

Why Are Rights Revolutions Rare?

By Sultan Mehmood, Shaheen Naseer, and Daniel L. Chen¹

August 2024

We show that when traditional attitudes are challenged, norm disruptors pay the price. A recent experiment fostered more progressive gender attitudes among female teachers in Pakistan. We find, however, that fostering progressive gender attitudes elevated stress hormone concentrations in blood plasma and increased domestic violence. Leveraging random variation in the fraction of teachers treated within a school, we show that when additional teachers hold progressive attitudes—a moral bandwagoning effect—the costs of adopting progressive gender attitudes are attenuated. These findings underscore the toll of defying traditional gender norms: heightened stress and violence, ultimately hindering the advancement of gender rights.

JEL Codes: I24, I28, J16, O12, C93.

Keywords: Social Norms, Cortisol, Stress, Domestic Violence.

¹ Corresponding Author: Shaheen Naseer. University of Oxford (E-mail: Shaheen.naseer@oxfordmartin.ox.ac.uk) Sultan Mehmood: New Economic School (E-mail: smehmood@nes.ru), Daniel Chen: Toulouse School of Economics (Email: daniel.chen@tse-fr.fr). Daniel L. Chen acknowledges IAST funding from the French National Research Agency (ANR) under the Investments for the Future (Investissements d'Avenir) program, grant ANR-17-EUR-0010. This research has benefited from the financial support of the research foundation TSE- Partnership and ANITI funding. We would like to thank the Director of Training and Development, Miss Sumera Morris, and the Director of Academics and Curriculum Development Miss Faria Hafeez at the Progressive Education Network for their invaluable support and cooperation throughout the experiment. We also thank Avner Seror, Matthew Ridley, Alberto Bisin, Paola Giuliano, Chris Bidner, Ekaterina Zhuravskaya, Usman Mahar, Matt Lowe, Eric Reinhardt, Hafeez-ul-Hassan, Ali Bakhtawar and Gautam Rao for their comments and feedback. Sajwaar Khalid, Hafeez-ul-Hassan, and Bakhtawar Ali provided excellent research assistance and on-the-field support. The trial was Pre-Registered at AEA RCT Registry with ID *AEARCTR-0007465*. A local IRB obtained from the Lahore School of Economics with IRB Number *RERC-062021-03*. A discussion of ethics of this experiment following framework Asiedu et al. (2021) is also provided in Appendix C2.

“Do not fear to be eccentric in opinion, for every opinion now accepted was once eccentric.”
in The Autobiography of [Bertrand Russell \(1967\)](#).

I. Introduction

Many models of social norms involve multiple equilibria. For instance, 'traditional' and 'progressive' attitudes are viewed as distinct equilibria within this framework (see e.g., Cole et al., 1992). A critical question then arises: What mechanisms maintain the stability of these equilibria? Why do individuals face substantial costs when attempting to shift from traditional to progressive norms? To explore these questions, we investigate the societal costs imposed on those who challenge prevailing norms, using one of the most prominent naturally-occurring gender rights initiatives in recent times as our case study. Specifically, we examine the impact of an award-winning movie, *Bol*, co-developed at Johns Hopkins University ([John Hopkins News, 2011](#)), which presents a visual narrative advocating for gender rights in Pakistan.

The transition from entrenched social norms to a more equitable status quo is not merely an ideological shift but a foundational step towards the economic and political empowerment of vulnerable groups. Recognition, or the mutual acceptance of others as possessing equal rights, is a critical precursor to this empowerment process ([Smith, 1759](#); [Hegel, 1820](#)). Consequently, campaigns advocating for equal rights are often framed as "struggles for recognition" ([Taylor, 1992, p. 26](#)). Despite the evident benefits of progressive gender attitudes—such as increased female labor force participation and overall economic advancement ([Field et al., 2010](#); [Doepke et al., 2012](#); [Hazan et al., 2019](#))—social attitudes are notoriously slow to evolve ([Giuliano and Nunn, 2021](#)). This raises pivotal questions: How can movements for rights revolutions gain momentum, and what barriers do they face? In this paper, we present evidence from an experiment designed to foster progressive gender attitudes and examine the consequences faced by those who challenge the status quo.

A large body of literature has established that the expansion of women's rights and progressive attitudes has coincided with increases in economic development ([Duflo, 2012](#)

provides a review). More rights for women can lead to more spending on health and education ([Doepke et al., 2012](#)), financial development ([Hazan et al., 2019](#)), and increased labor force participation ([Gay, 2023](#)). Given the substantial economic benefits of embracing progressive gender attitudes, what prevents societies from adopting them? Norm disruptors attempting to challenge the status quo may face denunciations, social stigma, and even outright violence ([Bursztyn et al., 2020](#); [Giuliano, 2020](#); [Anderson, 2021](#)).

In this paper, we leverage a unique opportunity to investigate the cost of implementing rights innovations through an experiment conducted at the Progressive Education Network (PEN), one of the largest networks of charter schools in the world, operating schools across Pakistan. The PEN experiment randomly assigned teachers to treatment arms designed to promote progressive gender attitudes, as previously studied by [Mehmood, Naseer, and Chen \(2024\)](#).² It documented the student-to-teacher transmission of progressive gender attitudes, investigated its impact on student academic achievement and cooperation, and explored the mechanism through which students' attitudes change with students cross-randomized into mixed vs. same-gender study groups. This paper integrates existing data on participating teachers, including their gender attitudes and the proportion of teachers in the school exposed to the progressive attitudes treatment. It combines this with novel data on teachers' stress levels measured through blood cortisol concentrations and their experiences of and attitudes towards domestic violence as defined by the United Nations ([United Nations, 2021](#)). In doing so, it provides the first experimental evidence on the potential unintended consequences of promoting progressive gender attitudes, wherein such efforts may inadvertently elevate stress levels and incidences of domestic violence. Previous research has shown that visual narratives, such as movies and edutainment, can foster more progressive gender attitudes ([Banerjee, et al. 2019](#); [Mehmood et al., 2024](#); [Riley, 2022](#)). Similarly, in the PEN experiment, teachers became significantly more progressive in gender attitudes, implicit attitudes, and political action. Since these, and other studies have found positive effects of progressive gender rights, we use new data to document backlash effects. We also investigate the impact of moral bandwagoning on mitigating the costs of gender rights revolutions.

² Because our analysis builds on the pre-registered work of [Mehmood et al. \(2024\)](#), we address the risk of Type I error inflation by adjusting p-values for multiple hypothesis testing. We utilize both Anderson's (2008) two-stage Benjamini-Hochberg procedure for FDR control and the Romano-Wolf step-down method. Our results are robust in both instances.

Our findings suggest two broad explanations for the persistence of traditional gender attitudes: fostering progressive gender attitudes is costly, with norm subverters paying through increased stress and domestic violence. Understanding which of these costs may be more amenable to change is crucial for understanding the stickiness of traditional attitudes and devising policies to shift potentially harmful attitudes. We measure costs associated with fostering more progressive attitudes by using self-reported stress, hormonal responses to stress in blood plasma, and domestic violence, complementing important new work on the interplay between gender rights and domestic violence ([Anderson, 2021](#)).

We measure the costs associated with holding novel gender attitudes regarding stress and violence. We assess stress in two ways. First, we measure self-reported stress via a survey. Second, we borrow from neuroscience an objective measure of stress: pre-breakfast blood cortisol concentration ([Hessl et al., 2002](#); [Adam et al., 2006](#)). Cortisol is a hormone produced in response to stress ([El-Farhan et al., 2017](#)). A large body of literature in neuroscience finds cortisol concentrations are a prominent “biomarker” of stress (see, e.g., [Hellhammer et al., 2009](#)) where prolonged stress states elevate the cortisol concentration in blood ([Lupien et al., 2009](#)).³ Several recent studies also link high levels of blood cortisol with clinical depression (see e.g., [Qin et al., 2016](#)). These measures capture mental well-being, increasingly recognized as an important indicator for economic development ([Schilbach et al., 2016](#); [Ridley et al., 2020](#); [Haushofer et al., 2020](#); [Bossuroy et al., 2022](#)). We measure violence in the form of domestic violence ([Macmillan and Gartner, 1999](#)). Violence against women, especially those who are more educated or are working, is especially prevalent in traditional societies, with three-quarters of all violence against women perpetrated by domestic partners ([Aizer, 2010](#); [Anderson, 2021](#)). Indeed, domestic violence may stall the cultivation of progressive gender attitudes. The majority of primary school teachers in Pakistan are women, and all PEN teachers are female, making the study of domestic violence particularly important. We conducted two measurements related to domestic violence. We measure whether the teacher reports having, over the past year, been a victim of domestic violence defined as “someone hurting or trying to hurt by hitting, kicking, burning, grabbing, pinching, shoving, slapping, hair-pulling, biting, denying medical care or

³ The pre-breakfast blood cortisol concentration measurements is a more persistent and reliable tracker during the course of the day relative to cortisol concentration in saliva.

forcing alcohol and/or drug use, or using other physical force” ([UN, 2021](#)).⁴ We also measure to what extent the teacher views domestic violence as justified. These measures help alleviate concerns that the rise in reported domestic violence is merely due to a change in the reporting threshold (e.g., [Kishor and Johnson, 2004](#); [Flake, 2005](#); [Burazeri et al., 2005](#)).

Our first set of results indicates that a year after the intervention, the cultivation of progressive gender attitudes elevated stress hormone concentrations in the blood plasma by 0.3 standard deviations and domestic violence by 0.35 standard deviations. To put these measurements in perspective, the effect sizes are comparable to recent interventions aiming to reduce stress and domestic violence. For instance, [Shreekumar and Vautrey \(2022\)](#) find that randomly assigning a popular mindfulness app can reduce stress by 0.4 standard deviations and [Castilla \(2022\)](#) randomly assigns couples therapy to find it reduces domestic violence by 0.25 standard deviations. [Taylor and Holston \(2016\)](#) document that a few sessions of music therapy can reduce cortisol concentrations by 0.2 standard deviations, [Barker et al. \(2022\)](#) show that cognitive behavioral therapy improves mental well-being by 0.29 standard deviations, [Wagman et al. \(2015\)](#) document that four community support sessions can reduce physical intimate partner violence up to 0.69 standard deviations, while [Shah et al. \(2022\)](#) find that a 90-minute goal setting session by a trained expert can reduce physical intimate partner violence by 0.2 standard deviations.

Moreover, the effects on stress and domestic violence are compounded when the visual narrative is combined with self-persuasion and is consistent with the male backlash hypothesis, that is, as women rally for more rights, they become more vulnerable to domestic violence ([Macmillan and Gartner, 1999](#)). When local traditions mandate a woman’s role to be one of subordination, redefining that role is taxing. As a result, women deviating from traditional gender attitudes may experience domestic violence and heightened stress.

Our second set of results utilizes random variation in the proportion of treated teachers within a school to explore whether an increase in the adoption of progressive attitudes by teachers can mitigate domestic violence and stress. We find that as more teachers adopt

⁴ It is important to note that we focus on domestic violence, which is distinct from intimate partner violence. We do this for two reasons. First, physical abuse from family members was flagged as a key issue by the PEN administration. Second, the majority of teachers are unmarried in our sample, essentially precluding an analysis of intimate partner violence.

progressive gender attitudes, the deleterious effects of stress attenuate. This is consistent with theoretical literature on herding and bandwagoning ([Banerjee, 1992](#); [Shiller, 1995](#); [Sunstein, 1996](#)), which often cite information cascades, social conformity, and reputation for driving changes from one equilibrium to another. Our design allows us to estimate this bandwagoning effect because we induce exogenous variation in the fraction of treated teachers within the school. While the exact mechanism is not observed, the estimates suggest that when roughly half the teachers within a school are treated, the negative effect of progressive gender attitudes on stress disappears.⁵ The greater the alignment of ideas among teachers, the less stressed they feel. However, the adverse impact of adopting progressive gender attitudes on domestic violence does not diminish. This implies that novel and contentious ideas in society can be fostered but it comes at costs for the norm subverter but at least some of these costs decrease once the new ideas diffuse.

