



ARTICLE




<https://doi.org/10.1057/s41599-021-00730-z>

OPEN

Modelling collective action to change social norms around domestic violence: social dilemmas and the role of altruism

Lu Gram ¹, Rolando Granados¹, Eva M. Krockow², Nayreen Daruwalla³ & David Osrin¹

Interventions promoting collective action have been used to prevent domestic violence in a range of settings, but their mechanisms of operation remain unclear. We formalise and combine feminist theoretical approaches to domestic violence into a game-theoretic model of women's collective action to change gendered social norms and outcomes. We show that social norms create a social dilemma in which it is individually rational for women to abstain from action to prevent domestic violence among neighbours, but all women suffer negative consequences if none take action. Promoting altruism among women can overcome the social dilemma. Discouraging women from tolerating domestic violence, imposing additional external punishment on men for perpetrating violence, or lowering costs to women of taking action against violence may not work or even backfire. We invite researchers on community mobilisation to use our framework to frame their understandings of collective action to prevent domestic violence.

¹Institute for Global Health, University College London, London, UK. ²Department of Neuroscience, Psychology and Behaviour, University of Leicester, Leicester, UK. ³Prevention of Violence against Women and Children, Society for Nutrition, Education and Health Action (SNEHA), Mumbai, India.
email: lu.gram.13@ucl.ac.uk

Introduction

Worldwide, domestic violence is a critical concern for virtually all aspects of society, with severe human, emotional, and economic costs (Garcia-Moreno and Watts, 2011). One form of domestic violence, intimate partner violence, is estimated to affect 30% of women at least once in their lifetime, and is an important cause of mental, physical, and reproductive harm (Devries et al., 2013; Dillon et al., 2013; Hill et al., 2016). International declarations including the Sustainable development goals have committed national governments to eliminating domestic violence (UN, 2017). However, investments in prevention and services for survivors of violence remain inadequate (Garcia-Moreno and Watts, 2011).

Community mobilisation interventions have long been of interest to policymakers and practitioners as a means of addressing otherwise intractable societal and environmental barriers to improving health (Rosato et al., 2008). They can be defined as interventions in which local individuals collaborate with external agents in identifying, prioritising, and tackling the causes of ill-health based on principles of bottom-up leadership and empowerment (Rosato et al., 2008). For example, interventions in South Africa and Uganda have trained volunteer activists to take action against domestic violence, engaged community groups in reflection and action on gender norms, and organised large-scale campaigns and marches (Abramsky et al., 2014; Pronyk et al., 2006; Wagman et al., 2015).

Randomised controlled trials have shown that such interventions can reduce domestic violence in contexts of severe poverty and gender inequality (Bourey et al., 2015). However, the mechanism through which interventions achieve impact remains poorly understood (Gram et al., 2019). Existing theoretical frameworks display 'positive a priori bias' (Abimbola, 2019), in which interventions are assumed to produce positive engagement with communities without complications. The failure of recent interventions in Rwanda (Chatterji et al., 2020), Afghanistan (Gibbs et al., 2018), and Nepal (Clark et al., 2020) to show comparable impacts on domestic violence to those observed in Uganda (Abramsky et al., 2014; Wagman et al., 2015) underlines the need to understand contexts and mechanisms.

Community mobilisation interventions are complex interventions that involve long causal chains from implementation to outcome, multiple recursive feedback loops, and emergent outcomes (Anderson, 2008). Social scientists often use mathematical models to make explicit their assumptions behind implicit, verbal models of social phenomena (Epstein, 2008; Heckman, 2005; Oliver and Myers, 2002). This enables them to derive predictions using algebra or computer simulation, check for logical

consistency of verbal explanations, and illuminate core uncertainties in existing evidence. Here, the goal is not to forecast the future or mirror reality as closely as possible, but rather to generate explanatory clarity (Epstein, 2008). Intervention researchers have proposed mathematical models as a tool to study the system dynamics of complex public health interventions (Davey et al., 2018).

We use a game-theoretic model to formalise oft-hypothesised processes of action to challenge unequal gender norms in community mobilisation interventions in low- and middle-income contexts. In so doing, we follow an established tradition in the social sciences of using mathematical models as thought experiments to generate logically consistent explanations for social phenomena (Centola et al., 2005; Oliver, 1993). We show how we can define conditions under which the prevention of domestic violence takes on the characteristics of a *social dilemma* (Gram et al., 2019), a situation in which it is individually rational for women to take as little action against domestic violence as possible, even though all women would be better off collectively if everybody invested in such action. We show that elaborations to the model based on assuming a degree of altruism in women can suggest solutions and provide new directions for research and policy.

Background

According to the socio-ecological framework (Heise, 1998), domestic violence results from the interplay of multiple factors at different levels of a social ecology: at the individual level, childhood experiences such as child abuse or marital violence between one's own parents are major factors; at the relationship level, alcohol use, marital conflict and male control over decision-making play a role; at the community level social isolation of women and delinquent peer associations of men contribute; at the societal level, ideologies of male superiority and rigid gender roles also contribute. The framework has been adapted to centre the role of patriarchal ideology, male entitlement, and household gender roles (Jewkes et al., 2002). Figure 1 shows such an adapted framework.

Social ecological frameworks provide an overview of the complexity of domestic violence, but do not generally provide specifics as to how societal ideologies manifest at the individual level or how individual actions can reproduce ideologies. Two feminist theories, the theory of Hegemonic Masculinity (Connell, 1987) and Male Peer Support Theory (DeKeseredy and Schwartz, 2013) propose a mechanism: Men belong to a gendered power

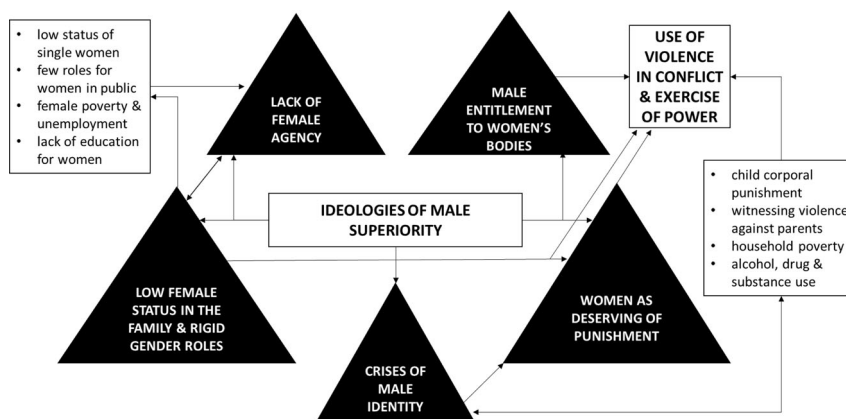


Fig. 1 Overall framework for the determinants of domestic violence adapted from Jewkes et al. (2002). Ideologies of male superiority combine with structural and individual risk factors to promote men's use of violence in domestic relationship.

Table 1 Actions taken by community members to address violence against women in a community-based intervention in Mumbai, India. Adapted from (Daruwalla et al., 2019).

Incident and action	Form of violence
A local volunteer helped a bride-to-be deal with dowry demands from the groom’s family which risked escalating into violence. The volunteer explained the law and her rights to the bride-to-be and discussed these with her family as well.	Domestic violence
A local volunteer intervened when a couple were fighting in the street, accompanied by their two young children. They had doused themselves in kerosene and were threatening to set fire to themselves. A volunteer talked them out of their plan, took their matchsticks away from them and called the police. The couple returned later to thank the volunteer.	Domestic violence
A woman fled to her mother’s home after her partner hit her, but he came after her and attacked both her and her mother. They shouted for help and members of the local woman’s group detained the man and called the police, who arrested him. The women persuaded the couple enter into a counselling programme.	Domestic violence
A woman told a community organiser that her neighbour was being beaten and locked in the house by her partner. The community organiser organised campaigns in the area and gathered a group who visited the house repeatedly, heard the woman inside, and got the police to effect entry. The local women and the community organiser persuaded the woman and her partner to enter into a counselling programme.	Domestic violence
A local volunteer heard her neighbour’s 7-year-old daughter shouting for help when a local man attempted to rape her. She gained access to the house, prevented him from leaving, and called for help. A group of neighbours took him to the police station, where he was arrested.	Child sexual abuse
A local volunteer led a women’s group to repeatedly confront a group of drug users who were harassing women. The harassment stopped and the volunteer became a community leader in a male-dominated area.	Sexual harassment
A young married woman was being harassed by two gang members. When she rejected their advances, they beat her up, set her on fire and locked her in her house. At the public mourning after her death, a large group of relatives and friends banded together and communicated with the police to ensure due process. The perpetrators were jailed.	Assault and murder

hierarchy in which they are rewarded by other men for conforming to stereotypically masculine gender roles, while men who break these norms are punished. Men are socialised into believing that gender non-conforming behaviour from female partners threatens their performance of masculine gender. This makes men use domestic violence to punish female partners for gender non-conforming behaviours. In other words, ‘social norms’, defined as expectations of informal reward or sanction from others for conforming to or deviating from a behaviour, play a key role (Hechter and Opp, 2001).

Qualitative and quantitative studies find that women’s ‘disobedience’ is cited as one of the most frequent reasons for domestic violence, whether due to neglect of household work, refusal to have sex, or arguments with the husband and in-laws (Neal and Edwards, 2017). Survey studies have found that societal acceptance of wife-beating as punishment correlates with experience of violence (García-Moreno et al., 2005; Heise and Kotsadam, 2015), while accepting attitudes among peers correlates with perpetration of violence (Mulawa et al., 2018). Other studies have found one of the most important predictors of male perpetration of violence to be the presence of male peers who explicitly verbally encourage it (DeKeseredy and Schwartz, 2013). In qualitative studies, men who engage with violence prevention programmes have encountered considerable peer resistance ranging from mockery to ostracism (Gibbs et al., 2015; Mogford et al., 2015), and fear of losing peer support has been a formidable barrier to behaviour change (Daruwalla et al., 2017).

Individual behaviour change interventions primarily seek to furnish women and men with the resources, knowledge, and skills to negotiate violence in their own relationships (Bandiera et al., 2016; Green et al., 2015; Hossain et al., 2014). These may fail to work in contexts in which pervasive social norms condone or reward male perpetrators’ use of violence to maintain control over female partners (Daruwalla et al., 2017; DeKeseredy and Schwartz, 2013). A trial of a government intervention in Papua New Guinea to prevent domestic violence through increased police presence showed no impact, as men mobilised in response to the programme to maintain power over women (Cooper, 2018). Trials of interventions to financially empower women through economic self-help groups without highly participatory

group education components have generally failed to show impact on intimate partner violence (Gram et al., 2020).

Community mobilisation interventions have been developed to challenge domestic violence-related social norms through a range of activities (Abramsky et al., 2014; Pronyk et al., 2006; Wagman et al., 2015). This includes engaging police, courts, shelter homes, physical and mental health services, and non-specialist community members (Daruwalla et al., 2019). They actively involve community members in decisions about intervention delivery, including the development of strategies to tackle local priority issues (Rosato et al., 2008). Their open-ended approach makes such interventions effective, but also unpredictable (Gram, et al., 2019). For example, a documentation exercise for a community-based intervention to prevent violence against women in Mumbai, India, found a plethora of community actions (Daruwalla et al., 2019) (Table 1). These ranged from negotiating with the family of a woman facing dowry harassment to organising a campaign to free a woman who was locked in her house and beaten by her partner. Amid such heterogeneity, common threads are hard to discern.

