Estimates show the difference-in-difference effects by experimental treatment relative to the control treatment (Supplementary Information Section 4). To ensure results robust to missing data, we analysed the unweighted complete cases, the inverse probability weighting of complete cases and multiple imputations. Data included 6,647 observations from 3,541 participants. Under multiple imputations, we imputed 435 observations for each of 200 imputations. For treatment assignment, 122 communities were combined into 88 groups according to geographic proximity. Robust standard errors were calculated by clustering on these 88 groups.