Three aspects of our experiment offer the interpretation that our results are unlikely to be driven by experimenter demand. First, we use blood cortisol concentration to measure stress. Cortisol is secreted by adrenal glands involuntarily in response to stress ([Hellhammer et al., 2009](#)). Second, we observe that views regarding the justification of domestic violence are unaffected by the treatments. Third, we follow [Dhar et al., \(2018\)](#)'s introduction of the Marlowe-Crowne social desirability scale in economics, a survey module developed by social psychologists to rigorously measure a person's propensity to misreport in surveys ([Crowne and Marlowe, 1960](#)). When we discard the individuals who score high on this scale, the results are essentially identical.

Finally, we note two additional treatment arms in our experiment. The utilitarian training intervention was designed to underscore the practical benefits of empathy in education. By associating empathy with self-interest, the program aimed to motivate teachers to embrace and foster empathetic and gender-positive behaviors. The training sought to appeal to teachers' rational decision-making, positing that empathy, as a skill, can be developed to encourage more gender-progressive attitudes. Similarly, the empathy training intervention, based on the idea that empathy is a cultivable trait, sought to promote gender-progressive attitudes by leveraging the

⁵ The director of training at PEN reported anecdotes of teachers forming mental support groups which may explain this result.

natural human capacity for understanding and compassion, potentially affecting perceptions of gender equality. Despite the theoretical rationale that these interventions could induce shifts toward gender-progressive attitudes, the empirical evidence revealed null effects for both interventions.

We contribute to two key literatures. First, a large literature has investigated how statistical information helps social norms equilibria be stable or unstable ([Bursztyn et al., 2020a](#); [Bursztyn et al., 2020b](#)). We provide a different interpretation of the stability of social norms using the classic literature on bandwagoning ([Becker, 1991](#); [Bikhchandani and Hirshleifer, 1992](#); [Banerjee, 1992](#); [Shiller, 1995](#)). Our experimental results complement observational analyses of the backlash that may arise with female empowerment ([Anderson and Bidner, 2015](#); [Aizer, 2010](#); [Anderberg et al., 2016](#); [Alsan et al., 2017](#)). We also show how stress dissipates as more peers shift their attitudes, with these costs attenuating once the new ideas take root. However, domestic violence remains persistent regardless of the fraction of peer women who adopt the progressive gender attitudes. We contribute to the largely theoretical literature on bandwagoning by providing experimental evidence that moral bandwagoning can attenuate costs associated with the backlash generated from cultivating novel nonconformist ideas. When individuals are challenging the status quo, the fact they deviate from the norm can have costly consequences for themselves. We contribute to this literature by documenting these costs and prescribing a way to mitigate some of these costs by increasing the number of peers who hold similar views.

We also contribute to a growing literature on mental health and economic development. Stress and domestic violence can create a vicious cycle. Poverty and low psychological well-being mutually reinforce each other ([Lund et al., 2011](#); [Ridley et al., 2020](#); [Biasi et al., 2021](#); [Ravensteijn et al., 2017](#)), which dovetails with the burgeoning literature documenting substantial economic consequences of mental health ([Mani et al., 2013](#); [Schilbach et al., 2016](#); [Rao et al., 2021](#); [Barker et al. 2022](#); [Bossuroy et al., 2022](#)). The lowest income groups are up to 3 times more likely than the rich to experience depression or anxiety ([Ridley et al., 2020](#)). We measure mental well-being with surveys and blood cortisol concentrations for individuals in a traditional society. For native members of a particular society or culture, their attitudes are usually a matter of course and, much like the rules of grammar, go unnoticed until they are violated. Sanctions are responses to norm deviance. One may distinguish sanctions by whether

they address the body or the psyche. Whereas physical sanctions, such as domestic violence, bring about physical pain, psychological ones address feelings and emotions. We show that both types of sanctions punish norm subverters as they struggle for their rights. The accompanying stress and violence, as vulnerable groups struggle for their rights, may stymie rights revolutions.

The rest of the paper is organized as follows. Section II provides background and the experimental design. Section III outlines data and empirical specifications, while Section IV presents key findings. Section 5 reports a series of robustness tests and a discussion of ethics. A final section concludes. Online Appendix A discusses heterogeneity. Online Appendix B offers supplementary Figures and Tables. Appendix C provides additional information on the data. Together, the appendices provide detailed accounts of the data collection processes, randomization process, ethical standards, and accounting for multiple hypothesis testing, experimenter demand, and spillover effects.

II. Background and Study Design

Background and Design.—As reported in [Mehmood, Naseer, and Chen \(2024\)](#), we work with the Progressive Education Network (PEN) to embed the field experiment within their regular teachers training.⁶ The PEN schools work to improve the quality of education via a public-private partnership, similar to charter schools in the United States ([Angrist and Pischke, 2014](#)). These schools, which are an initiative of United States expatriate Pakistanis, were formed to replicate the "success" of charter schools in the United States and have grown exponentially over the last two decades, from almost none in 2000 to over 5,000 charter schools by the end of 2024. These schools are managed using public funds by the private actors in a public-private partnership. In particular, we implement a randomized evaluation in all PEN schools chartered in Punjab, the largest province of Pakistan, where the network employs 607 teachers responsible for around 15000 students across 52 schools. We randomly assign the 607 teachers to the following treatment arms: (i) utilitarian treatment (121 teachers); (ii) malleability treatment (121 teachers); (iii) visual narrative treatment (122 teachers) (iii) visual narrative joint with self-persuasion treatment (121 teachers) and (iv) control group receives training on procedures to open a bank account in Pakistan (122 teachers). It should be noted that all teachers in our sample already had

⁶Appendix D details the principal findings regarding the influence of our interventions on gender attitudes.

a bank account, making the control training likely to have a null effect. The baseline, midline, and endline surveys were conducted in February 2021, September 2021, and March 2022, respectively, with treatment roll out occurring in March 2021. In the earlier paper, we provide three principal insights: (1) generic interventions aimed at enhancing empathy do not substantially alter gender attitudes or boost student achievement, underscoring the limitations of non-specific non-gender themed strategies. (1) even minimal gender-centric interventions, such as screening a film with gender themes, can foster more progressive gender attitudes, underscoring the potency of focused measures. Lastly, to achieve a broader and more profound effect on educational outcomes, we need to pair gender-rights film with a related curriculum, which we show has an impact on both teacher and student attitudes, as well as on student achievement.

Bol Movie Visual Narrative Treatment.— Our first treatment group watched a movie with a structured discussion on the gender-related themes of the movie. We arranged a live movie screening within this treatment arm along with the structured discussion that followed the screening. The emotionally charged movie was about 180 minutes long.⁷ The screening of the movie followed an hour-long discussion about gender rights. The movie, *Bol* (literally, *to speak up*), is a Pakistani Urdu-language social drama with a strong female lead on death row telling her story to a journalist of why she found it necessary to murder her father. The theme revolves around how the female lead and her sisters deal with their father's obsession to have a son and his staunch reluctance to recognize his daughters and wife as equals. The movie was directed by celebrated Pakistani director Shoaib Mansoor and produced by Johns Hopkins Bloomberg School of Public Health Center. This award-winning movie is believed to be a critical milestone in Pakistani cinema, reaching over 5.7 million individuals in Pakistan within a year of its release, and is believed to increase awareness about traditional gender attitudes in Pakistan ([John Hopkins News, 2011](#)). The movie eventually became one of the highest-grossing Pakistani films of all time.

Joint Visual Narrative and Self-Persuasion Treatment.— Our second treatment intensified the visual narrative of the movie and structured discussion with a semester-long gender attitudes course taught by the treated teachers to PEN primary school students. This is inspired by the

⁷ This is roughly the same length as the Utilitarian and Malleability treatment lectures.

evidence on the efficacy of social-emotional learning and teaching as an instrument of self-persuasion ([Yeager et al., 2019](#); [Eskreis-Winkler et al., 2019](#)). The teachers with students envisioned and reflected on women's rights in contemporary Pakistan. These gender study classes were held once a week for two hours from February to May 2021. The classroom exercises involved readings, drawing, and other activities encouraging students to reflect on gender attitudes and women's lack of rights and freedoms.

Utilitarian and Malleability Treatments.— Another way to increase equitable gender attitudes is to shift empathy - putting oneself in another's shoes plausibly increases equitable attitudes, similar to the ideas of John Rawls. We considered two methods based on utilitarianism and on social psychology that can enhance empathy. The third treatment provided training that emphasized the benefits of empathy and how it can positively affect their teaching. The main message of the treatment: *“All types of evidence backs the idea that empathy is good for you. It is not just the right thing to do but also the most sensible thing to do for your performance as a teacher.”* The fourth treatment arm provided training that emphasized the malleability of empathy and was based on identity theory. The malleability training's main message was: *“All types of evidence backs the idea that empathy is not fixed but malleable. It is a skill that can be developed.”* The complete transcript of these trainings is presented in the companion paper. The utilitarian training intervention aimed to highlight the practical advantages of cultivating empathy in educational environments. The intervention's design aimed to synchronize the teachers' inherent self-interest with the broader educational goal of fostering empathetic behavior towards others, which is anticipated to lead to more gender-positive attitudes. The utilitarian aspect of the teacher training specifically targets the rational aspects of decision-making among educators. Concurrently, the malleability treatment is predicated on the premise that empathy is not a fixed trait but rather one that can be actively nurtured and strengthened. This aspect of the training posits that by enhancing empathy, a pathway to more gender-progressive attitudes may be established. Collectively, these treatments seek to capitalize on the universal human ability for empathy, positing that its deliberate cultivation can be a transformative force in promoting gender equality.

We did not find significant impacts of these two trainings so we will interpret these as control treatments along with our placebo. We, however, always control for these treatments in all our main regressions. For more details on all the treatments, including full transcripts of the

Utilitarian and Malleability training, please refer to the companion paper ([Mehmood et al., 2024](#)). The table provides a comprehensive summary of all studies involving the Progressive Education Network sample, alongside related research findings. Refer to Figure B3 for a detailed overview.

III. Data and Empirical Strategy

A. *The Data*

Sample.— The sample consists of 607 teachers and their 13,932 pupils across 52 charter schools of the Progressive Education Network (PEN) in the province of Punjab. All of these teachers are female and teach every subject from Kindergarten to Grade 6. PEN organizes training workshops for teachers periodically and we embed our experiment within the PEN teacher-training drives. Since our experimental intervention is embedded within PEN’s training we have no attrition: all 607 PEN teachers in the province of Punjab participated in both the experiment and the surveys.⁸ Baseline survey and treatment took place in January 2021 and endline in December 2021. A balance of our treatments over pre-treatment stress, domestic violence, and baseline teacher characteristics is reported in Table B1.

Outcomes on Stress.— We assess stress in several ways. Our first two variables are self-reported perceptions of stress, measured via two survey questions. The first one is a rating on a 5-point Likert scale to the question, “Overall, how stressed are you?” The second question is a binary response of yes and no to the question, “Are you stressed?” Noise, social-desirability bias, and concerns about misreporting, however, make the interpretation of self-reported measures complicated. Therefore, we use pre-breakfast cortisol concentration in blood—the stress hormone excreted in response to stress—to get a more accurate estimate of stress ([El-Farhan et al., 2017](#)). A large body of literature in neuroscience finds blood cortisol levels are a prominent “biomarker” of stress (see e.g., [Hellhammer et al., 2009](#)).⁹ Cortisol is released in response to psychological or physiological strain on the body. For instance, it increases following injuries, intense physical exertion, or during public speaking, performing taxing mental arithmetic, or enduring unpleasant situations such as waiting in the operation theater before a

⁸ We thank the Progressive Education network for their invaluable support and cooperation to ensure this remarkable take-up. We especially wanted to thank the Director of Training, Miss Sumera Morris and her staff at PEN, who provided invaluable cooperation and support throughout this intervention. It is also worth clarifying, we did not communicate our research question or the treatment status of the teacher, with the PEN administration.