However, most theories of change for community mobilisation intervention emphasise community-driven attitude and social norm change as key drivers of reductions in levels of domestic violence (Gram et al., 2019). Process evaluations have evidenced the key role of ‘organised diffusion’—active, systematic efforts by community members themselves in diffusing anti-violence messages the wider community – in achieving impact (Cislaghi et al., 2019). The measurement and tracking of attitudes to violence against women is a standard component of impact evaluations (Abramsky et al., 2016; Chatterji et al., 2020; Kim et al., 2007). Social norms are increasingly monitored in high-income (Fabiano et al., 2003; Kilmartin et al., 2008) and low- and middle-income settings (Clark et al., 2018) too.

The next section formalises literature from this section into a stylised game-theoretic model¹. Game theory is a formal theoretical framework for analysing interactive decision-making (Ferguson, 2013). A ‘game’ is any situation with multiple decision-makers (‘players’) whose choices impact on another. Game theory predicts the behaviour of players in a setting, where the ‘payoffs’ to strategies chosen by individuals depends on strategies adopted by

other individuals in the same population (McAdams et al., 2020). Game theory is widely applied in the psychological and social sciences to model trust, cooperation, and collective action (Ferguson, 2013). Public health researchers have used it to model physician prescribing behaviour in the face of rising antimicrobial resistance (Colman et al., 2019), population behaviour under voluntary vaccination (Bauch and Earn, 2004), and social distancing behaviour during COVID-19 (McAdams, 2020), but not domestic violence prevention. The following sections assume basic familiarity with game theory, but our final discussion of research and policy implications can be grasped without prior knowledge.

Setting up the model

Stylised assumptions. Suppose a set of women w_1, \dots, w_n live in a locality with male² partners h_1, \dots, h_n , where woman w_i has partner h_i . We distinguish between ‘positive assumptions’, which concern the world as it is, and ‘normative assumptions’, which concern the world as it should be. For ease of reference, we have listed common variables used in mathematical formulae throughout this article in Table 2. Proofs of all theorems and propositions are available in a supplementary Technical Appendix. We make the following positive assumptions:

- I. Attitudes matter: men with more pro-violent attitudes are—everything being equal—more likely to perpetrate domestic violence (Abramsky et al., 2016; Flood and Pease, 2009; Fulu et al., 2013; García-Moreno et al., 2005). We assume that each male partner has attitude $a_i \in (-\infty, \infty)$ to violence, where higher values denote more pro-violence attitudes and lower values denote more anti-violence attitudes. Each male partner derives utility $n_{ij}a_i v_i$ from perpetrating violence at level v_i for constant $n_{ij} > 0$. For notational convenience, we call the payoff rate n_{ij} so that it matches the notation below for social norm payoffs. To avoid corner cases that do not change our main argument, we assume $v_i \in (-\infty, \infty)$.
- II. Social norms matter: men reward or punish peers for being violent. Men with more pro-violent attitudes are more likely to reward rather than punish violence in peers (Dworkin et al., 2013; Gibbs et al., 2015; Mogford et al., 2015; Peacock and Levack, 2004; Verma et al., 2006). We assume male partner i perpetrating violence at level v_i receives norm-based reward or sanction $\sum_{j \neq i} n_{ij} a_j v_i$ from

- other men h_j for constants $n_{ij} > 0$. Positive values of $n_{ij} a_j$ indicate reward, while negative values indicate sanction. Positive values do not necessarily indicate reward for violence per se, but can also indicate reward for ‘enforcing discipline’ and fulfilling masculine gender roles in the household (Dworkin et al., 2013; Gibbs et al., 2015; Mogford et al., 2015; Peacock and Levack, 2004; Verma et al., 2006).
- III. Perpetration of violence incurs costs: perpetration of violence is not completely without psychological, social or legal cost (Bloch and Rao, 2002; Busby, 1999; Go et al., 2003; Haushofer et al., 2019). Everything else being equal, men do not prefer infinite levels of domestic violence, nor does society tolerate infinite amounts. We assume man h_i perpetrating violence at level v_i experiences cost $\frac{1}{2} c_i v_i^2$, where $c_i > 0$ is a constant.
- IV. Costly preventive action is possible: women taking action against domestic violence are able to change men’s attitudes, including those of their own partners (Abramsky et al., 2014; Daruwalla et al., 2019; Pronyk et al., 2006; Wagman et al., 2015), but this requires time and effort. Let $e_{ij} \in [0, \infty)$ denote effort spent by woman j in changing the attitude of man i , with larger values denoting greater effort. Then we let $a_i(e_{i1}, \dots, e_{in}, 0)$ be strictly decreasing, strictly convex functions on domain $[0, \infty)^n$. We assume, the a_i are infinitely smooth functions. We model costs of effort for woman w_j as $\frac{1}{2} \sum_k d_{kj} e_{kj}^2$ for constants $d_{kj} > 0$. Note, efforts to change men’s attitudes do not exclusively involve women directly speaking to men and changing their minds (Cislighi et al., 2019; Starmann et al., 2018), but can also involve women asking for support from men’s family members, asking local authorities to speak to a man, collectively protesting a man’s behaviour, or persuading couples to enter counselling programmes (Goodman and Smyth, 2011; Mannell et al., 2018).
- V. Violence creates suffering for survivors, but survivors themselves may tolerate it: women surviving violence suffer severe harm, whether mental, physical, sexual or economic (Devries et al., 2013; Dillon et al., 2013; Hill et al., 2016). However, survivors vary in the extent to which they perceive violence towards themselves as unacceptable or unjust (Busby, 1999; Go et al., 2003; Krause et al., 2016).

Table 2 List of variables. Identifiers refer to specific agents in the world, that is men and women. Parameters refer to fixed model parameters. State variables are states of the world that may change due to agents’ choices.

Variables	Range	Type	Description
n	0,1,2,3,...	Parameter	Total number of couples
w_i	N/A	Identifier	Woman number i , partner to man i
h_i	N/A	Identifier	Man number i , partner to woman i
a_i	$(-\infty, \infty)$	State variable	Attitude to violence of man i
e_{ij}	$[0, \infty)$	State variable	Effort of woman j to change man i
n_{ij}	$(0, \infty)$	Parameter	For $i \neq j$: Norm-based payoffs imposed on man i by man j in response to violence perpetrated by man i For $i = j$: Intrinsic payoffs experienced by man i in response to violence perpetrated by himself
v_i	$(-\infty, \infty)$	State variable	Level of violence perpetrated by man i
c_i	$(0, \infty)$	Parameter	Marginal cost of violence to man i
d_{ij}	$(0, \infty)$	Parameter	Marginal cost of effort by woman j to change man i
s_i	$(0, \infty)$	Parameter	Rate of suffering incurred by violence to woman i
t_i	$[0, 1]$	Parameter	Tolerance for violence expressed by woman i
U_{wi}	$(-\infty, \infty)$	State variable	Utility level of woman i
U_{hi}	$(-\infty, \infty)$	State variable	Utility level of man i
W_i	$(-\infty, \infty)$	State variable	Level of welfare of woman i
b	$(0, \infty)$	Parameter	Intrinsic motivation to take action against violence
η	$(0, \infty)$	Parameter	Rate of empathy towards other women’s suffering
ρ	$(0, \infty)$	Parameter	Degree of reciprocity in women’s investment of effort

We assume women experience disutility $s_i(1 - t_i)v_i$ from violence and level v_i . $s_i > 0$ represents the degree of suffering incurred by violence. $0 \leq t_i < 1$ represents the extent to which women tolerate and view as acceptable violence in their own lives.

Additionally, we make the following normative assumption:

- VI. Women’s welfare should be assessed assuming zero tolerance for violence: long-term abuse is known to create ‘adaptive preferences’ in women that lead them to accept conditions of oppression (Sen, 1999). A broad consensus exists that survivors’ claims to domestic violence being justified should not be taken at face value, but evaluated in light of potential adaptive preferences (Nussbaum, 2001). As such, we evaluate woman w_i ’s welfare using $W_i(v_i, e_{1i}, \dots, e_{ni}) = -s_i v_i - \frac{1}{2} \sum_k d_{kj} e_{kj}^2$, which corresponds to setting t_i to 0 in her utility function.

Game specification and solution. Let G be a two-stage multi-player game with players $\{w_1, \dots, w_n, h_1, \dots, h_n\}$. Assume $n \geq 2$. The stages of the game are as follows:

1. Each woman w_i chooses a level of effort e_{ij} in changing each man h_j ’s attitudes to violence, including those of their own partner.
2. Each man h_i chooses a level of violence v_i to perpetrate against his partner w_i .

Payoffs are awarded to men and women depending on their actions using utility functions $U_{w_i}(v_i, e_{1i}, \dots, e_{ni})$ for women and $U_{h_i}(v_i, e_{1i}, \dots, e_{ni})$ for men, where:

$$U_{h_i} = n_{ii}a_i(e_{i1}, \dots, e_{in})v_i + \sum_{j \neq i} n_{ij}a_j(e_{j1}, \dots, e_{jn})v_i - \frac{1}{2} c_i v_i^2 \tag{1}$$

$$U_{w_i} = -s_i(1 - t_i)v_i - \frac{1}{2} \sum_k d_{kj} e_{kj}^2 \tag{2}$$

Players make moves seeking to maximise payoffs U_{w_i} and U_{h_i} . However, women’s welfare deriving from these moves is evaluated using W_i (Assumption VI). One can show:

Proposition 1 (Nash equilibrium). Suppose n_{ij} and a_j have bounded first-order derivatives, i.e., there exists a constant $M > 0$ such that $\left| \frac{\partial n_{ij}}{\partial a_j} \right|, \left| \frac{\partial a_j}{\partial e_{ji}} \right| < M$ for all i, j . Then a unique subgame perfect equilibrium exists, in which all levels of violence v_j^* and effort e_{ji}^* satisfy:

$$v_i^* = \frac{1}{c_i} \left(n_{ii}a_i(e_{i1}, \dots, e_{in}) + \sum_{j \neq i} n_{ij}a_j(e_{j1}, \dots, e_{jn}) \right) \tag{3}$$

$$e_{ji}^* = - \frac{s_i(1 - t_i)n_{ij} \partial a_j}{d_{ki}c_i \partial e_{ji}} \tag{4}$$

For a given set of baseline model parameters, we can represent a community that has not yet received a new policy or intervention by the resulting equilibrium levels of effort, pro-violent attitude, and perpetration of violence. Changes to equilibria resulting from altering model parameters represent simulated impacts of new policies or interventions. For example, raising costs of violence might lower perpetration of violence relative to a set of baseline parameters. This would suggest that a new intervention might be able to prevent violence in the modelled community by imposing additional costs to domestic violence.