⁹ Several recent studies also find that high-levels of cortisol in blood is correlated with clinical depression (see e.g., [Qin et al., 2016](#)). This is unsurprising since stress is found to cause fundamental changes in neuronal structure and functioning of the brain ([Lupien et al., 2009](#); [Goldfarb, 2020](#)).

surgical procedure ([Kirschbaum et al., 1993](#); [Ferracuti et al., 1994](#)). The blood cortisol levels can vary throughout the day but plasma cortisol is more stable than salivary cortisol and *less responsive* to time of day relative to salivary cortisol. Nonetheless, we are particularly careful in observing the protocol of cortisol sample timing: we measure cortisol concentration in blood plasma in the morning at 7 a.m. (following [Kische et al., 2021](#)). To minimize gaps between teachers getting their cortisol tests done at different times and hence prevent a timing imbalance, we ensured all tests were made within a 30-minute time window. That is, by 7.30 am we collected samples of all 607 teachers. We thank the competent team of volunteers and lab attendants from the Chughtai Labs who made this possible. The readings are made in micrograms per deciliter (mcg/dl), using the standard Chemiluminescence Immunoassay (CLIA) technique. The cortisol concentration we observe in our sample is about 11.15 mcg/dl with a standard deviation of 3.31. To put this into perspective, persons suffering from severe anxiety, and adrenal or pituitary gland disorders typically have cortisol readings of above 20 mcg/dl ([Mantella et al., 2008](#); [Armario et al., 1996](#)).

Outcomes on Domestic Violence.— We measure domestic violence in two ways based on United Nations definitions. First, we evaluate whether the teacher has been a victim of domestic violence in the past. In our survey, we define both who is most likely to commit domestic violence (brother, father, or husband) and what is included in domestic violence (United Nations, 2021). Specifically, we use the following text: “Over the past year, have you been a victim of domestic violence i.e., physical violence by your husband, father or/and brother? By physical abuse, we mean hurting or trying to hurt a partner by hitting, kicking, burning, grabbing, pinching, shoving, slapping, hair-pulling, biting, denying medical care or forcing alcohol and/or drug use, or using other physical force.” To disentangle experiencing domestic violence from differential reporting on domestic violence, which may be a function of teachers’ definition of or views on domestic violence, we also assess the extent to which teachers find domestic violence justified. In particular, we field the following statement: “Domestic violence by husbands cannot be justified (rate from 1 to 5)”. The second measure allows us to investigate whether potential treatment effects are entirely explained by changes in reporting or expansion in the definition of domestic violence.

Main Explanatory Variables.— Our key explanatory variables are dummies for the four treatments. VN_i and $VN \& SP_i$ are dummies that switch on if the teachers were assigned the

visual narrative of the movie *Bol* and joint treatment augmenting visual narrative with the curriculum. U_i and M_i denote dummies that switch on if the teachers were assigned the Utilitarian and Malleability treatment respectively. A control group receives information on procedures on how to open a bank account in Pakistan. It is worth noting that all 607 teachers in our experiment already have their individual bank accounts on which they receive their salaries.

B. *The Empirical Specification*

The experimental design allows us to run a simple OLS regression:

$$Y_i = \alpha + \beta U_i + \gamma M_i + \delta VN_i + \omega VN \& SP_i + X_i \mu + \epsilon_i \quad (1)$$

where Y_i is the respective outcome on gender attitudes or sanction to the individual i , VN_i is the dummy variable equal to one if the teacher is assigned to the visual narrative treatment arm of the movie *Bol*, while $VN \& SP_i$ complements the visual narrative with the gender studies curriculum; U_i is a dummy that switches on if the teacher is assigned to the Utilitarian treatment arm; M_i when assigned to the malleability of empathy treatment arm. X_i is a vector of individual-level controls. Since we randomly assign teachers to our treatment groups, we cluster standard errors at the teacher level. In our investigation of moral bandwagoning, where the variation in the fraction of teachers treated only varies across schools, not within them, we show the results are robust to clustering the standard errors at the school level. It is noteworthy that our core findings remain consistent regardless of whether we cluster standard errors at the teacher or school level.

To confront the risk associated with multiple hypothesis testing, we present our principal results with additional multiple hypothesis-adjusted p-values, utilizing two rigorous statistical methods. The FDR is controlled through a two-stage adaptation of the Benjamini-Hochberg procedure by Anderson (2008), while FWER is managed via the Romano-Wolf step-down approach. We report Newey-West robust p-values alongside FDR q-values and FWER-adjusted p-values.

IV. Main Results

The Background Experiment.— The PEN experiment examined the transmission of progressive gender attitudes from teachers to students and evaluated its effects on student academic performance and cooperation in strategic dilemmas. By randomly assigning students to mixed or same-gender study groups, the study also delved into the mechanisms facilitating this transmission of gender attitudes. The findings suggest that visual narrative of the Bol movie hold the potential to cultivate more progressive gender attitudes, subsequently influencing students' gender attitudes, academic test scores, and the gender gap in mathematics test scores. Both teachers' and students' gender attitudes show a shift toward more progressive views. Notably, the observed increase in cooperation between genders appears to be a key driver of these outcomes. For further details, please refer to the paper ([Mehmood, Naseer and Chen, 2024](#)). For more information, refer to Figure B3 that offers a consolidated summary table of all research conducted with the Progressive Education Network sample, including descriptions of associated studies.

Effects on Stress of Teachers.— Our standalone visual narrative and joint treatment increased stress. In Column 1 of Table 1, we observe that the joint visual narrative and self-persuasion treatment increased self-reported stress levels by about half a point on a 5-point scale. This is equivalent to an 18% increase over the sample mean. On the other hand, the visual narrative alone increased stated stress by about a third of a point on this 5-point scale, a 13% increase over the mean dependent variable. Similar results are found in the binary variable answers to the question “Are you stressed?” The joint visual narrative and curriculum group is about 25 percentage points more likely to report they are stressed. The self-reported stress results hold for a more objective measure of stress: plasma cortisol—a hormone produced in response to stress—is also found to be impacted.¹⁰ The joint treatment increases blood cortisol levels by 1.13 micrograms per deciliter (Table 1, Column 3). This is equivalent to a 10% increase over the sample mean cortisol concentration of 11.15 mcg/dl. Column 4 of Table 1 reports this result in terms of standard deviations: even one year after the treatment, the joint visual narrative and

¹⁰ The blood samples were collected and transported to a prominent lab in Lahore, where measurement of plasma cortisol concentrations are recorded using the standard Chemiluminescent immunoassay (CLIA). All 607 teachers had their sample collected within our specified 30-minute window with the help of a team of volunteers and lab attendants. Vacunas tubes were deployed to mitigate the discomfort of a venipuncture syringe. We obtained results of blood plasma concentration readings from the lab of all teachers within 24 hours.

self-persuasion treatment increased stress hormonal response in plasma by about 0.35 standard deviations, while the visual narrative with structured discussion elevated blood plasma concentrations by about 0.2 standard deviations.¹¹ These results are not limited to changes in means. In Figure B1, we observe that distributions of visual narrative and joint treatment shifts relative to the placebo-treated teachers. We observe that both visual narrative and joint treatment groups first-order stochastically dominate the blood cortisol concentrations in the control group. The results in Table 1 and Figure B1 tell a cautionary tale: progressive gender attitudes can be cultivated as in our case ([Mehmood et al., 2024](#)) and those documented in other studies ([Dhar et al., 2018](#); [Banerjee, et al. 2019](#); [Riley, 2022](#)), but this norm subversion may not be costless.

Bandwagoning Effect on Stress as more Teachers are Treated.— Our empirical design allows us to experimentally explore a mitigating factor that diminishes the cost of cultivating progressive gender attitudes. We exploit the variation of treatment intensity within schools to investigate a moral bandwagoning effect: if enough teachers cultivate progressive gender attitudes within a school, the deleterious effects on stress are mitigated. Table 3 reports these results. We estimate that when about 45% of the teachers within a school are treated with the joint treatment of visual narrative and self-persuasion, the adverse impact on blood cortisol concentrations disappears. These results suggest a “bandwagoning effect” where greater community level adoption of new ideas mitigate the potential costs associated with holding internalized norms ([Becker, 1991](#); [Banerjee, 1992](#); [Shiller, 1995](#)). In this specification, since we utilized treatment intensity variation is at the school level, we, therefore, cluster standard errors at the school level in Table B9 of Appendix B, where the interaction terms in Columns 3 and 4 are significant at the 10% level. The distribution of fraction of treated teachers within schools is also reported in Figure B2 of Appendix B. The figure indicates a range in the variation of the proportion of teachers who received treatment within the school. The two distributions corresponding to the treatments appear to be consistent with randomization.

Effects on Experiencing Domestic Violence.— We investigate and find evidence that holding more progressive gender attitudes came at the expense of increased domestic violence experienced by the teachers. We conducted a survey inquiring whether teachers experienced domestic violence over the past 12 months and investigated if the treatments impacted their answers. Domestic violence was defined as “someone hurting or trying to hurt by hitting,

¹¹ The Utilitarian and Malleability treatments, however, have no effect on stress or any measure of gender attitudes.

kicking, burning, grabbing, pinching, shoving, slapping, hair-pulling, biting, denying medical care or forcing alcohol and/or drug use, or using other physical force” ([UN, 2021](#)). Table 2, Columns 1 and 2, reveal these findings. Teachers exposed to the visual narrative are approximately 0.25 standard deviations more likely to report recent domestic violence incidents compared to the control group (see Panel A). This represents about double the incidence of domestic violence compared to the sample average (see Panel B). Worryingly, this effect is magnified to about a third of a standard deviation when the visual narrative is combined with the curriculum. Taken together, our results indicate that norm subversion elevates both stress and teachers’ stated experience of domestic violence.

Effects on Views about Domestic Violence.— The treatments increasing teachers’ experience of domestic violence may also be explained by the alternate channel of a change in attitudes. That is, the treatment may have shifted teachers’ views regarding the acceptability of domestic violence. For instance, holding more progressive gender attitudes may prompt teachers to view domestic violence as more unjustified and less likely to report its incidence. In Columns 3 and 4 of Table 2, we explore this channel by investigating whether our treatments impacted acceptability of domestic violence on a 5-point Likert scale. Panel A of Table 2 reports the results in standardized units, Panel B reports them in the original Likert scale. We find the shift in views regarding the justification of domestic violence to be statistically indistinguishable from zero. In contrast, as we noted above, the visual narrative and jointly treated teachers were more likely to state they experienced an incident of domestic violence suggesting that the incidence of domestic violence may have increased as a result of our treatment.

Effects on Stated Domestic Violence Due to Misreporting.— A related possibility is that the treated teachers are more likely to misreport that they have been victims of domestic violence. This is because domestic violence is often misreported in surveys ([Flake, 2005](#)). To address this issue, we follow [Dhar et al., \(2018\)](#), who suggest using the Crowne-Marlowe scale, a survey module developed by social psychologists that rigorously measures and purges individuals that are most likely to answer inaccurately ([Crowne and Marlowe, 1960](#)). Table B10’s Columns 1 and 2 reveal that when we drop the teachers who are most likely to misreport in their survey responses, according to the Crowne-Marlowe scale, the results are essentially

identical.¹² Taken together, results from Table 2 strongly suggest that the cost of holding progressive gender attitudes, the increase in stress, comes alongside a rise of domestic violence that is not due to our treatment changing views or reporting of domestic violence.