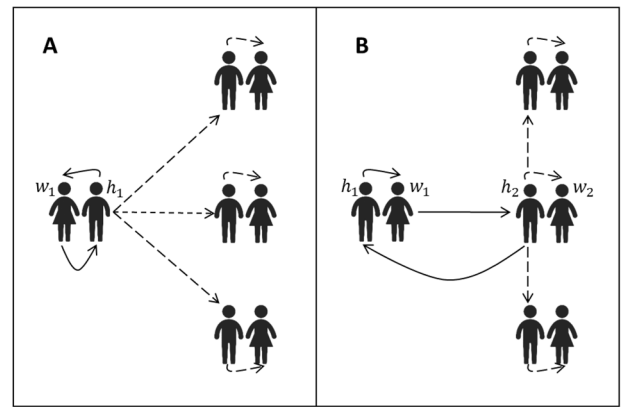


Fig. 2 Externalities to women’s efforts to change men’s attitudes to domestic violence. **A** Woman w_1 ’s efforts to change her partner h_1 ’s attitudes to violence affect her welfare, because she becomes less likely to perpetrate domestic violence (solid lines), but this produces externalities for other women (dotted lines). **B** Woman w_1 ’s efforts to change neighbouring man h_2 ’s attitudes to violence affect her welfare, because h_2 will be less willing to encourage h_1 to perpetrate domestic violence (solid lines). However, her efforts produce externalities for other women’s relationships, including the relationship between h_2 and w_2 (dotted lines).

In general, parameter changes do not result in straightforward changes to equilibrium values. For example, making all women less tolerant of violence may not increase all women’s level of effort: Higher effort expended by woman w_k on man h_j may disincentivise woman w_i from engaging with the same man, if he has already been persuaded by the efforts of w_k . The net effect might be an increase in effort for some women, but a decrease for others. If the net effect of changing model parameters on women’s effort to prevent violence is ambiguous, then so might be the net effect of changing model parameters on final levels of violence. However, we can show that the latter effect is mostly unambiguous.

Corollary 1. Suppose $\frac{\partial a_j}{\partial e_{ji}}(e_{j1} \dots e_{jn}) = \alpha h_{ji}(e_{j1} \dots e_{jn})$ and $s_i = \sigma \tilde{s}_i, t_i = \tau \tilde{t}_i, d_i = \gamma \tilde{d}_i, c_i = \kappa \tilde{c}_i, n_{ij} = \nu \tilde{n}_{ij}$, for constants $\tilde{s}_i, \tilde{t}_i, \tilde{c}_i, \tilde{d}_i, \tilde{n}_{ij}$, functions $h_{ji}(e_{j1} \dots e_{jn}) < 0$ and scale factors $\sigma, \tau, \gamma, \kappa, \nu, \alpha > 0$. Then equilibrium levels of violence decrease in σ and increase in τ and γ . The effects of changing κ and ν on levels of violence are ambiguous. Suppose for any fixed set of $e_{ji}, \frac{\partial a_j}{\partial \alpha}(e_{j1} \dots e_{jn}) < 0$. Then equilibrium levels of violence decrease in α too.

Social dilemmas. Women’s efforts to change the attitudes of their own or neighbouring partners produce benefits for other women’s relationships that they do not themselves directly experience (Fig. 2). These benefits are conventionally called ‘positive externalities’ (Cornes and Sandler, 1996). Women may find it individually rational to abstain from taking preventive action, even though they would be better off if all women collectively agreed to act, a situation called a ‘social dilemma’ (Gram et al., 2019). We can prove that a subset of women always exists for whom increasing levels of effort to prevent violence beyond equilibrium levels would improve their individual welfare:

Theorem 1 (Social dilemmas). Suppose two women w_1 and w_2 have equilibrium levels of effort e_{11}^* and e_{22}^* with regard to influencing the attitude of men h_i and h_j for some, $j \in 1, \dots, n$. Then there exist $\delta e_{11}, \delta e_{22} > 0$ such that the welfare of w_1 and w_2 , W_1 and W_2 , would be greater if they simultaneously exerted effort levels $e_{11}^* + \delta e_{11}$ and $e_{22}^* + \delta e_{22}$ respectively compared to e_{11}^* and e_{22}^* .

The social dilemma is likely to be particularly acute when the community is large ($n \gg 1$) and social norms are only weakly responsive to individual women's efforts to change them ($\frac{\partial n_{ji}}{\partial a_i} \rightarrow 0$ for $i \neq j$). Such a situation approximates classic collective action problems studied in economics and political science (Olson, 1971). Here, the impact of an individual woman trying to persuade men to become less violent is negligible, even though significant reductions in violence could be obtained were all women in the community to contribute simultaneously. One woman investing huge effort to change men's attitudes would have little impact relative to many contributing small amounts each.

Scenario 1. We illustrate our results using examples. Interventions that make it easier for women to take preventive action—i.e., reduce the costs of effort or increase the effectiveness of effort—may seem like obvious solutions to inaction at equilibrium. However, Theorem 1 shows that such interventions do not eliminate the presence of social dilemmas as they do not change the relationship between individual and collective rationality. This limits their impact on violence and women's welfare: if individual benefits to action are almost zero relative to cost, it might require unrealistically large changes in model parameters to generate impact.

Let us model a scenario where the social dilemma of preventing violence is severe: social norms encouraging domestic violence are widespread and influential relative to women's ability to change men's attitudes to violence. We parameterise attitudes as a linear function of efforts $a_j(e_{j1}, \dots, e_{jn}) = 2 - \sum_j f_{ji} e_{ji}$ and set $f_{ji} = 0.1$ to indicate a weak ability to alter men's attitudes. We set the number of couples $n = 10$ to model a small community and $t_i = 0.5$ to indicate a degree of tolerance for violence among women. We set $s_i = c_i = n_{ji} = d_{ji} = 1$. As all couples have identical parameters, all women experience the same level of violence, exert the same level of effort and experience the same level of welfare; we can also consider any plotted values to be the values of a 'representative average woman'.

Figure 3 shows the effect of varying simulated parameters³ on equilibrium levels of violence, effort, and welfare. Interventions that reduce women's tolerance for experiencing violence or increase their effectiveness at changing men's attitudes have almost no discernible impact on any outcome. Interventions to reduce the cost of effort only begin to have material impact when the cost has been reduced to almost zero. Interventions to increase costs of perpetrating violence do have an impact, but this impact is independent of women's efforts to prevent violence, which stay near constant throughout. These results show the complexity of mobilising collective action to prevent domestic violence.

Scenario 2. Let us consider an alternative scenario. Corollary 1 indicates that increases in the cost of perpetrating violence have ambiguous effects. Counterintuitively, it is theoretically possible for more severe punishment for violence to exacerbate levels of violence. In our model, this effect arises from women themselves anticipating the deterrent effects of externally imposed punishment, which leads women to reduce effort in changing men's attitudes to violence, thus weakening or nullifying any possible deterrent effect of external punishment.

We can simulate this perverse effect with a slight modification to Scenario 1. As before, we parameterise attitudes as a linear function of efforts $a_j(e_{j1}, \dots, e_{jn}) = 2 - \sum_j f_{ji} e_{ji}$. As before, we set the effectiveness of attitude change to $f_{ji} = 0.1$, tolerance for violence to $t_i = 0.5$ and cost of preventive effort to $d_{ji} = 1$. However, we now increase the strength of norm-based reward and punishment

to $n_{ji} = 5$, the number of couples to $n = 20$, and the degree of suffering incurred by violence to $s = 1.5$. We set the cost of perpetrating violence to $c_i = 0.4$ and then look at the effect of gradually increasing it.

Figure 4 shows the result. We see that increases in the cost of violence initially result in greater levels of violence before slowly reducing in intensity after the cost of violence has more than doubled. In contexts with virtually no external sanction for violence, women's own efforts at attitude change are the primary force keeping violence at bay. Increasing costs of violence at a point where external sanction is still weak can lead women to overly relax their own preventive efforts, which results in the observed rise in violence.

Modelling solutions

Given the inherent challenges in motivating collective action to prevent violence in the face of social dilemmas, it is important to consider possible solutions. As discussed above, community mobilisation interventions have achieved varied success in reducing violence (Abramsky et al., 2014; Chatterji et al., 2020). Differences in impact reflect differences in context, implementation, and time allotted to the intervention (Gram et al., 2019). One indicator of quality of implementation may be the ability of such interventions to encourage community participation beyond baseline levels. Indeed, if no social dilemma were present, one would need to explain why community members had not already organised effective collective action prior to the introduction of an external intervention.

Conventional economic solutions to social dilemmas based on direct material incentives—tax credits, financial subsidies, property rights, or legal contracts (Oliver, 1993; Olson, 1971)—are largely infeasible. Financial rewards for laywomen to challenge patriarchal attitudes among their neighbours are unlikely to work if men know that women have adopted feminist values simply because they were paid to do so. Solutions based on property rights and legal contracts are also unworkable when the externalities in question—reductions in domestic violence due to changed social norms—are difficult-to-observe, intangible goods.

Instead, we might consider solutions based on altruism.

Process-based altruism. Process-based altruism refers to an intrinsic motivation to take action irrespective of any instrumental benefit (Kyriacou, 2010). This can be due to righteous anger, a sense of injustice, moral duty, or principle (Schuessler, 2000; Smith et al., 2012). It could also be due to a sense of empowerment or ability to lead on a moral issue. We can model process-based altruism by adding an intrinsic benefit term to women's utility Eq. (2). Let

$$U_{wi} = \sum_k b e_{ki} - s_i (1 - t_i) v_i - \frac{1}{2} \sum_k d_{ki} e_{ki}^2 \quad (5)$$

for constant $b > 0$. $\sum_k b e_{ki}$ only depends on women's effort, not their experience of violence. This avoids incentive incompatibilities caused by women's motivation for effort to prevent violence being conditional on such effort directly benefiting themselves and their own relationship to their partner. However, as motivation to take action is no longer tied to expectations about impact on violence, this mechanism might also theoretically result in over-investment in action against violence. Women might take action at great personal cost with little impact on violence, because 'doing something is better than nothing'. Finding the optimal value for b is nontrivial, but we might obtain insight by considering the following:

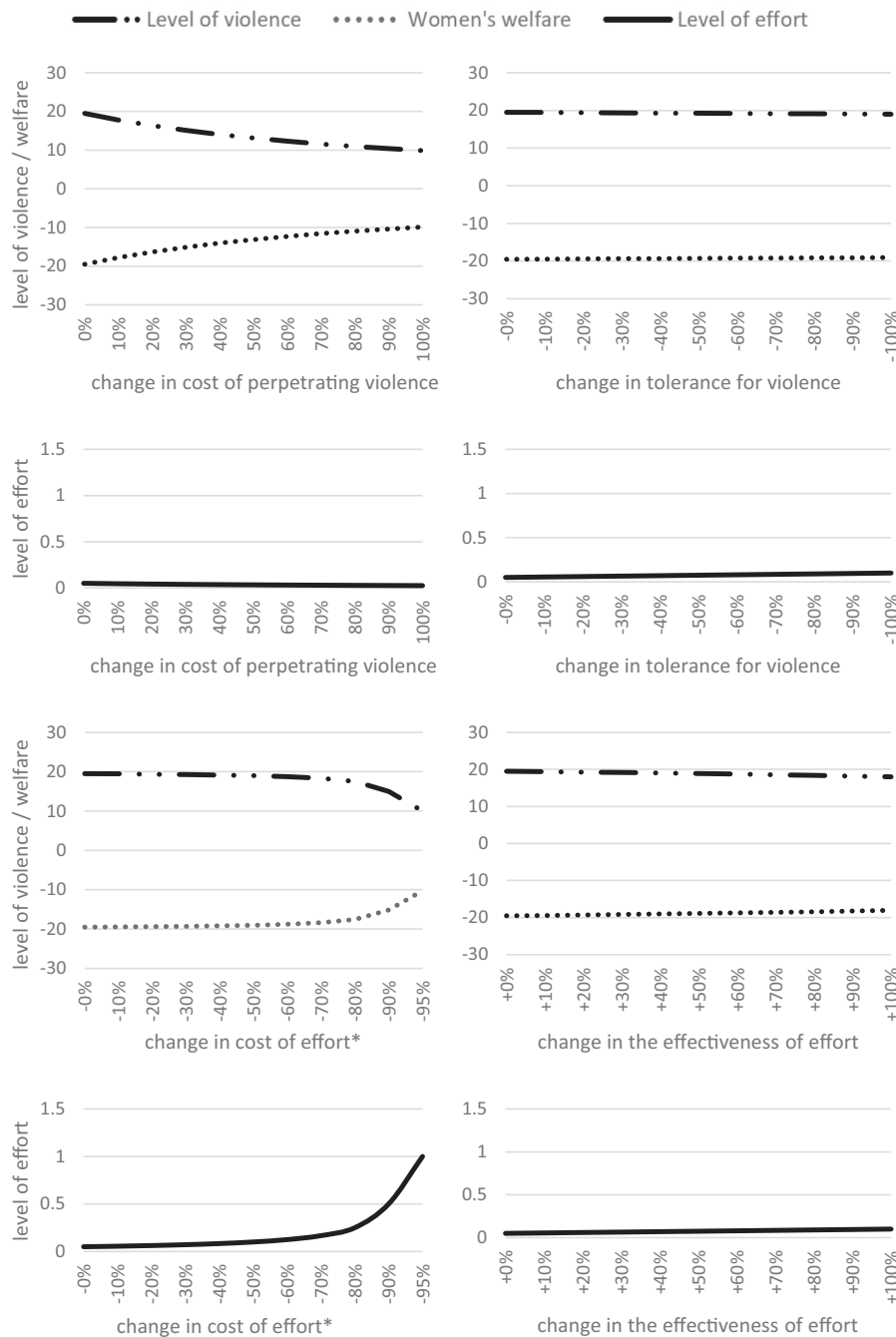


Fig. 3 Simulated impacts of changes to model parameters on equilibrium outcomes (Scenario 1). Only increased cost of violence perpetration materially improves women’s welfare for moderate-sized parameter changes. *Cost of effort must be strictly positive, so the last data point is –95% instead of –100%.

Proposition 2. *Women’s welfare is increasing in b if and only if for all i :*

$$b \sum_k d_{ki} \frac{\partial e_{ki}^*}{\partial b} < - \sum_k \sum_j \tau_{ji} \frac{s_i n_{jk}}{c_i} \frac{\partial a_k}{\partial e_{kj}} \frac{\partial e_{kj}^*}{\partial b} \tag{6}$$

where $\tau_{ji} = \begin{cases} t_i & \text{if } j = i \\ 1 & \text{if } j \neq i \end{cases}$

Proposition 2 states that we can always improve woman w_i ’s welfare by promoting greater intrinsic motivation to act among all women, as long as her marginal cost of effort is outweighed by additional benefit received from other women becoming more

motivated to act. This suggests b can be large without causing over-investment in the context of social dilemmas, where the marginal benefit of individual effort is negligible, but that of collective effort is substantial.

Empathetic altruism. We use ‘empathetic altruism’ to refer to a motivation to help people in need caused by feelings of concern, sympathy, compassion or identification with their pain (Penner et al., 2005). This differs from process-based altruism, which refers to a desire to act due to the perceived intrinsic goodness of such action irrespective of its consequences for oneself or others. We can model empathetic altruism by adding other women’s

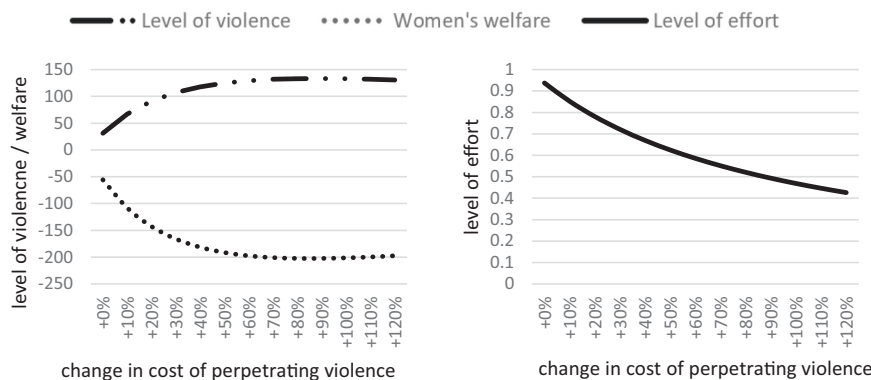


Fig. 4 Simulated impacts of increases in the cost of violence perpetration on equilibrium outcomes (Scenario 2). Moderate-sized increases in cost of perpetrating violence can result in *greater* levels of violence.

suffering to women’s utility function⁴:

$$U_{w_i} = -s_i(1 - t_i)v_i - \eta \sum_{j \neq i} s_j v_j - \frac{1}{2} \sum_k d_{ki} e_{ki}^2 \quad (7)$$

where $0 \leq \eta$ represents weight given to other women’s suffering. η is upwardly unbounded, so a woman might put larger weight on others’ suffering than on her own. Women can now account for the positive externalities of their own actions when deciding on effort levels. As with process-based altruism, this mechanism may increase effort levels toward optimality, but may also overshoot and compel women to expend costly effort without adequate compensation to themselves for their own involvement in collective action.

Proposition 3. *Women’s welfare is increasing in η if and only if for all i :*

$$-\eta \sum_k \sum_{j \neq i} \frac{s_j n_{jk}}{c_j} \frac{\partial a_k}{\partial e_{ki}} \frac{\partial e_{ki}^*}{\partial \eta} < - \sum_k \sum_j \tau_{ji} \frac{s_i n_{ik}}{c_i} \frac{\partial a_k}{\partial e_{kj}} \frac{\partial e_{kj}^*}{\partial \eta} \quad (8)$$

$$\text{where } \tau_{ji} = \begin{cases} t_i & \text{if } j = i \\ 1 & \text{if } j \neq i \end{cases}$$

Proposition 3 states that increasing η improves women’s welfare, as long as the degree of benefit produced by each woman for neighbouring women is outweighed the degree of benefit received by that woman from the actions of neighbouring women. This suggests that empathetic altruism works best in homogeneous populations where flows of benefit between women are more likely to be equal in size.

Reciprocal altruism. Reciprocal altruism⁵ refers to a motivation to provide favours for other individuals in anticipation of being able to receive future favours in return (Trivers, 2006). It differs from empathetic altruism in that individuals help others in proportion to the degree they expect others to help them in return, rather than the degree to which they expect others’ suffering to be alleviated by such help. It differs from process-based altruism in that individuals do take the consequences of their actions into account in deciding whether to act. We model reciprocal altruism by assuming women are motivated to match the effort levels of other women (Gächter, 2006):

$$U_{w_i} = -s_i(1 - t_i)v_i + \rho \sum_k \sum_{j \neq i} e_{ki} e_{kj} + \frac{1}{2} \rho \sum_k e_{ki}^2 - \frac{1}{2} \sum_k d_{ki} e_{ki}^2 \quad (9)$$

$0 \leq \rho$ represents weight given to matching other women’s efforts with higher values indicating greater willingness to match⁶. We can show:

Proposition 4. *Suppose $0 \leq \rho < \frac{1}{n}$. Women’s welfare increases in ρ if and only if for all i :*

$$-\frac{\rho}{1 - n\rho} \sum_k \sum_j \frac{s_j(1 - t_j)n_{jk}}{c_j} \frac{\partial a_k}{\partial e_{kj}} \frac{\partial e_{ki}^*}{\partial \rho} < - \sum_k \sum_j \tau_{ji} \frac{s_i n_{ik}}{c_i} \frac{\partial a_k}{\partial e_{kj}} \frac{\partial e_{kj}^*}{\partial \rho} \quad (10)$$

where $\tau_{ji} = \begin{cases} t_i & \text{if } j = i \\ 1 & \text{if } j \neq i \end{cases}$. For $\rho \geq \frac{1}{n}$, a subgame perfect equilibrium may not exist.

Proposition 4 shows that the marginal cost of increasing ρ may rise rapidly, as women’s commitment to match each other’s investments in collective action creates mutually reinforcing feedback loops: $\frac{\rho}{1 - n\rho} \rightarrow \infty$ as $\rho \rightarrow \frac{1}{n}$ from the left. This suggests that even small amounts of reciprocity can entail large commitments in effort.

Scenario 3. Let us now consider the impact of increasing altruism on women’s outcomes. For simplicity, we use the same functional forms and initial values that we used in Scenario 1. Recall that all couples have identical parameters, so all women experience the same level of violence, exert the same level of effort and experience the same level of welfare, i.e., we can consider values plotted to concern a representative ‘average woman’.

Figure 5 shows the simulated effect of increasing process-based, empathetic and reciprocal altruism on equilibrium levels of violence, effort to prevent violence, and women’s welfare. In contrast to Scenario 1, we find an immediate increase in women’s effort levels and a concomitant decrease in levels of violence in response to even small amounts of altruism. For example, an increase in average levels of effort from zero to one unit results in a halving of experienced violence. If a woman wanted to achieve comparable improvements to her own relationship through her own efforts alone—i.e., by investing all her effort in changing her own male partner’s attitudes without involving other women—it would require nearly a hundred-fold greater level of effort.

We found little difference in our simulated scenario between the effects of process-based and empathetic altruism. However, reciprocal altruism had little effect on effort and violence levels until the degree of reciprocity neared $\frac{1}{n}$, when levels of effort changed rapidly, as predicted by Proposition 4.

Scenario 4. Let us now consider the effect of introducing heterogeneity between couples. We begin with the ‘representative woman’ model in Scenario 1 and then vary the parameter values for each couple by an independent random amount ranging from a threefold reduction to a threefold increase (see Supplementary Appendix for further details). We next gradually vary levels of

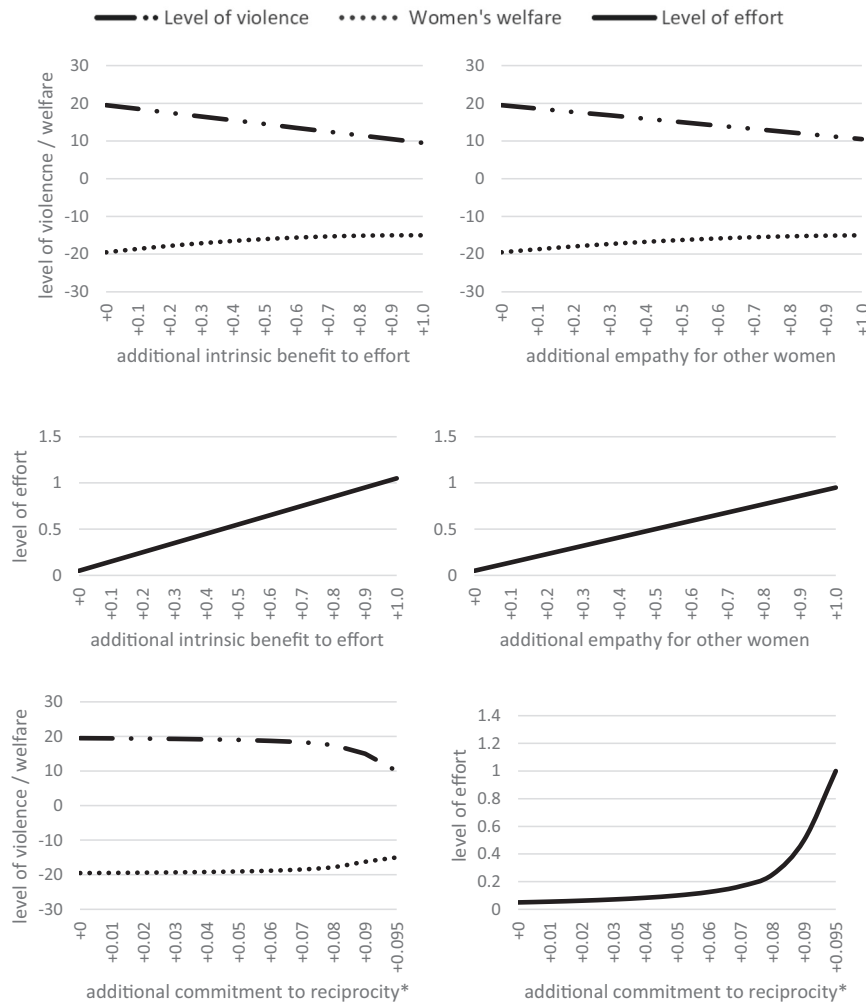


Fig. 5 Simulated impacts of changes to women's level of altruism on equilibrium outcomes (Scenario 3). Increases in intrinsic benefit, empathy and reciprocity reduce levels of violence and increase women's welfare. *No equilibrium exists for reciprocity level $\rho = 0.1$, so the last data point is +0.095, not +0.1.

altruism and inspect their effect on violence-related outcomes. The minimum, median, and maximum change in such outcomes in response to changes in levels of altruism is plotted in Fig. 6. We have not displayed the effect of changing women's reciprocal altruism, as it offers essentially the same lessons as the other two forms of altruism.

The median changes in violence-related outcomes mirror results from Scenario 3. The median drop in violence after increasing intrinsic benefit by +1.0 is still approximately ten units. The median change in welfare is still approximately five units. The median increase in effort is the same as before⁷. However, there is considerable variation between individual couples. Long before the welfare of the median woman peaks, the welfare of the minimum woman peaks. If all women experienced +1.0 units greater process-based altruism, the median and maximum woman would experience improved welfare, but the minimum woman would be worse off compared to the situation where no altruism existed. The results for empathetic altruism are similar.

Implications

Community mobilisation researchers have long asserted that coordinated, collective action is necessary to prevent domestic violence, but have not fully explained why this is the case (Gram

et al., 2019). One of the main benefits of explicit mathematical models is an enhanced ability to interrogate the assumptions behind our theories. We show that, under Assumptions I–VI, women's collective action to prevent domestic violence becomes a social dilemma, which requires coordinated action to overcome. Should any one of the assumptions fail to hold, social dilemmas may not exist and interventions to address domestic violence may not need to involve collective action at all. Should they hold, community mobilisation researchers might need to re-interrogate existing ideas of how to best encourage collective action.

Not all contexts call for collective action to prevent violence.

Assumptions I and II in our model require the existence of social norms that impose sanctions or rewards on perpetrators of violence. Without these assumptions, our model does not predict externalities or social dilemmas to arise. Domestic violence-related social norms have been documented in contexts in which community members live in close geographic and social proximity such as college campuses in the United States (DeKeseredy and Schwartz, 2013), urban slums in India (Verma et al., 2006) and Uganda (Kyegombe et al., 2014), social clubs in Tanzania (Mulawa et al., 2018), or rural communities in Kenya (Hatcher et al., 2013). For example, residents of Mumbai slums generally keep their door open for ventilation and socialisation given the

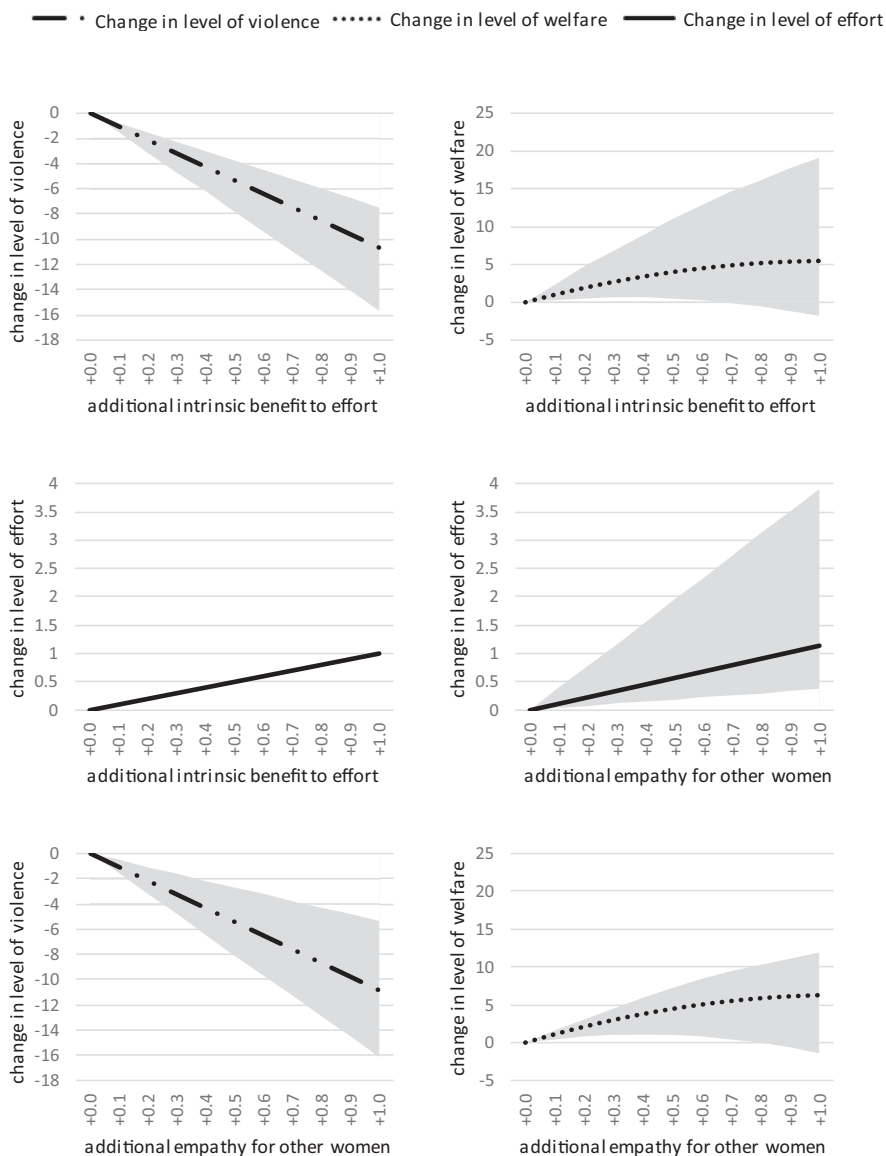


Fig. 6 Simulated impacts of changes to women's altruism in the presence of heterogeneity between individual couples (Scenario 4). Heterogeneity in model parameters for individual couples create heterogeneity in outcomes. Lines represent median changes among all women. Grey bands indicate the full range of changes among all women.

small amount of space available to them inside (Gram et al., 2020). Under such conditions, residents generally felt they would immediately notice if anyone closed their door and perpetrated domestic violence. Studies of violence prevention programmes in Kenya (Haushofer et al., 2019) and Mexico (Bobonis et al., 2015) found reductions in domestic violence within one household was accompanied by decreases in violence among neighbouring households.

However, contexts in which such social norms are weak or non-existent do not obviously present risks of social dilemmas. For example, rural US households have been argued to be socially isolated due to large physical distances between homes and a culture of silence surrounding household affairs (Anderson et al., 2014). A social network analysis of rural households in Honduras found attitudes to violence were almost exclusively determined by household members rather than community members (Shakya et al., 2016). Men and women in Beijing, Seoul, and Hong Kong have been found to keep domestic violence secret from their closest friends and relatives for fear of losing face (Bouhours and

Broadhurst, 2015; Emery and Wu, 2020). In these contexts, it may be challenging for men to reward or punish other men for perpetrating domestic violence or enforcing gender normative behaviour in their partners, as they do not have access to such information.⁸

If social dilemmas do not exist, it is unclear to what extent implementers are justified in pushing for higher levels of collective action beyond individually rational levels, as women living in such contexts do not face significant returns to coordinated action. Extensive participation might take time and energy away from activities that women themselves genuinely value without proportionate reductions in their experience of violence. Collective action-oriented community mobilisation might be inefficient in such cases relative to individual interventions such as couples counselling (Stern and Niyibizi, 2018) or economic empowerment (Green et al., 2015).

Conventional solutions to prevent violence may not work. Surprisingly, our model suggests that a number of conventional

techniques for preventing violence in community mobilisation interventions may not always work. Theorem 1 showed that changes to women's tolerance for experience violence do not eliminate the presence of social dilemmas. Scenario 1 showed that lower tolerance for violence may matter to neither women's decision to take action nor men's decisions to perpetrate violence. Intervention designers should not automatically expect consciousness-raising efforts encouraging women to reject violence (Watts and Hipolito-Delgado, 2015) to lead to collective action.

Similarly, Scenario 1 showed that changes to the cost of women's preventive action or the effectiveness with which they changed men's attitudes might not materially alter women's level of participation in collective action or men's level of perpetration of violence. Theorem 1 showed that such changes do not remove the presence of social dilemmas. Interventions that make it easier, less risky or less time-consuming for women to take action against domestic violence should not always be expected to stimulate significant collective action.

Scenario 2 even showed that increasing the cost of violence perpetration can backfire and cause greater levels of violence, if women anticipate its deterrent effect and invest less effort in prevention as a result. This mirrors models of criminal punishment, which predict null effects of harsher penalties on crime because police officers anticipate their deterrent effect and relax law enforcement efforts as a result (Tsebelis, 1990). Behavioural evidence partially bears this out by showing that higher punishments do indeed lead to relaxed efforts to control crime in lab experimental settings (Rauhut, 2009, 2015). However, caution is needed before extrapolating from these findings, as no field experiments have been done.

We proposed solutions based on process-based, empathetic, and reciprocal altruism, with roots in the literature on social movements and collective action. Scenarios 3 and 4 showed that such mechanisms could stimulate action and reduce violence. Studies of feminist social movements have documented the importance of emotions in activating process-based altruism towards other women, particularly righteous anger arising from awareness of patriarchal oppression (Hercus, 1999). Collective identity has been found to foster empathetic altruism, as women's understanding of themselves as a social group makes them feel that 'violence against one woman constitutes automatically also a threat against... others' (Kreft, 2019). Researchers studying sexual assault in the US have evidenced the role of empathy in promoting bystander action (Beeble et al., 2008; Yule et al., 2020). Finally, reciprocal altruism has been observed in contexts where women already had reciprocal economic ties, such as microcredit groups in which women mobilised to support other group members against domestic violence (Sanyal, 2009).