Effect on Violence as more Teachers are Treated.— We also leverage our experimental set-up to experimentally explore as more teachers are treated to the *Joint Visual Narrative and Self-Persuasion* treatment within a school whether the costs associated with social sanctions such as domestic violence fall. As opposed to reduction of costs associated with internalized norms as measured by blood cortisol concentrations (Table 3’s Columns 3 and 4), we do not find much evidence that a greater share of treated teachers within a school mitigates the sanction of domestic violence (Column 5 of Table B6). This result also suggests that the increase in stress is not one-to-one linked to domestic violence and that internal and external sanctions play separate roles for limiting rights revolutions.

Heterogeneity by Previously Seeing Bol Movie.— Since about half of the teachers stated that they had watched the movie when it was released a decade earlier in 2011, we investigate if those teachers who had previously watched Bol are more likely to be impacted by the visual narrative or the joint treatment. Table B5 reports these results. We find mixed evidence for the heterogeneous effect of treatment on those who had previously watched the movie Bol. Significant heterogeneity appears for the question “are you stressed” in Column 2 indicating that those who previously watched Bol are less affected by the treatment, but insignificant heterogeneous effects appear for domestic violence, self-reported stress as a continuous measure, and blood cortisol levels. This may be due to many factors. For instance, if the teachers watched the movie a decade earlier when it premiered, and the effects of the movie on teachers may have dissipated after 10 years. It could also be the case that the structured discussion on the gender rights themes of the movie among peers reinforced the message of the movie beyond just watching the movie.¹³

V. Ethics and Robustness Analysis

¹² Specifically, Crowne-Marlowe assesses if the baseline results change, when teachers who answer yes to the following statements are discarded: 1) I am never jealous of another person’s good fortune 2) I am always a good listener. 3) I am never angry. Intuitively, the method drops subjects from the experiment who may be most prone to misreport in surveys.

¹³ All available outcome variables, for teachers and students, with the 10 hypotheses pre-registered in the Pre-Analysis Plan (<https://www.socialscienceregistry.org/versions/161842/docs/version/document>) are reported in Table B1.1 and B1.2 of the Appendix B, together in two tables.

Research Ethics Approvals.— Our study protocols were also reviewed and approved by Lahore School of Economics with IRB Number RERC-062021-03. The Lahore School of Economics review board actively engaged with the study participants, made several random spot visits to our experimental site and ensured that ethical protocols, for instance, consent from teachers and PEN administration was sought rigorously. The Lahore School of Economics review board actively engaged with the study participants, made several random spot visits to our experimental site and ensured that ethical protocols, for instance, consent from teachers and PEN administration was sought rigorously. Majority of the research team is Pakistani with Dr. Shaheen, who was based in Pakistan at the time. The study closely collaborated with the PEN training department to design and implement the experiment.¹⁴ We also discuss ethics in the Appendix C2, following the framework suggested in [Asiedu et al. \(2021\)](#).

Balance and Attrition.—The fact that the experiment was embedded within PEN’s regular training implied that we had no attrition. PEN training staff was headed by the training director and ensured every teacher responded. Nevertheless, a lack of balance might still complicate causal interpretation of our results. Before proceeding to presenting our main results, we first show in Table B1 that our randomization was successful in creating balance among teachers. We also conduct a Joint Orthogonality Test for balance as suggested by [Bruhn and McKenzie \(2009\)](#) to also find similar balance (see Table B3 in Appendix B). Differences across treatment groups are small in magnitude, and almost all p-values are larger than 0.10. For example, teachers’ educational specialization, years of education, experience, and marital status are balanced, so are their pretreatment measurements of stress, domestic violence and gender attitudes. Table B2 presents further balance checks concerning the proportion of teachers who received visual narrative alone, as well as the combined approach of visual narrative with self-persuasion (film plus curriculum).

Spillovers.—Our experiment randomly assigned the treatments among 607 teachers across 52 schools in Pakistan. The randomization at the teacher level provided advantages such as the ability to match an individual teacher to the class and to collect rich granular data such as plasma cortisol concentrations of individual teachers. However, because treated and control

¹⁴ Shortly after the rollout of the experiment, the Director of Training and Development at Progressive Education Network suggested that we also track domestic violence and stress of teachers. On behalf of our counterpart, we included measurements on stress and *physical* domestic violence, although these were not included when the experiment was registered at the AEA RCT registry with ID AEARCTR-0007465.

group teachers may interact within a school, we might have spillover effects with some of the control teachers becoming partially treated. While such spillover effects certainly exist, there are at least three arguments that support the idea that spillover effects are likely not to be large enough to completely explain our results. First, conceptually, if there are spillovers within a school with some of our control teachers who do end up getting treated, then our estimates are likely to be an underestimate of the true treatment effects. Second, our experimental design allows us to partially test for the extent of potential spillover effects: we exploit the variation in treated teachers across the 52 schools in our sample to see how it impacts our outcomes. Table B6 presents these results. We find that the treatment effect on stress and domestic violence is essentially identical as more teachers get treated within a school. That is, regardless of the fraction of treated teachers in a school, the visual narrative and self-persuasion treatment impacts stress and domestic violence similarly. Third, we leverage the fact that our Utilitarian and Malleability treatments had no impact on stress or domestic violence over the control group. Therefore, we investigate whether the fraction of schools treated with visual narrative or joint treatment causes the control group to increase in progressive gender attitudes. Under the assumption that a higher fraction of treated teachers leads to a greater likelihood for interactions between treated teachers and control teachers, we assess the impact of fraction of treated teachers on the control teacher's outcomes. However, we find little impact of fraction of treated teachers among control teachers on domestic violence or stress (Table B8 and Table B9). Taken together, the evidence strongly suggests that spillover effects between treated and control teachers, even if they exist, are likely to be small in magnitude.

Experimental Demand.—There are at least five reasons why our results are unlikely to be explained by experimental demand effects. First, the elevated concentration of cortisol observed for the treated teachers cannot be the result of experimental demand since cortisol is secreted involuntarily in response to stress. Second, we track behavioral change over a fairly long timeframe. The persistence of the effects even a year after the treatment strengthens the inference that the treatment had real impacts beyond experimental demand. Finally, we follow [Dhar et al., \(2018\)](#), who introduce the Marlowe-Crowne method, a survey module developed by social psychologists to rigorously measure a person's propensity to give socially-desirable answers ([Crowne and Marlowe, 1960](#)). In Table B10 of Appendix B, when we discard the teachers who

score high on this social desirability scale, our results are essentially identical. These features of our experiment indicate that experimenter demand is unlikely to explain our results.

Multiple Hypothesis Testing. — Our collaboration and cooperation with the Progressive Education Network allowed us access to a rich set of outcomes to investigate the impact of our treatments. Nevertheless, since we have many outcome variables, our results might be explained by false positives, we provide all our main results using Multiple Hypothesis adjusted p-values. Under the assumption that the treatments have no effect on any of our outcomes, i.e., all our null hypotheses are true, then the probability of at least one false rejection when using a critical value of 0.05 is about 80%. Consequently, in this robustness check, we adjust for the fact that we are testing for multiple hypotheses by using sharpened False Discovery Rate (FDR) q-values ([Anderson, 2008](#)). The sharpened q-values are reported in square brackets in Table 1, Table 2 and Table 3, which also shows standard p-values from our baseline regressions in parentheses for comparison. Similar results are obtained when we deploy [List et al., \(2019\)](#)'s familywise error rate correction (FWER); this extends the False Discovery Rate (FDR) method by using a bootstrapping approach, incorporating point-dependence structure of different treatments and controlling for the familywise error rate i.e., the probability of one or more false rejections. Specifically, we apply the most strident test that pools p-values across both outcomes and treatments in a single family. Similar results are obtained if we pooled outcomes into families of stress (self-reported stress and cortisol concentration) and domestic violence (victim of domestic violence and views on domestic violence). The results, reported in Table 1, Table 2 and Table 3, suggest that false positives are statistically unlikely to drive our results.

Sample Size and Randomization Inference. — Even though our sample size is about 600 teachers and 14000 students, we follow [Imbens and Rubin \(2015\)](#) suggestion to use randomization inference. That is, we scramble the data, reassign treatments, and compare the distribution of control estimates with the estimates from the experiment. The resulting p-values for 1000 iterations of this process is reported in Table B13 of Appendix B. The results are essentially similar with the treatment effects still statistically significant at conventional levels.

School-level Clustering.—In our examination of the moral bandwagoning effect, which hinges on variation (treatment intensity) across schools rather than individual teachers, a

potential concern arises regarding the potential underestimation of standard errors due to clustering at the individual teacher level, which is identical to just using the Newey-West robust standard errors. Nonetheless, it's important to note that even when we cluster the standard errors at the school level, our findings related to moral bandwagoning, although slightly less precise, remain marginally significant, with p-values of at least 0.10. Table 3 presents these results.

Any movie treatment.—Our analysis reveals significant effects for interventions associated with the film "Bol" or visual narrative treatments. Consequently, we have consolidated these two interventions into a single treatment group for comparison purposes. The control group comprises participants from three other treatments, essentially contrasting any intervention involving the "Bol" movie or visual narrative against other treatments. The outcomes, as reported in Table B11, indicate robust and statistically significant results.

Married vs Unmarried teachers.—By capitalizing on the demographic detail that half of the teachers in our study are married, we are in a position to statistically analyze the varying responses to our treatments between married versus unmarried teachers. Table B12 presents our findings, indicating that married teachers experience a more substantial increase in stress as a result of the treatment, with these differences being both larger in magnitude and statistically significant. This may be potentially due to the differing social roles, expectations and duties associated with marriage for women that may amplify an increase in stress. In contrast, when examining the impact on domestic violence, the data does not allow us to draw a statistically significant distinction between the married and unmarried teachers.

Ex-Post Addition of Stress and Domestic Violence.— Finally, we hope to clarify why stress and domestic violence were not included in the original preregistration outcomes. The original aim of the PEN experiment was to understand if we could foster more progressive gender attitudes by increasing empathy and through visual narratives, among teachers and see if such norms could be transmitted to students as reported in the companion paper [Mehmood et al., 2024](#). Before launching the experiment, we conducted three focus groups which showed no indication that our treatment may increase domestic violence or stress. However, a month after the roll-out the director training in a logistics-related meeting mentioned that she got some teachers complain about domestic violence and whether we would want to figure out a way to reduce such incidents in a future project (she did not directly link it with our treatment).

Thinking back on the sociological literature on male backlash, we worried this may be due to our gender attitudes treatments giving rise to male backlash. Therefore, following discussion with the PEN administration, the directors, and teacher representatives, we decided to measure domestic violence and stress. These were not pre-registered but were included following the experimental roll-out but before data collection. There were no additional variables included in our analysis and we report each survey statement and variables collected verbatim in the Data Appendix C of the paper. We, nevertheless, have now scheduled free mental health support by a mental health expert and connected all teachers with a helpline against domestic violence. We have also scheduled mindfulness exercises and pharma-psychological support for teachers. We hope, these measures, may at least partially offset the negative effects. We note that a priori at the roll out of the experiment, we relied on the best available evidence that fostering more progressive gender attitudes would likely have only positive not negative repercussions (see e.g. [Banerjee et al., 2019](#); [Ashraf et al., 2020](#); [Dhar et al., 2022](#); [Riley, 2022](#)).