Valuing the benefits of community participation is nontrivial.

Cost-effectiveness analyses of community mobilisation interventions to prevent domestic violence have hitherto valued the cost of time for community volunteers as a linear function of local wage rates, while equating benefits with the total number of cases averted by the whole intervention (Jan et al., 2011; Michaels-Igbokwe et al., 2016). A counterfactual analysis of the causal impact on domestic violence of social norms-based interventions in Uganda explicitly avoided accounting for non-linear returns to action (Kadengye et al., 2019). However, simply dividing the total number of cases averted by the total time spent on intervention activities provides an incomplete picture of the relationship between benefits and costs, when there are externalities to time spent on intervention activities. Time spent by women engaging with a community mobilisation intervention can produce benefits, not only for themselves, but also for other women in the

community through social norm change. Our analysis highlights the need to identify and measure externalities when costing community mobilisation interventions (Brouwer, 2019).

Researchers have justified lack of collective action in community mobilisation interventions with reference to women's verbal statements about lacking time and energy to invest (Hargreaves et al., 2009). Bystander action researchers have attributed inaction in the face of violence to fears of getting hurt, causing embarrassment, or being targeted by gossip for intervening in "others' business" (Banyard, 2011). Community members may also believe that getting involved achieves little or even harms survivors of violence (Latta and Goodman, 2011). Yet, if women are stakeholders in a social dilemma, it is simultaneously true that (1) women are individually justified in abstaining from action, because it is not worth the time, costs and risks involved and (2) women are denying themselves benefits that more than outweigh the time, costs and risks involved by collectively abstaining. This moral aspect of social dilemmas has been debated in research on environmental (Hourdequin, 2010) and feminist (Slote, 2007) ethics, but not in community mobilisation.

Discussion

By modelling collective action to address domestic violence, we walk the less trodden path. Previous mathematical models of domestic violence have focused on household-level determinants of violent behaviour, modelling domestic violence as a function of intra-household bargaining power (Bloch and Rao, 2002; Eswaran and Malhotra, 2011; Farmer and Tiefenthaler, 1997; Haushofer et al., 2019; Tauchen et al., 1991). In such models, men perpetrate violence due to intrinsic preferences for violence or for the sake of extracting resources from their female partners. Women can only prevent domestic violence by leaving the relationship, strengthening their bargaining power or acquiescing to their male partners' demands for resources. Other models have examined intergenerational impacts on domestic violence of witnessing violence between parents as a child (Koç and Erkin, 2012; Pollak, 2004) or modelled women's choice to report violent partners to the police (Aizer and Dal Bo, 2009). None of these modelled women's collective efforts to change social norms.

Like all models, our model relied on assumptions for deriving predictions and explanations of behaviour. Most fundamentally, we modelled individual decision-making as a process of rational utility-maximisation. We believe this assumption is more realistic than models in which individuals act reflexively without considering costs or benefits of action (Drigo et al., 2012; Rigby and Johnson, 2017). Our assumption is consistent with evidence on social mobilisation (Rogers et al., 2018), prosocial (Bierhoff, 2005), and help-seeking (Fugate et al., 2005) behaviour, which demonstrates that individuals are sensitive to costs and benefits of such behaviour. In contrast to strict rational choice models, we included significant social and psychological components. In particular, we modelled men's preferences for violence as being amenable to change through the sustained efforts of women in their community. This aligns with recent assessments of the evidence base in economics calling for a move towards a view of individuals as 'enculturated actors' whose preferences are subject to social influences (Hoff and Stiglitz, 2016).

Nonetheless, our model has limitations. First, we only modelled women's participation in collective action. Men's participation in action to prevent violence against women does not straightforwardly entail the presence of social dilemmas as they do not directly benefit from reduced violence; indeed, active male engagement in violence prevention has been weaker compared to female engagement in programmes involving both genders⁹ (Chakraborty et al., 2018; Hatcher et al., 2020; Jejeebhoy and

Santhya, 2018). Second, our model focused on the interaction between attitudes, social norms, and behaviour, but did not explicitly model beliefs. Models of efficacy beliefs (Lemoine, 2019) can assess the impact of women under- or over-estimating their actual influence over men's attitudes. Women may for example refrain from action because they lack confidence in their own strength. Models of 'pluralistic ignorance' (Centola et al., 2005), in which individuals over- or under-estimate support for a social norm, may be needed if such misperceptions prevent collective action. We did not incorporate belief updating to avoid adding complexity to an already-complex model.

Conclusion

We integrated widely known notions of patriarchal gender norms, male peer support, and hegemonic masculinity from the literature on violence against women into a mathematical model of social change through community mobilisation. We showed that, under a broad range of conditions, women's decision to participate in collective action to prevent domestic violence becomes a social dilemma. In such a situation, our model predicts process-based, empathetic, and reciprocal altruism to work in mobilising collective action and reducing levels of domestic violence. Our model provides a basis for developing of theories of violence prevention that employ mathematical modelling to explore the complex system dynamics of collective action. Future research might look for predicted relationships between altruism and collective action. Such work would benefit from collaborative work with game theorists, gender scholars, and public health researchers. Interdisciplinary studies combining a solid theoretical basis with empirical investigation should trial such solutions in an effort to improve violence prevention work.

Data availability

All data generated or analysed during this study are included in this published article.

Received: 23 September 2019; Accepted: 5 February 2021;

Published online: 01 March 2021

Notes

- Note, we are not claiming that our model captures the full complexity of preventing domestic violence through community mobilisation. Rather, we are putting a magnifying glass to social interactions between community members in order to analyse the contribution of grassroots collective action to the prevention of domestic violence in the context of a community mobilisation intervention.
- For simplicity, we describe the agents in our model as women and men living together as couples, for example as husband and wife in a nuclear family. We can fit our model to other contexts by re-interpreting w_i and h_i . For example, in an Indian context where younger married women live in households with their husband and his parents, in-laws may well be instigators or perpetrators of violence themselves (Bentley, 2018). In such a context, we can re-interpret w_i to refer to young, married women and h_i to refer to their husband and in-laws jointly.
- Note, we are purely choosing parameter values in example sections to facilitate presentation of our argument: they are not derived from real-world empirical data.
- Our model of empathetic altruism here matches economic conceptions of 'pure altruism' (Andreoni, 1990). 'Impure altruism' or 'warm glow giving' where the altruistic behaviour is motivated by intrinsic psychological benefits of warmth and fellow-feeling is better modelled as a form of process-based altruism (Andreoni, 1990).
- Our model of reciprocal altruism only refers to direct reciprocity. Indirect reciprocity, where participants contribute to the collective good as a costly signal of integrity and trustworthiness, has been modelled as an exogenous, expected benefit to prosocial behaviour (Heinz and Schumacher, 2017). In this case, we can re-use our model for intrinsic benefits by re-interpreting b to refer to expected future benefit from improving one's reputation.
- For presentation purposes, we assume women also 'match' their own efforts at rate ρ using the term $\frac{1}{2}\rho \sum_k e_{ki}^2$. The equilibria do not change if we omit this, but the relevant formulae become more difficult to interpret.

- Our model of process-based altruism fixes a direct relationship between intrinsic benefit and effort, so there is no variation in this respect in Fig. 6.
- Nonetheless, rural US communities have also been described as lacking in anonymity and confidentiality and even perpetuating violence against women through fraternal 'boy networks that shield perpetrators from punishment (Anderson et al., 2014). 'Face culture' has also been argued to protect against sexual aggression in contexts where men expect such behaviour to become publicly known and thus cause them shame (Hall et al., 2005). It may be difficult to make general statements about cultures where social norms concerning domestic violence do or do not exist.
- Complex ethical concerns surround programmes with a focus on men's collective action to prevent violence against women, including risks of shifting funding away from programmes that support survivors of violence, delegitimising women-only or women-focused programmes, or marginalising women's voices by placing men in leadership positions (Macomber, 2018). Masculinity researchers argue these concerns can be addressed in men's engagement programmes (Flood, 2015), but emphasise the need to dispel ideas that men necessarily benefit from dismantling patriarchy, that men are best placed to talk to other men about domestic violence, and that men fitting the masculine ideal wield most influence over other men's attitudes to gender equality.

References

- Abimbola S (2019) Beyond positive a priori bias: reframing community engagement in LMICs. *Health Promot Int* 35(3):598–609
- Abramsky T, Devries K, Kiss L, Nakuti J, Kyegombe N, Starmann E, Cundill B, Francisco L, Kaye D, Musuya T (2014) Findings from the SASA! Study: a cluster randomized controlled trial to assess the impact of a community mobilization intervention to prevent violence against women and reduce HIV risk in Kampala, Uganda. *BMC Med* 12(1):122
- Abramsky T, Devries KM, Michau L, Nakuti J, Musuya T, Kiss L, Kyegombe N, Watts C (2016) Ecological pathways to prevention: how does the SASA! community mobilisation model work to prevent physical intimate partner violence against women? *BMC Public Health* 16(1):1–21
- Aizer A, Dal Bo P (2009) Love, hate and murder: commitment devices in violent relationships. *Journal Public Econ* 93(3–4):412–428
- Anderson KM, Renner LM, Bloom TS (2014) Rural women's strategic responses to intimate partner violence. *Health Care Women Int* 35(4):423–441
- Anderson R (2008) New MRC guidance on evaluating complex interventions. *Br Med J*, 337. <https://doi.org/10.1136/bmj.a1937>
- Andreoni J (1990) Impure altruism and donations to public goods: a theory of warm-glow giving. *Econ J* 100(401):464–477
- Bandiera O, Buehren N, Burgess R, Goldstein M, Gulesci S, Rasul I, Sulaiman M (2016) Empowering adolescent girls: evidence from a randomized control trial in Uganda. London School of Economics. Retrieved 12/1/2016, from <http://sticerd.lse.ac.uk/dps/eopp/eopp50.pdf>
- Banyard VL (2011) Who will help prevent sexual violence: creating an ecological model of bystander intervention. *Psychol Violence* 1(3):216
- Bauch CT, Earn DJ (2004) Vaccination and the theory of games. *Proc Natl Acad Sci* 101(36):13391–13394
- Beeble ML, Post LA, Bybee D, Sullivan CM (2008) Factors related to willingness to help survivors of intimate partner violence. *J Interpers. Violence* 23(12):1713–1729
- Bentley A (2018) The role of in-laws as perpetrators of violence against women in Mumbai. *Int J Soc Sci Interdiscip Stud* 3(1):4
- Bierhoff HW (2005) Prosocial behaviour. Psychology Press
- Bloch F, Rao V (2002) Terror as a bargaining instrument: a case study of dowry violence in rural India. *Am Econ Rev* 92(4):1029–1043
- Bobonis GJ, Perez Castro R, Morales JS (2015) Conditional cash transfers for women and spousal violence: evidence of the long-term relationship from the Oportunidades program in Rural Mexico. <https://www.econstor.eu/bitstream/10419/146440/1/IDB-WP-632.pdf>
- Bouhours B, Broadhurst R (2015) Violence against women in Hong Kong: results of the international violence against women survey. *Violence Against Women* 21(11):1311–1329
- Bourey C, Williams W, Bernstein EE, Stephenson R (2015) Systematic review of structural interventions for intimate partner violence in low-and middle-income countries: organizing evidence for prevention. *BMC Public Health* 15(1):1
- Brouwer WB (2019) The inclusion of spillover effects in economic evaluations: not an optional extra. *Pharmacoeconomics* 37(4):451–456
- Busby C (1999) Agency, power and personhood: discourses of gender and violence in a fishing community in South India. *Crit Anthropol* 19(3):227–248
- Centola D, Willer R, Macy M (2005) The emperor's dilemma: a computational model of self-enforcing norms. *Am J Sociol* 110(4):1009–1040
- Chakraborty P, Osrin D, Daruwalla N (2018) "We learn how to become good men" working with male allies to prevent violence against women and girls in urban informal settlements in Mumbai, India. *Men Masc* 23(3–4):749–771
- Chatterji S, Stern E, Dunkle K, Heise L (2020) Community activism as a strategy to reduce intimate partner violence (IPV) in rural Rwanda: results of a community randomised trial. *J Global Health* 10(1):1–17

- Cislaghi B, Denny EK, Cissé M, Gueye P, Shrestha B, Shrestha PN, Ferguson G, Hughes C, Clark CJ (2019) Changing social norms: the importance of “organized diffusion” for scaling up community health promotion and women empowerment interventions. *Prev Sci* 20(6):936–946
- Clark CJ, Ferguson G, Shrestha B, Shrestha PN, Oakes JM, Gupta J, McGhee S, Cheong YF, Yount KM (2018) Social norms and women’s risk of intimate partner violence in Nepal. *Soc Sci Med* 202:162–169
- Clark CJ, Shrestha B, Ferguson G, Shrestha PN, Calvert C, Gupta J, Batayah B, Bergenfeld I, Oakes JM (2020) Impact of the change starts at home trial on Women’s experience of intimate partner violence in Nepal. *SSM-Popul Health* 10:100530
- Colman AM, Krockow EM, Chattoe-Brown E, Tarrant C (2019) Medical prescribing and antibiotic resistance: a game-theoretic analysis of a potentially catastrophic social dilemma. *PLoS ONE* 14(4):e0215480
- Connell RW (1987) *Gender and power: society, the person and sexual politics*. Stanford University Press
- Cooper J (2018) *State capacity and gender inequality: experimental evidence from Papua new Guinea*. Columbia University, New York, NY
- Cornes R, Sandler, T (1996) *The theory of externalities, public goods, and club goods*. Cambridge University Press
- Daruwalla N, Hate K, Pinto P, Ambavkar G, Kakad B, Osrin D (2017) You can’t burn the house down because of one bedbug: a qualitative study of changing gender norms in the prevention of violence against women and girls in an urban informal settlement in India. *Wellcome Open Res*, 2(48). <https://doi.org/10.12688/wellcomeopenres.11805.2>
- Daruwalla N, Jaswal S, Fernandes P, Pinto P, Hate K, Ambavkar G, Kakad B, Gram L, Osrin D (2019) A theory of change for community interventions to prevent domestic violence against women and girls in Mumbai, India. *Wellcome Open Res* 4:1–32
- Davey C, Hargreaves J, Hassan S, Cartwright N, Humphreys M, Masset E, Prost A, Gough D, Oliver S, Bonell C (2018) Designing evaluations to provide evidence to inform action in new settings. *CEDIL Inception Paper*, Issue. <https://cedilprogramme.org/wp-content/uploads/2018/10/Designing-evaluations-to-provide-evidence.pdf>
- DeKeseredy WS, Schwartz MD (2013) *Male peer support and violence against women: the history and verification of a theory*. Northeastern University Press
- Devries KM, Mak JY, Garcia-Moreno C, Petzold M, Child JC, Falder G, Lim S, Bacchus LJ, Engell RE, Rosenfeld L (2013) The global prevalence of intimate partner violence against women. *Science* 340(6140):1527–1528
- Dillon G, Hussain R, Loxton D, Rahman S (2013) Mental and physical health and intimate partner violence against women: a review of the literature. *Int J Fam Med* 1–15
- Drigo M, Ehlschlaeger CR, Sweet EL (2012) Modeling intimate partner violence and support systems. *Ecologist-developed spatially-explicit dynamic landscape models*. Springer. pp. 235–253
- Dworkin SL, Hatcher AM, Colvin C, Peacock D (2013) Impact of a gender-transformative HIV and antiviolence program on gender ideologies and masculinities in two rural, South African communities. *Men Mas* 16(2):181–202
- Emery CR, Wu S (2020) A tale of two Confucian capitals: the role of friends and secrecy in Beijing and Seoul. *Violence Against Women* 26(5):458–481
- Epstein JM (2008) Why model? *J Artif Soc Soc Simul* 11(4):12
- Eswaran M, Malhotra N (2011) Domestic violence and women’s autonomy in developing countries: theory and evidence. *Can J Econ* 44(4):1222–1263
- Fabiano PM, Perkins HW, Berkowitz A, Linkenbach J, Stark C (2003) Engaging men as social justice allies in ending violence against women: evidence for a social norms approach. *J Am Coll Health* 52(3):105–112
- Farmer A, Tiefenthaler J (1997) An economic analysis of domestic violence. *Rev Soc Econ* 55(3):337–358
- Ferguson W (2013) *Collective action and exchange: a game-theoretic approach to contemporary political economy*. Stanford University Press
- Flood M (2015) Work with men to end violence against women: a critical stocktake. *Cult Health Sex* 17(sup2):159–176
- Flood M, Pease B (2009) Factors influencing attitudes to violence against women. *Trauma Violence Abuse* 10(2):125–142
- Fugate M, Landis L, Riordan K, Naureckas S, Engel B (2005) Barriers to domestic violence help seeking: implications for intervention. *Violence Against Women* 11(3):290–310
- Fulu E, Jewkes R, Roselli T, Garcia-Moreno C (2013) Prevalence of and factors associated with male perpetration of intimate partner violence: findings from the UN Multi-country Cross-sectional Study on Men and Violence in Asia and the Pacific. *Lancet Global Health* 1(4):e187–e207
- Gächter S (2006) Conditional cooperation: behavioral regularities from the lab and the field and their policy implications. <https://www.econstor.eu/bitstream/10419/67977/1/512169225.pdf>
- García-Moreno C, Jansen HA, Ellsberg M, Heise L, Watts C (2005) WHO multi-country study on women’s health and domestic violence against women: initial results on prevalence, health outcomes and women’s responses. *WHO*
- García-Moreno C, Watts C (2011) Violence against women: an urgent public health priority. *Bull World Health Organ* 89(1):2–2. 2011
- Gibbs A, Corboz J, Shafiq M, Marofi F, Mecagni A, Mann C, Karim F, Chirwa E, Maxwell-Jones C, Jewkes R (2018) An individually randomized controlled trial to determine the effectiveness of the Women for Women International Programme in reducing intimate partner violence and strengthening livelihoods amongst women in Afghanistan: trial design, methods and baseline findings. *BMC Public Health* 18(1):164
- Gibbs A, Jewkes R, Sikweyiya Y, Willan S (2015) Reconstructing masculinity? A qualitative evaluation of the Stepping Stones and Creating Futures interventions in urban informal settlements in South Africa. *Cult Health Sex* 17(2):208–222
- Go VF, Johnson SC, Bentley ME, Sivaram S, Srikrishnan AK, Celentano DD, Solomon S (2003) Crossing the threshold: engendered definitions of socially acceptable domestic violence in Chennai, India. *Cult Health Sex* 5(5):393–408
- Goodman LA, Smyth KF (2011) A call for a social network-oriented approach to services for survivors of intimate partner violence. *Psychol Violence* 1(2):79–91
- Gram L, Chakraborty P, Nayreen D, Osrin, D (2020) Social and psychological readiness to take collective action against violence against women: a mixed methods study of informal settlements in Mumbai, India. *Violence Against Women*. 1–21. <https://journals.sagepub.com/doi/full/10.1177/1077801220971360>
- Gram L, Daruwalla N, Osrin D (2019) Understanding participation dilemmas in community mobilisation: can collective action theory help? *J Epidemiol Community Health* 73(1):90–96
- Gram L, Desai S, Prost A (2020) Classroom, club or collective? Three types of community-based group intervention and why they matter for health. *BMJ Global Health* 5(12):e003302
- Gram L, Fitchett A, Ashraf A, Daruwalla N, Osrin D (2019) Promoting women and children’s health through community-based groups in low- and middle-income countries: a mixed methods systematic review of mechanisms, enablers and barriers. *BMJ Global Health* 4:e001972
- Green EP, Blattman C, Jamison J, Annan J (2015) Women’s entrepreneurship and intimate partner violence: a cluster randomized trial of microenterprise assistance and partner participation in post-conflict Uganda (SSM-D-14-01580R1). *Soc Sci Med* 133:177–188
- Hall GCN, Teten AL, DeGarmo DS, Sue S, Stephens KA (2005) Ethnicity, culture, and sexual aggression: risk and protective factors. *J Consult Clin Psychol* 73(5):830
- Hargreaves J, Hatcher A, Strange V, Phetla G, Busza J, Kim J, Watts C, Morison L, Porter J, Pronyk P (2009) Process evaluation of the Intervention with Microfinance for AIDS and Gender Equity (IMAGE) in rural South Africa. *Health Educ Res* 25(1):27–40
- Hatcher AM, McBride R-S, Rebombo D, Munshi S, Khumalo M, Christofides N (2020) Process evaluation of a community mobilization intervention for preventing men’s partner violence use in peri-urban South Africa. *Eval Program Plann* 78:101727
- Hatcher AM, Romito P, Otero M, Bukusi EA, Onono M, Turan JM (2013) Social context and drivers of intimate partner violence in rural Kenya: implications for the health of pregnant women. *Cult Health Sex* 15(4):404–419
- Haushofer J, Ringdal C, Shapiro JP, Wang XY (2019) Income changes and intimate partner violence: evidence from unconditional cash transfers in Kenya. https://www.nber.org/system/files/working_papers/w25627/w25627.pdf
- Hechter M, Opp KD (eds) (2001) *Social norms*. Russell Sage Foundation
- Heckman JJ (2005) The scientific model of causality. *Soc Methodol* 35(1):1–97
- Heinz M, Schumacher H (2017) Signaling cooperation. *Eur Econ Rev* 98:199–216
- Heise LL (1998) Violence against women an integrated, ecological framework. *Violence Against Women* 4(3):262–290
- Heise LL, Kotsadam A (2015) Cross-national and multilevel correlates of partner violence: an analysis of data from population-based surveys. *The Lancet Global Health* 3(6):e332–e340
- Hercus C (1999) Identity, emotion, and feminist collective action. *Gender Soc* 13(1):34–55
- Hill A, Pallitto C, McCleary-Sills J, Garcia-Moreno C (2016) A systematic review and meta-analysis of intimate partner violence during pregnancy and selected birth outcomes. *Int J Gynecol Obstet* 133(3):269–276
- Hoff K, Stiglitz JE (2016) Striving for balance in economics: towards a theory of the social determination of behavior. *J Econ Behav Organ* 126:25–57. <https://doi.org/10.1016/j.jebo.2016.01.005>
- Hossain M, Zimmerman C, Kiss L, Abramsky T, Kone D, Bakayoko-Topolska M, Annan J, Lehmann H, Watts C (2014) Working with men to prevent intimate partner violence in a conflict-affected setting: a pilot cluster randomized controlled trial in rural Côte d’Ivoire. *BMC Public Health* 14(1):339
- Hourdequin M (2010) Climate, collective action and individual ethical obligations. *Environ Values* 19(4):443–464
- Jan S, Ferrari G, Watts CH, Hargreaves JR, Kim JC, Phetla G, Morison LA, Porter JD, Barnett T, Pronyk PM (2011) Economic evaluation of a combined microfinance and gender training intervention for the prevention of intimate partner violence in rural South Africa. *Health Policy Plann* 26(5):366–372