VI. Conclusion

What hinders the progress of gender equality movements? When societal norms have multiple potential states, individuals who challenge the traditional state may encounter adverse consequences. Our investigation brought to light two central challenges for those advocating gender equality: the specter of domestic violence and stress. However, we found that the impact of stress decreases when more people share progressive gender attitudes. In other words, it's stressful to be the only one with non-conformist gender beliefs in a large group.

In much of the world, women still have worse labor market and educational opportunities, lower physical mobility, less autonomy to run for political office or to make their own decisions ([Doepke and Tertilt, 2009](#); [Duflo, 2012](#); [Fernández, 2014](#); [Fernandez and Wong, 2014](#); [Field et al., 2021](#)). This paper shows how tumultuous the transition to a new social norm can be. We implement a randomized control trial testing different methods of shifting teachers' gender attitudes. We find that teacher-training based on visual narrative and self-persuasion is effective in shifting teachers' gender attitudes; they, however, come at a cost.

Embracing more progressive gender attitudes increase levels of blood cortisol concentrations among teachers and is linked with a higher incidence of domestic violence. Our findings serve as a stark reminder that promoting progressive gender attitudes, while feasible, exacts a toll on those challenging prevailing norms. Nevertheless, some of these costs wane as societal norms converge.

References

1. Adam, E.K., Hawkley, L.C., Kudielka, B.M. and Cacioppo, J.T., 2006. Day-to-day dynamics of experience–cortisol associations in a population-based sample of older adults. *Proceedings of the National Academy of Sciences*, 103(45), pp.17058-17063.
2. Aizer, A., 2010. The gender wage gap and domestic violence. *American Economic Review*, 100(4), pp.1847-59.
3. Alan, S., Duysak, E., Kubilay, E. and Mumcu, I., 2020. Social Exclusion and Ethnic Segregation in Schools: The Role of Teacher's Ethnic Prejudice. *The Review of Economics and Statistics*, pp.1-45.
4. Alan, S., Baysan, C., Gumren, M. and Kubilay, E., 2021. Building social cohesion in ethnically mixed schools: An intervention on perspective taking. *The Quarterly Journal of Economics*, 136(4), pp.2147-2194.
5. Alesina, A., Brioschi, B. and La Ferrara, E., 2016. *Violence against women: A cross-cultural analysis for Africa* (No. w21901). National Bureau of Economic Research.
6. Anderberg, D., Rainer, H., Wadsworth, J. and Wilson, T., 2016. Unemployment and domestic violence: Theory and evidence. *The Economic Journal*, 126(597), pp.1947-1979.
7. Anderson, M.L., 2008. Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American statistical Association*, 103(484), pp.1481-1495.

8. Anderson, S. and Bidner, C., 2015. Property rights over marital transfers. *The Quarterly Journal of Economics*, 130(3), pp.1421-1484.
9. Anderson, S., 2021. Intimate partner violence and female property rights. *Nature human behaviour*, 5(8), pp.1021-1026.
10. Alsan, M., Xing, A., Wise, P., Darmstadt, G.L. and Bendavid, E., 2017. Childhood illness and the gender gap in adolescent education in low-and middle-income countries. *Pediatrics*, 140(1).
11. Angrist, J.D. and Pischke, J.S., 2014. *Mastering 'metrics: The path from cause to effect*. Princeton university press.
12. Armario, A., Marti, O., Molina, T., De Pablo, J. and Valdes, M., 1996. Acute stress markers in humans: response of plasma glucose, cortisol and prolactin to two examinations differing in the anxiety they provoke. *Psychoneuroendocrinology*, 21(1), pp.17-24.
13. Ashraf, N., Bau, N., Low, C. and McGinn, K., 2020. Negotiating a better future: How interpersonal skills facilitate intergenerational investment. *The Quarterly Journal of Economics*, 135(2), pp.1095-1151.
14. Asiedu, E., Karlan, D., Lambon-Quayefio, M. and Udry, C., 2021. A call for structured ethics appendices in social science papers. *Proceedings of the National Academy of Sciences*, 118(29), p.e2024570118.
15. Banerjee, A.V., 1992. A simple model of herd behavior. *The quarterly journal of economics*, 107(3), pp.797-817.
16. Becker, G.S., 1991. A note on restaurant pricing and other examples of social influences on price. *Journal of political economy*, 99(5), pp.1109-1116.
17. Benabou, R. and Ok, E.A., 2001. Social mobility and the demand for redistribution: the POUM hypothesis. *The Quarterly Journal of Economics*, 116(2), pp.447-487.
18. Bénabou, R., Falk, A. and Tirole, J., 2018. *Narratives, imperatives, and moral reasoning* (No. w24798). National Bureau of Economic Research.
19. Bertrand, M. and Duflo, E., 2017. Field experiments on discrimination. *Handbook of economic field experiments*, 1, pp.309-393.

20. Biasi, B., Dahl, M.S. and Moser, P., 2021. *Mental Health, Creativity, and Wealth* (No. w29422). National Bureau of Economic Research.
21. Boisjoly, J., Duncan, G.J., Kremer, M., Levy, D.M. and Eccles, J., 2006. Empathy or antipathy? The impact of diversity. *American Economic Review*, 96(5), pp.1890-1905.
22. Barker, N., Bryan, G., Karlan, D., Ofori-Atta, A. and Udry, C., 2022. Cognitive Behavioral Therapy among Ghana's Rural Poor Is Effective Regardless of Baseline Mental Distress. *American Economic Review: Insights*, 4(4), pp.527-545.
23. Bossuroy, T., Goldstein, M., Karimou, B., Karlan, D., Kazianga, H., Parienté, W., Premand, P., Thomas, C.C., Udry, C., Vaillant, J. and Wright, K.A., 2022. Tackling psychosocial and capital constraints to alleviate poverty. *Nature*, 605(7909), pp.291-297.
24. Bruhn, M. and McKenzie, D., 2009. In pursuit of balance: Randomization in practice in development field experiments. *American economic journal: applied economics*, 1(4), pp.200-232.
25. Bursztyn, L., Egorov, G. and Fiorin, S., 2020. From extreme to mainstream: The erosion of social norms. *American economic review*, 110(11), pp.3522-48.
26. Bursztyn, L., González, A.L. and Yanagizawa-Drott, D., 2020. Misperceived social norms: Women working outside the home in Saudi Arabia. *American economic review*, 110(10), pp.2997-3029.
27. Burazeri, G., Roshi, E., Jewkes, R., Jordan, S., Bjegovic, V. and Laaser, U., 2005. Factors associated with spousal physical violence in Albania: cross sectional study. *Bmj*, 331(7510), pp.197-201.
28. Castilla, C., 2022. Tipples and Quarrels in the Household: The Effect of an Alcohol Mitigation Intervention on Intimate Partner Violence.
29. Cheema, A., Khwaja, A.I., Naseer, F. and Shapiro, J.N., 2019. *Glass walls: Experimental evidence on access constraints faced by women*. Mimeo, Harvard University.
30. Cole, H.L., Mailath, G.J. and Postlewaite, A., 1992. Social norms, savings behavior, and growth. *Journal of Political economy*, 100(6), pp.1092-1125.

31. Crowne, D.P. and Marlowe, D., 1960. A new scale of social desirability independent of psychopathology. *Journal of consulting psychology*, 24(4), p.349.
32. Dhar, D., Jain, T. and Jayachandran, S., 2018. Reshaping adolescents' gender attitudes: Evidence from a school-based experiment in India (No. w25331). *National Bureau of Economic Research*. Forthcoming American Economic Review.
33. Doepke, M. and Tertilt, M., 2009. Women's Liberation: What's in it for Men?. *The Quarterly Journal of Economics*, 124(4), pp.1541-1591.
34. Doepke, M., Tertilt, M. and Voena, A., 2012. The economics and politics of women's rights. *Annu. Rev. Econ.*, 4(1), pp.339-372.
35. Duflo, E., 2012. Women empowerment and economic development. *Journal of Economic literature*, 50(4), pp.1051-79.
36. Eigen, Z.J. and Listokin, Y., 2012. Do lawyers really believe their own hype, and should they? A natural experiment. *The Journal of Legal Studies*, 41(2), pp.239-267.
37. El-Farhan, N., Rees, D.A. and Evans, C., 2017. Measuring cortisol in serum, urine and saliva—are our assays good enough?. *Annals of clinical biochemistry*, 54(3), pp.308-322.
38. Eskreis-Winkler, L., Milkman, K.L., Gromet, D.M. and Duckworth, A.L., 2019. A large-scale field experiment shows giving advice improves academic outcomes for the advisor. *Proceedings of the National Academy of Sciences*, 116(30), pp.14808-14810.
39. Fernández, R., 2014. Women's rights and development. *Journal of Economic Growth*, 19(1), pp.37-80.
40. Fernández, R. and Wong, J., 2014. Unilateral divorce, the decreasing gender gap, and married women's labor force participation. *American Economic Review*, 104(5), pp.342-47.
41. Ferracuti, S., Seri, S., Mattia, D. and Cruccu, G., 1994. Quantitative EEG modifications during the cold water pressor test: hemispheric and hand differences. *International journal of psychophysiology*, 17(3), pp.261-268.

42. Field, E., Pande, R., Rigol, N., Schaner, S. and Troyer Moore, C., 2021. On Her Own Account: How Strengthening Women's Financial Control Impacts Labor Supply and Gender Norms. *American Economic Review*, 111(7), pp.2342-75.
43. Field, E., Jayachandran, S. and Pande, R., 2010. Do traditional institutions constrain female entrepreneurship? A field experiment on business training in India. *American Economic Review*, 100(2), pp.125-129.
44. Flake, D.F., 2005. Individual, family, and community risk markers for domestic violence in Peru. *Violence against women*, 11(3), pp.353-373.
45. Fujiwara, T., Bursztyn, L. and Pallais, A., 2017. 'Acting Wife': Marriage Market Incentives and Labor Market Investments. *American Economic Review*, 107(11), pp.3288-3319.
46. Gay, V., 2023. The Intergenerational Transmission of World War I on Female Labor. Forthcoming *The Economic Journal*.
47. Giuliano, P., 2020. Gender and culture. *Oxford Review of Economic Policy*, 36(4), pp.944-961.
48. Giuliano, P. and Nunn, N., 2021. Understanding cultural persistence and change. *The Review of Economic Studies*, 88(4), pp.1541-1581.
49. Goldfarb, E.V., 2020. Participant stress in the COVID-19 era and beyond. *Nature Reviews Neuroscience*, 21(12), pp.663-664.
50. Greenwald, A.G., Poehlman, T.A., Uhlmann, E.L. and Banaji, M.R., 2009. Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of personality and social psychology*, 97(1), p.17.
51. Hazan, M., Weiss, D. and Zoabi, H., 2019. Women's liberation as a financial innovation. *The Journal of Finance*, 74(6), pp.2915-2956.
52. Haushofer, J., Mudida, R. and Shapiro, J.P., 2020. *The comparative impact of cash transfers and a psychotherapy program on psychological and economic well-being* (No. w28106). National Bureau of Economic Research.
53. Heckman, J.J. and Karapakula, G., 2019. *The Perry Preschoolers at late midlife: A study in design-specific inference* (No. w25888). National Bureau of Economic Research.
54. Hegel, G.W.F. 1820. *The philosophy of Rights*. Hackett Publishing.

55. Hellhammer, D.H., Wüst, S. and Kudielka, B.M., 2009. Cortisol as a biomarker in stress research. *Psychoneuroendocrinology*, 34(2), pp.163-171.
56. Hessler, D., Glaser, B., Dyer-Friedman, J., Blasey, C., Hastie, T., Gunnar, M. and Reiss, A.L., 2002. Cortisol and behavior in fragile X syndrome. *Psychoneuroendocrinology*, 27(7), pp.855-872.
57. Imbens, G.W. and Rubin, D.B., 2015. *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.
58. Jayachandran, S., 2015. The roots of gender inequality in developing countries. *Annual Review of Economics*, 7(1), pp.63-88.
59. Jayachandran, S., 2021. Social norms as a barrier to women's employment in developing countries. *IMF Economic Review*, 69(3), pp.576-595.
60. Kane, J.C., Elafros, M.A., Murray, S.M., Mitchell, E.M., Augustinavicius, J.L., Causevic, S. and Baral, S.D., 2019. A scoping review of health-related stigma outcomes for high-burden diseases in low-and middle-income countries. *BMC medicine*, 17(1), pp.1-40.
61. Kirschbaum, C., Pirke, K.M. and Hellhammer, D.H., 1993. The 'Trier Social Stress Test'—a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology*, 28(1-2), pp.76-81.
62. Kische, H., Ollmann, T.M., Voss, C., Hoyer, J., Rückert, F., Pieper, L., Kirschbaum, C. and Beesdo-Baum, K., 2021. Associations of saliva cortisol and hair cortisol with generalized anxiety, social anxiety, and major depressive disorder: An epidemiological cohort study in adolescents and young adults. *Psychoneuroendocrinology*, 126, p.105167.
63. Kishor, S. and Johnson, K., 2004. *Profiling domestic violence: A multi-country study*. MEASURE DHS+, ORC Macro.
64. List, J.A., Shaikh, A.M. and Xu, Y., 2019. Multiple hypothesis testing in experimental economics. *Experimental Economics*, 22(4), pp.773-793.
65. Lowe, M., 2021. Types of contact: A field experiment on collaborative and adversarial caste integration. *American Economic Review*, 111(6), pp.1807-44.

66. Lund, C., De Silva, M., Plagerson, S., Cooper, S., Chisholm, D., Das, J., Knapp, M. and Patel, V., 2011. Poverty and mental disorders: breaking the cycle in low-income and middle-income countries. *The lancet*, 378(9801), pp.1502-1514.
67. Lupien, S.J., McEwen, B.S., Gunnar, M.R. and Heim, C., 2009. Effects of stress throughout the lifespan on the brain, behaviour and cognition. *Nature reviews neuroscience*, 10(6), pp.434-445.
68. Macmillan, R. and Gartner, R., 1999. When she brings home the bacon: Labor-force participation and the risk of spousal violence against women. *Journal of Marriage and the Family*, pp.947-958.
69. Mani, A., Mullainathan, S., Shafir, E. and Zhao, J., 2013. Poverty impedes cognitive function. *Science*, 341(6149), pp.976-980.
70. Mantella, R.C., Butters, M.A., Amico, J.A., Mazumdar, S., Rollman, B.L., Begley, A.E., Reynolds, C.F. and Lenze, E.J., 2008. Salivary cortisol is associated with diagnosis and severity of late-life generalized anxiety disorder. *Psychoneuroendocrinology*, 33(6), pp.773-781.
71. Mehmood, S., Naseer, S. and Chen, D.L., 2024. Transmitting Rights: Effective Cooperation, Inter-Gender Contact, and Student Achievement. *American Economic Journal: Economic Policy*.
72. Moomal, H., Jackson, P.B., Stein, D.J., Herman, A., Myer, L., Seedat, S., Madela-Mntla, E. and Williams, D.R., 2009. Perceived discrimination and mental health disorders: the South African Stress and Health study: mental health. *South African Medical Journal*, 99(5), pp.383-389.
73. Pakistan Demographic and Health Survey, 2018. Available at: <https://dhsprogram.com/pubs/pdf/FR354/FR354.pdf> (Accessed: 20 February 2022).
74. Qin, D.D., Rizak, J., Feng, X.L., Yang, S.C., Lü, L.B., Pan, L., Yin, Y. and Hu, X.T., 2016. Prolonged secretion of cortisol as a possible mechanism underlying stress and depressive behaviour. *Scientific reports*, 6(1), pp.1-9.
75. Rao, G., 2019. Familiarity does not breed contempt: Generosity, discrimination, and diversity in Delhi schools. *American Economic Review*, 109(3), pp.774-809.

76. Rao, G., Redline, S., Schilbach, F., Schofield, H. and Toma, M., 2021. Informing sleep policy through field experiments. *Science*, 374(6567), pp.530-533.
77. Ridley, M., Rao, G., Schilbach, F. and Patel, V., 2020. Poverty, depression, and anxiety: Causal evidence and mechanisms. *Science*, 370(6522), p.eaay0214.
78. Russell, B., 1967. The Autobiography of Bertrand Russell, 3 volumes, London: George Allen and Unwin; Boston: Little Brown and Company.
79. Schilbach, F., Schofield, H. and Mullainathan, S., 2016. The psychological lives of the poor. *American Economic Review*, 106(5), pp.435-40.
80. Shiller, R.J., 1995. Conversation, information, and herd behavior. *The American economic review*, 85(2), pp.181-185.
81. Sen, A., 2014. Development as freedom (1999). *The globalization and development reader: Perspectives on development and global change*, 525.
82. Shah, M., Seager, J., Montalvo Machado, J.H. and Goldstein, M.P., 2022. *Two Sides of Gender: Sex, Power, and Adolescence* (No. 10072). The World Bank.
83. Smith, A., 1759. The theory of moral sentiments. *J. Bell, Edinburgh*.
84. Sunstein, C.R., 1996. Social norms and social roles. *Colum. L. Rev.*, 96, p.903.
85. Taylor, C., 1992, "The Politics of Recognition," in Multiculturalism: Examining the Politics of Recognition, A. Gutmann (ed.), *Princeton: Princeton University Press*, pp. 25–73.
86. Taylor, J.Y. and Holston, E.C., 2016. An exploratory study using cortisol to describe the response of incarcerated women IPV survivors to MAMBRA intervention. *Nursing Research and Practice*, 2016.
87. United Nations, 2021. What Is Domestic Abuse?. Available at: <https://www.un.org/en/coronavirus/what-is-domestic-abuse> (Accessed: 2nd February 20, 2022)
88. Wagman, J.A., Gray, R.H., Campbell, J.C., Thoma, M., Ndyababo, A., Ssekasanvu, J., Nalugoda, F., Kagaayi, J., Nakigozi, G., Serwadda, D. and Brahmabhatt, H., 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. *The Lancet Global Health*, 3(1), pp.e23-e33.

89. Yeager, D.S., Hanselman, P., Walton, G.M., Murray, J.S., Crosnoe, R., Muller, C., Tipton, E., Schneider, B., Hulleman, C.S., Hinojosa, C.P. and Paunesku, D., 2019. A national experiment reveals where a growth mindset improves achievement. *Nature*, 573(7774), pp.364-369.

Tables

Table 1: The Cost of holding Progressive Gender Attitudes

	<i>Stress Likert</i>	<i>Stress Dummy</i>	<i>Cortisol Levels</i>	<i>Standardized Cortisol</i>
	(1)	(2)	(3)	(4)
<u><i>Visual narrative (movie)</i></u>	0.306**	0.189***	0.694*	0.210*
p-value	(0.046)	p<0.001	(0.099)	(0.099)
Sharpened q-value	[0.150]	[0.001]	[0.222]	[0.222]
Romano-Wolf corrected p-value	{0.113}	{0.001}	{0.343}	{0.343}
<u><i>Joint Movie-Curriculum</i></u>	0.444***	0.231***	1.138***	0.344***
p-value	(0.005)	p<0.001	(0.008)	(0.008)
Sharpened q-value	[0.049]	[0.001]	[0.053]	[0.053]
Romano-Wolf corrected p-value	{0.015}	{0.001}	{0.020}	{0.020}
<u><i>Utilitarian</i></u>	-0.109	-0.0181	0.0206	0.006
p-value	(0.365)	(0.205)	(0.961)	(0.961)
Sharpened q-value	[0.545]	[0.352]	[0.999]	[0.999]
Romano-Wolf corrected p-value	{0.945}	{0.742}	{0.999}	{0.999}
<u><i>Malleability</i></u>	0.0331	-0.0329*	0.374	0.113
p-value	(0.817)	(0.057)	(0.362)	(0.362)
Sharpened q-value	[0.930]	[0.172]	[0.545]	[0.545]
Romano-Wolf corrected p-value	{0.999}	{0.161}	{0.945}	{0.945}
Controls and School FEs	Yes	Yes	Yes	Yes
Observations	607	607	607	607
R-squared	0.156	0.293	0.145	0.145
Mean of Dep. Variable	2.269	0.091	11.152	0.000
P-value (BM = BMC)	0.45	0.435	0.322	0.322

Note: In Column (1), the dependent variable is the answer to the question “Overall, how stressed are you?”, on a 5-point Likert scale, with one being not stressed at all and 5 being very stressed. In Column (2), the dependent variable is a response to a separate question, but this time formulated as “Are you stressed?”, with one being yes and zero being no. In Column 3 the dependent variable is the cortisol concentration in blood, measured in micrograms per deciliter, while in Column (4) we standardize the cortisol concentration in blood to mean zero and standard deviation one. Dummy variables for *Utilitarian* and *Malleability* treatments are always added as controls in the regressions. *Visual Narrative* represents the visual narrative treatment of the movie Bol augmented with the structured discussion. *Visual Narrative & Self-Persuasion* is the dummy that switches to one for participants who received the augmented visual narrative of the movie together with the gender studies curriculum. P-values computed using the Newey-West estimator are reported in parentheses, along with the multiple hypothesis-adjusted FDR q-values in square brackets and FWER-adjusted p-values in curly braces. Considering 48 hypotheses with 4 treatments X 12 outcomes. The teacher level controls include years of teaching experience, educational qualification, professional qualification, average teaching hours, class size, and marital status. School fixed effects are also included. *** p<0.01, ** p<0.05, * p<0.1.

Table 2: Impact on Domestic Violence

	Victims of Domestic Violence		Beliefs about Domestic Violence	
	(1)	(2)	(3)	(4)
<u>Visual narrative (movie)</u>	0.285**	0.273**	-0.216*	-0.205
p-value	(0.027)	(0.033)	(0.092)	(0.112)
Sharpened q-value	[0.107]	[0.118]	[0.222]	[0.243]
Romano-Wolf corrected p-value	{0.061}	{0.078}	{0.308}	{0.404}
<u>Joint Movie-Curriculum</u>	0.375***	0.357**	0.100	0.114
p-value	(0.009)	(0.014)	(0.461)	(0.401)
Sharpened q-value	[0.053]	[0.066]	[0.659]	[0.566]
Romano-Wolf corrected p-value	{0.021}	{0.029}	{0.973}	{0.945}
<u>Utilitarian</u>	-0.016	-0.017	0.126	0.132
p-value	(0.858)	(0.850)	(0.333)	(0.312)
Sharpened q-value	[0.930]	[0.930]	[0.545]	[0.539]
Romano-Wolf corrected p-value	{0.999}	{0.999}	{0.932}	{0.914}
<u>Malleability</u>	0.050	0.033	0.175	0.198
p-value	(0.584)	(0.715)	(0.206)	(0.156)
Sharpened q-value	[0.756]	[0.888]	[0.352]	[0.332]
Romano-Wolf corrected p-value	{0.994}	{0.999}	{0.742}	{0.571}
Individual controls	No	Yes	No	Yes
School Fixed Effects	Yes	Yes	Yes	Yes
Observations	607	607	607	607
R-Squared	0.093	0.101	0.123	0.130
P-value (BM = BMC)	0.6	0.629	0.017	0.016

Note: In Columns 1 and 2, the variable is an answer to the question: “Over the past year, have you been a victim of domestic violence i.e., physical violence by your husband, father or/and brother? By physical abuse, we mean hurting or trying to hurt a partner by hitting, kicking, burning, grabbing, pinching, shoving, slapping, hair-pulling, biting, denying medical care or forcing alcohol and/or drug use, or using other physical force”. The dependent variable in Columns 3 and 4 is the statement on a scale of 1 to 5 for “To what extent is domestic or physical violence by your husband, father or brother justified.” In Panel A, all dependent variables are standardized to mean zero and standard deviation one, while in Panel B, the dependent variables are in their original units. Dummy variables for *Utilitarian* and *Malleability* treatments are always added as controls in the regressions. *Visual Narrative* represents the visual narrative treatment of the movie *Bol* augmented with the structured discussion. *Visual Narrative & Self-Persuasion* is the dummy that switches to one for participants who received the augmented visual narrative of the movie together with the gender studies curriculum. In P-values computed using the Newey-West estimator are reported in parentheses, along with the multiple hypothesis-adjusted FDR q-values in square brackets and FWER-adjusted p-values in curly braces. Considering 48 hypotheses with 4 treatments X 12 outcomes. The teacher level controls include years of teaching experience, educational qualification, professional qualification, average teaching hours, class size, and marital status. *** p<0.01, ** p<0.05, * p<0.1.