- Jejeebhoy SJ, Santhya K (2018) Preventing violence against women and girls in Bihar: challenges for implementation and evaluation. *Reprod Health Matters* 26(52):92–108
- Jewkes R, Levin J, Penn-Kekana L (2002) Risk factors for domestic violence: findings from a South African cross-sectional study. *Social Sci Med* 55(9):1603–1617
- Kadengye DT, Iddi S, Hunter L, McCoy SI (2019) Effectiveness of potential interventions to change gendered social norms on prevalence of intimate partner violence in Uganda: a causal inference approach. *Prev Sci* 20(7):1043–1053
- Kilmartin C, Smith T, Green A, Heinzen H, Kuchler M, Kolar D (2008) A real time social norms intervention to reduce male sexism. *Sex Roles* 59(3-4):264–273
- Kim JC, Watts CH, Hargreaves JR, Ndhlovu LX, Phetla G, Morison LA, Busza J, Porter JD, Pronyk P (2007) Understanding the impact of a microfinance-based intervention on women's empowerment and the reduction of intimate partner violence in South Africa. *Am J Public Health* 97(10):1794–1802
- Koç SA, Erkin HC (2012) On intergenerational transmission of domestic violence: pollak revisited. *Bogazici J* 26(1):45–65
- Krause KH, Gordon-Roberts R, VanderEnde K, Schuler SR, Yount KM (2016) Why do women justify violence against wives more often than do men in Vietnam? *J Interpers Violence* 31(19):3150–3173
- Kreft A-K (2019) Responding to sexual violence: women's mobilization in war. *J Peace Res* 56(2):220–233
- Kyegombe N, Starmann E, Devries KM, Michau L, Nakuti J, Musuya T, Watts C, Heise L (2014) 'SASA! is the medicine that treats violence'. Qualitative findings on how a community mobilisation intervention to prevent violence against women created change in Kampala, Uganda. *Global Health Action* 7(1):25082
- Kyriacou AP (2010) Intrinsic motivation and the logic of collective action: the impact of selective incentives. *American J Econ Sociol* 69(2):823–839
- Latta RE, Goodman LA (2011) Intervening in partner violence against women: a grounded theory exploration of informal network members' experiences. *Couns Psychol* 39(7):973–1023
- Lemoine D (2019) Rationally Misplaced Confidence. <https://ssrn.com/abstract=3138936>
- Macomber K (2018) I'm sure as hell not putting any man on a pedestal: male privilege and accountability in domestic and sexual violence work. *J Interpers Violence* 33(9):1491–1518
- Mannell J, Seyed-Raeisy I, Burgess R, Campbell C (2018) The implications of community responses to intimate partner violence in Rwanda. *PLoS ONE* 13(5):e0196584
- McAdams D (2020) Nash SIR: an economic-epidemiological model of strategic behavior during a viral epidemic. *Covid Econ* 1(16):115–134
- McAdams D, McDade KK, Ogbuoi O, Johnson M, Dixit S, Yamey G (2020) Incentivising wealthy nations to participate in the COVID-19 Vaccine Global Access Facility (COVAX): a game theory perspective. *BMJ Global Health* 5(11):e003627
- Michaels-Igbokwe C, Abramsky T, Devries K, Michau L, Musuya T, Watts C (2016) Cost and cost-effectiveness analysis of a community mobilisation intervention to reduce intimate partner violence in Kampala, Uganda. *BMC Public Health* 16(196):1–10
- Mogford E, Irby CA, Das A (2015) Changing men to change gender: combatting hegemonic masculinity through antiviolence activism in northern India. *Int J Sociol Fam* 41(2):71–93
- Mulawa MI, Reyes HLM, Foshee VA, Halpern CT, Martin SL, Kajula LJ, Maman S (2018) Associations between peer network gender norms and the perpetration of intimate partner violence among urban Tanzanian men: a multilevel analysis. *Prev Sci* 19(4):427–436
- Neal AM, Edwards KM (2017) Perpetrators' and victims' attributions for IPV: a critical review of the literature. *Trauma Violence Abuse* 18:239–267
- Nussbaum MC (2001) Symposium on Amartya Sen's philosophy: 5 adaptive preferences and women's options [01]. *Econ Philos* 17(1):67–88
- Oliver PE (1993) Formal models of collective action. *Ann Rev Sociol* 19(1):271–300
- Oliver PE, Myers DJ (2002) Formal models in studying collective action and social movements. In: Klandermans B, Staggenborg S (eds) *Methods of social movement research*. vol. 16. University of Minnesota Press. pp. 32–61
- Olson M (1971) *The logic of collective action* 2nd edn. vol. 124. Harvard University Press
- Peacock D, Levack A (2004) The men as partners program in South Africa: reaching men to end gender-based violence and promote sexual and reproductive health. *Int J Men's Health* 3(3):173–188
- Penner LA, Dovidio JF, Piliavin JA, Schroeder DA (2005) Prosocial behavior: multilevel perspectives. *Ann Rev Psychol* 56:365–392
- Pollak RA (2004) An intergenerational model of domestic violence. *J Popul Econ* 17(2):311–329
- Pronyk PM, Hargreaves JR, Kim JC, Morison LA, Phetla G, Watts C, Busza J, Porter JD (2006) Effect of a structural intervention for the prevention of intimate-partner violence and HIV in rural South Africa: a cluster randomised trial. *Lancet* 368(9551):1973–1983
- Rauhut H (2009) Higher punishment, less control? Experimental evidence on the inspection game. *Ration Soc* 21(3):359–392
- Rauhut H (2015) Stronger inspection incentives, less crime? Further experimental evidence on inspection games. *Ration Soc* 27(4):414–454
- Rigby SW, Johnson LF (2017) The relationship between intimate partner violence and HIV: a model-based evaluation. *Infect Dis Model* 2(1):71–89
- Rogers T, Goldstein NJ, Fox CR (2018) Social mobilization. *Ann Rev Psychol* 69(1):357–381. <https://doi.org/10.1146/annurev-psych-122414-033718>
- Rosato M, Laverack G, Grabman LH, Tripathy P, Nair N, Mwansambo C, Azad K, Morrison J, Bhutta Z, Perry H (2008) Community participation: lessons for maternal, newborn, and child health. *Lancet* 372(9642):962–971
- Sanyal P (2009) From credit to collective action: the role of microfinance in promoting women's social capital and normative influence. *Am Sociol Rev* 74(4):529–550
- Schuessler AA (2000) *A logic of expressive choice*. Princeton University Press
- Sen A (1999) *Development as freedom*. Oxford University Press
- Shakya HB, Hughes DA, Stafford D, Christakis NA, Fowler JH, Silverman JG (2016) Intimate partner violence norms cluster within households: an observational social network study in rural Honduras. *BMC Public Health* 16(1):233
- Slote M (2007) *The ethics of care and empathy*. Routledge
- Smith HJ, Pettigrew TF, Pippin GM, Bialosiewicz S (2012) Relative deprivation: a theoretical and meta-analytic review. *Person Soc Psychol Rev* 16(3):203–232
- Starmann E, Heise L, Kyegombe N, Devries K, Abramsky T, Michau L, Musuya T, Watts C, Collumbien M (2018) Examining diffusion to understand the how of SASA!, a violence against women and HIV prevention intervention in Uganda. *BMC Public Health* 18(1):616
- Stern E, Niyibizi LL (2018) Shifting perceptions of consequences of IPV among beneficiaries of Indashyikirwa: an IPV prevention program in Rwanda. *J Interpers Violence* 33(11):1778–1804
- Tauchen HV, Witte AD, Long SK (1991) Domestic violence: a nonrandom affair. *Int Econ Rev* 32(2):491–511
- Trivers R (2006) Reciprocal altruism: 30 years later. In: Kappeler PM, van Schaik CP (eds) *Cooperation in primates and humans*. Chapter 4, Springer. pp. 67–83
- Tsebelis G (1990) Penalty has no impact on crime: a game-theoretic analysis. *Ration Soc* 2(3):255–286
- UN (2017) Final list of proposed sustainable development goal indicators. United Nations. Retrieved 8/11/2017, from <https://sustainabledevelopment.un.org/content/documents/11803Official-List-of-Proposed-SDG-Indicators.pdf>
- Verma RK, Pulerwitz J, Mahendra V, Khandekar S, Barker G, Fulpagare P, Singh SK (2006) Challenging and changing gender attitudes among young men in Mumbai, India. *Reprod Health Matters* 14(28):135–143
- Wagman JA, Gray RH, Campbell JC, Thoma M, Ndyanabo A, Ssekasanvu J, Nalugoda F, Kagaayi J, Nakigozi G, Serwadda D (2015) Effectiveness of an integrated intimate partner violence and HIV prevention intervention in Rakai, Uganda: analysis of an intervention in an existing cluster randomised cohort. *Lancet Global Health* 3(1):e23–e33
- Watts RJ, Hipolito-Delgado CP (2015) Thinking ourselves to liberation?: advancing sociopolitical action in critical consciousness. *Urban Rev* 47(5):847–867
- Yule K, Hoxmeier JC, Petranu K, Grych J (2020) The chivalrous bystander: the role of gender-based beliefs and empathy on bystander behavior and perceived barriers to intervention. *J Interpers Violence*. <https://journals.sagepub.com/doi/full/10.1177/0886260520916277>

Acknowledgements

Thanks to Helen Harris-Fry and Komal Bhatia for close readings of the paper. Thanks to Neha Batura, Jolene Skordis-Worrall, Paula De Souza Leao Spinola, Gerard Abou Jaoude, and Ana Correa Ossa for great comments on economic theory. Thanks to Robbie Manning Smith and Jim Allen for advice on the mathematics of the paper. Thanks to Meghna Ranganathan, Ben Cislighi, and Lori Heise for insightful comments on theories of gender, norms and power. Thanks to members of the SNEHA team whose engaged fieldwork informed the paper: Unnati Machchhar, Shilpa Adelar, Bhaskar Kakad, Archana Bagra, Vanessa D'Souza, and Shanti Pantvaitya. Funding by Wellcome Trust (206417). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions

LG conceived of the paper, led all mathematical analyses and computer simulations, and wrote the first draft. RG reviewed micro economic theory in the paper and assisted with mathematical analyses. EK reviewed the paper from a social psychology perspective. DO and ND provided overall guidance. All authors contributed to the writing of the manuscript. All authors have given approval of this final version to be published and agree to be accountable for all aspects of the work. No others fulfil the criteria for authorship.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1057/s41599-021-00730-z>.

Correspondence and requests for materials should be addressed to L.G.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021