Table 3: Bandwagoning Effect on Stress

	(1)	(2)	(3)	(4)
	<i>Stress Likert</i>	<i>Stress Dummy</i>	<i>Cortisol Raw</i>	<i>Standardized Cortisol</i>
<u><i>Fraction of Joint Treated Teachers X Joint Treatment</i></u>	0.848	-0.134	-6.229*	-1.882*
p-value	(0.544)	(0.739)	(0.093)	(0.093)
Sharpened q-value	[0.722]	[0.900]	[0.222]	[0.222]
Romano-Wolf corrected p-value	{0.983}	{0.999}	{0.299}	{0.299}
<u><i>Visual Narrative & Self-Persuasion</i></u>	0.219	0.267**	2.792***	0.844***
p-value	(0.611)	(0.022)	(0.004)	(0.004)
Sharpened q-value	[0.756]	[0.093]	[0.046]	[0.046]
Romano-Wolf corrected p-value	{0.994}	{0.041}	{0.005}	{0.005}
<u><i>Fraction of Movie Treated Teachers X Movie Treatment</i></u>	0.048	0.131	2.247	0.679
p-value	(0.980)	(0.844)	(0.485)	(0.485)
Sharpened q-value	[0.999]	[0.930]	[0.659]	[0.659]
Romano-Wolf corrected p-value	{0.999}	{0.999}	{0.973}	{0.973}
<u><i>Visual Narrative</i></u>	0.291	0.157	0.149	0.045
p-value	(0.529)	(0.302)	(0.867)	(0.867)
Sharpened q-value	[0.722]	[0.539]	[0.930]	[0.930]
Romano-Wolf corrected p-value	{0.980}	{0.908}	{0.999}	{0.999}
Individual Controls	Yes	Yes	Yes	Yes
School Fixed Effects	Yes	Yes	Yes	Yes
Observations	607	607	607	607
R-squared	0.157	0.293	0.151	0.151
Mean of Dep. Variable	2.269	0.091	11.152	0.000

Note: Robust standard errors appear in brackets (clustered at the school level). In Column (1), the dependent variable is the answer to the question “Overall, how stressed are you?”, on a 5-point Likert scale, with one being not stressed at all and 5 being very stressed. In Column (2), the dependent variable is a response to another question, but this time formulated as “Are you stressed?”, with one being yes and zero being no. In Column 3 the dependent variable is the cortisol concentration in blood, measured in micrograms per deciliter, while in Column (4) we standardize the cortisol concentration in blood to mean zero and standard deviation one. The *Fraction of Joint Treated Teachers* is the proportion of teachers treated with the joint *Visual Narrative & Self-Persuasion* treatment within schools. Dummy variables for *Utilitarian* and *Malleability* treatments are always added as controls in the regressions along with the interactions with Fraction of Joint Treated Teachers in the school. *Visual Narrative* represents the visual narrative treatment of the movie Bol augmented with the structured discussion. *Visual Narrative & Self-Persuasion* is the dummy that switches to one for participants who received the augmented visual narrative of the movie together with the gender studies curriculum. P-values computed using the Newey-West estimator are reported in parentheses, along with the multiple hypothesis-adjusted FDR q-values in square brackets and FWER-adjusted p-values in curly braces. Considering 48 hypotheses with 4 treatments X 12 outcomes. The teacher level controls include years of teaching experience, educational qualification, professional qualification, average teaching hours, class size, and marital status. School fixed effects are also included. *** p<0.01, ** p<0.05, * p<0.1

Online Appendix to:

Why are Rights Revolutions Rare?

By Sultan Mehmood, Shaheen Naseer and Daniel Chen

Contents

- A. Heterogeneity Analysis
- B. Additional Figures and Tables
- C. Survey Instruments, Data Description and Ethics
- D. Impact on Gender Attitudes ([Mehmood et al., 2024](#))

Appendix A: Heterogeneity

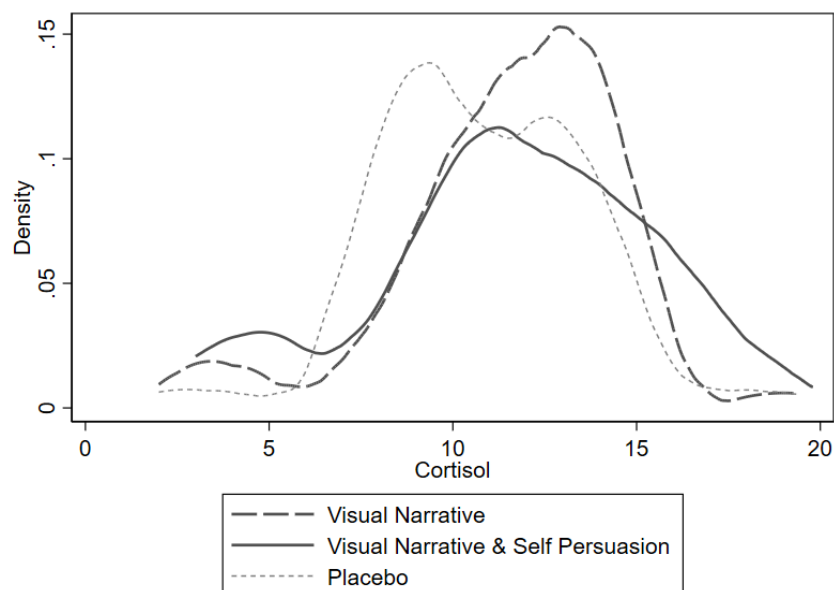
Recognition of One's Limited Rights.—A potential consequence of holding more progressive gender attitudes is the realization of one's inferior status in a hierarchy. A large and vibrant literature documents that perceived discrimination affects mental health ([Moomal et al., 2009](#); [Kane et al., 2019](#); [Rescicow et al., 2021](#)) Therefore, an alternative mechanism explaining the elevated stress may be that progressive gender attitudes come hand-in-hand with recognition of one's disadvantage in society. Our results suggest that this mechanism is not able to explain all the patterns in our data. First, we observed that as the share of treated teachers in a school increased, the impact of stress is slightly reduced (Table 3). Second, this increase in the share of teachers who were assigned to visual narrative and self-persuasion treatments within a school does not impact gender attitudes (see Table B6). Both of these patterns suggest that our results are unlikely to be driven by teachers recognizing their inferior status in the social hierarchy. This is because as more teachers are treated within a school, the effect of the treatment does not magnify the realization of one's inferior status (measured as gender attitudes and domestic violence) as documented in Table B6. It does, however, reduce stress as documented in Table 3.

Heterogeneity by Pretreatment Gender Index.— Finally, we explore whether the effect of stress and domestic violence varies by pretreatment gender attitudes. Although we do not have enough statistical power to statistically reject a homogenous effect of costs of holding progressive gender attitudes, results from Table B7 of Appendix B suggest that the costs are mostly borne out by –the norm subverters– who initially had more traditional gender attitudes. Specifically, the results indicate that teachers who had below median (or less progressive) gender attitudes appear to be most negatively affected: for the visual narrative and self-persuasion treatment group, the stress and domestic violence is about half a standard deviation higher. This hints at the fact that the costs of holding new norms are levied on those teachers who abandon their (previous) traditional gender views for more progressive ones.

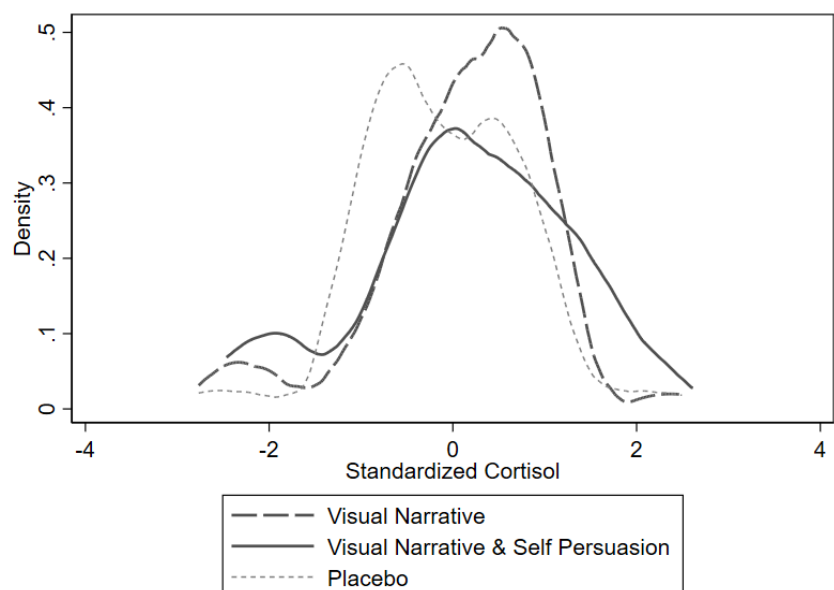
Appendix B: Additional Figures and Tables

Figure B1: Distributions of Teachers' Blood Cortisol Concentrations

Panel A: Distributions - Original Scale

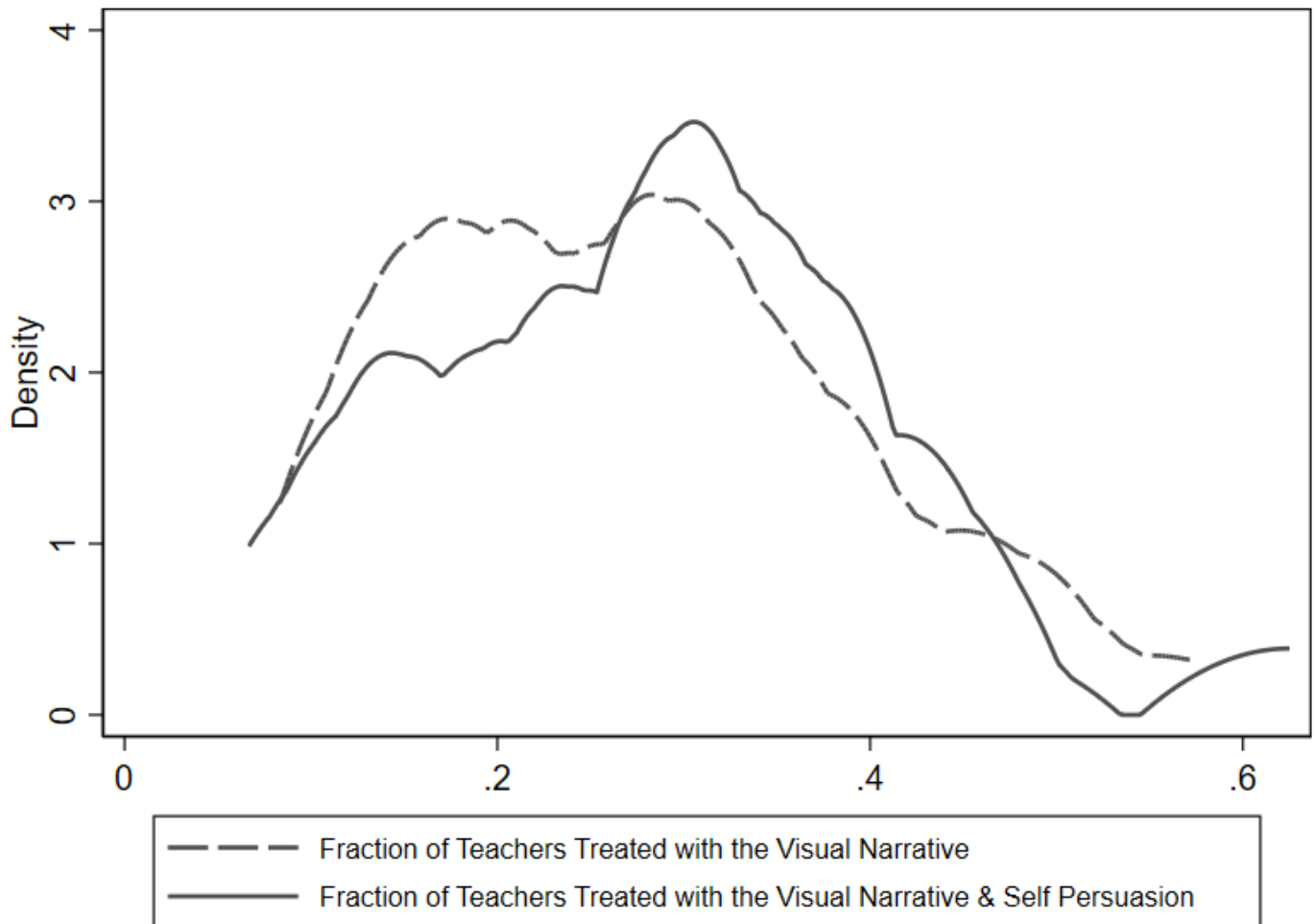


Panel B: Distributions - Standardized



Note: In Panel A, the distributions of cortisol level in blood, measured in micrograms per deciliter, using the Chemiluminescence Immunoassay (CLIA) technique are reported. In Panel B the distributions for blood cortisol readings are reported with the variable standardized to mean zero and standard deviation one. The distributions of augmented visual narrative and the visual narrative joint with self-persuasion treatments relative to the placebo groups are shown.

Figure B2: Distribution of Fraction of Teachers Treated Within Schools



Note: The figure above shows the distributions of the fraction of teachers treated within schools with the visual narrative of the movie Bol (long-dashed curve) and the fraction of teachers treated with the joint visual narrative and self-persuasion gender-rights curriculum (solid curve). The x-axis is the fraction of treated teachers schools.

Figure B3: Summary of All Studies with the Progressive Education Network

All Experiments with Teachers and Students of the Progressive Education Network (PEN)					
Title of Paper	Sample	Timing of Experiment and Treatment Rollout	Main Outcomes Observed	Treatments	Results
Transmitting Rights	607 teachers and their 13,932 pupils in 52 PEN schools in Punjab, Pakistan.	The Baseline, midline, and endline surveys were conducted in February 2021, September 2021, and March 2022, respectively. We conducted the first round of data collection 6 months after the intervention (labelled short term effects), while the second round of data was collected 12 months later. The treatment of Transmitting Rights and Rights Revolution happened at the same time but gender attitude outcomes for Transmitting Rights were collected earlier (February 2021, September 2021, and March 2022). The stress outcomes were collected later on request of PEN administration and were not pre-registered. Note all treatments were rolled out in March 2021.	Teachers' gender attitudes, IAT scores, Students' Attitudes and Math Test Scores	Utilitarian Treatment, Malleability Treatment, Visual Narrative Treatment, Joint Visual Narrative and Gender Rights Curriculum Treatment. Different practices of shifting the gender	Visual narratives and joint visual narratives combined with the curriculum foster more progressive gender attitudes among teachers. These gender attitudes are transmitted from teachers to students. Broad based interventions, such as those based on Utilitarianism and the Malleability of Empathy, have no effect.
Why Are Rights Revolutions Rare?			Stress Outcomes, Domestic Violence		A year after the intervention, the cultivation of progressive gender attitudes elevated stress hormone concentrations in the blood plasma. However, this reduces as more teachers within a school are treated and have more progressive gender attitudes, a moral bandwagoning effect.
Role Models and Theory of Mind: Teacher Vaccinations and Student Success		All treatments was rolled out in August 2021, with the baseline collected 6 month before treatment (February 2021), midline 12 months (September 2022) post treatment and endline 18 months post-treatment (March 2023), respectively. For mathematics, we have test scores for 6, 12 and 18 months after the treatment, and vaccinations and absenteeism data is available at the monthly level up to 18 months post treatment.	COVID-19 Vaccination Certificates, Mathematics Test Scores	Cash 15%, Cash 30% and Lottery Treatment, Celebrity Treatment, Role Model Treatment. Methods of encouraging people to get vaccinated	Role models positively impacts teachers' vaccination status as verified by their COVID-19 certificates. The math test scores of vaccinated teachers also rose. Reduced covid related absenteeism the likely mechanism. Monetary incentives appear not to have any effect.
Psychological Well-being and Educational Efficiency: Evidence from Civil Servant Intervention in Pakistan	850 teachers and their 24,752 students of PEN in province of Sindh and Punjab.	All treatments were rolled out in February 2022 (on going study). To monitor student performance, we utilized the February 2022 scores as the baseline, representing pre-intervention levels, and the September 2022 scores as the endline for assessing post intervention outcomes.	Stress, Test Scores, Teacher Attendance.	Cognitive Behavioral Therapy, Psychopharmacological Therapy, Mindfulness Meditation	Mindfulness Meditation (MM) and CBT effectively reduce stress, with psychopharmacological aid showing less impact.
Reshaping beliefs about ourselves and others: Experimental Evidence from civil servants in Pakistan	400 teachers and their 10,000 students of PEN in province of Sindh.	All treatments was rolled out in January 2022. The baseline survey was carried out in the second week of January, the midline survey in August, and the endline on 17th December 2022.	Growth Mindset Scores, Student Test Scores.	Growth Mindset Treatment, Narrative Treatment, Empirical Evidence Treatment	Growth mindset training shifts teachers' beliefs about malleability of intelligence, and reduces stereotypes against first generation learners and students from disadvantaged backgrounds. In contrast, exposure to narrative or empirical evidence about teacher value-added did not have statistically significant effects.

Note: The table presents a synthesis of five distinct studies carried out in collaboration with the Private Education Network (PEN) in Pakistan during the 2020s. The studies encompass various samples, timelines, outcomes, and results, which are detailed within the table. It is important to note that the samples across these studies are not identical, with the exception of two papers: "Transmitting Rights" and "Why Are Rights Revolutions Rare?" These two studies utilized the same sample of 52 PEN Schools in Pakistan's Punjab Province.

Table B1: Balance over pre-treatment stress, domestic violence and teacher Characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	<i>Pre-Treatment Stress</i>	<i>Pre-Treatment Domestic Violence</i>	<i>Pre-Treatment Gender Index</i>	<i>Married</i>	<i>Bol Movie Watched</i>	<i>Av. Teaching Hours</i>	<i>Teaching Experience</i>	<i>Years of Education</i>
<i>Visual Narrative</i>	0.0639 [0.145]	0.00533 [0.134]	-0.0720 [0.0775]	-0.0767 [0.0674]	0.0869 [0.0675]	0.418 [0.356]	0.0238 [0.400]	0.0719 [0.206]
<i>Visual Narrative & Self-Persuasion</i>	0.0942 [0.147]	0.0510 [0.136]	0.00463 [0.0789]	-0.0619 [0.0687]	0.0598 [0.0688]	0.130 [0.362]	0.481 [0.407]	0.138 [0.210]
<i>Utilitarian</i>	-0.0755 [0.118]	-0.0637 [0.111]	-0.121 [0.0839]	0.0107 [0.0654]	0.0265 [0.0667]	-0.0927 [0.251]	-0.0173 [0.343]	-0.112 [0.154]
<i>Malleability</i>	0.0472 [0.151]	-0.000256 [0.115]	-0.0429 [0.0884]	-0.0965 [0.0712]	0.0766 [0.0715]	0.132 [0.281]	0.441 [0.335]	-0.0771 [0.175]
School Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	607	607	607	607	607	607	607	607
R-squared	0.119	0.091	0.105	0.078	0.075	0.063	0.075	0.057
F Statistics (Joint Significance)	0.41	0.18	0.93	0.96	0.55	0.57	0.77	0.50
Mean of dependent var.	2.153	0.000	0.00	0.473	0.532	30.277	4.608	12.679

Note: Robust standard errors appear in brackets (clustered at the teacher level). The dependent variables are teacher level individual characteristics. Teaching Experience and years of education are teachers' experience and years of education, respectively. Av. Teaching Hours is the average number of hours the teacher teaches every week. Married is a dummy variable that switches on when the teacher is married and zero otherwise. Pre-treatment stress is measured on a 5-point Likert scale as is the dummy for pre-treatment victims of domestic violence variables. Both are measured retroactively. Bol Movie Watched is a dummy whether the teacher has previously seen the movie Bol. Dummies for *Utilitarian* and *Malleability* treatments are always added as controls in the regressions. *Visual Narrative* represents the visual narrative treatment of the movie Bol augmented with the structured discussion. *Visual Narrative & Self-Persuasion* is the dummy that switches to one for participants who received the augmented visual narrative of the movie together with the gender studies curriculum. VN is abbreviation for *Visual Narrative* and VN & SP for *Visual Narrative & Self-Persuasion*. The teacher level controls include years of teaching experience, educational qualification, professional qualification, average teaching hours, class size, and marital status. *** p<0.01, ** p<0.05, * p<0.1.

Table B2: Balance for fractions of teachers treated

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	<i>Pre-Treatment Stress</i>	<i>Pre-Treatment Domestic Violence</i>	<i>Pre-Treatment Gender Index</i>	<i>Married</i>	<i>Bol Movie Watched</i>	<i>Av. Teaching Hours</i>	<i>Teaching Experience</i>	<i>Years of Education</i>
<i>Fraction of Joint Treated Teachers X Joint Treatment</i>	0.409	0.0471	0.126	-0.0865	0.0817	-0.137	-0.247	0.216
	[0.559]	[0.0515]	[0.202]	[0.158]	[0.146]	[1.316]	[1.074]	[0.580]
<i>Fraction of Movie Treated Teachers X Movie Treatment</i>	0.167	0.0332	-0.0164	-0.326*	0.248	1.307	-1.238	0.00997
	[0.744]	[0.0528]	[0.142]	[0.187]	[0.174]	[1.293]	[0.846]	[0.710]
School Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	607	607	607	607	607	607	607	607
R-squared	0.118	0.091	0.100	0.077	0.074	0.062	0.072	0.054
Mean of dependent var.	2.153	0.000	0.00	0.473	0.532	30.277	4.608	12.679

Note: Robust standard errors appear in brackets (clustered at the school level). The dependent variables are teacher level individual characteristics. Teaching Experience and years of education are teachers' experience and years of education, respectively. Av. Teaching Hours is the average number of hours the teacher teaches every week. Married is a dummy variable that switches on when the teacher is married and zero otherwise. Pre-treatment stress is measured on a 5-point Likert scale as is the dummy for pre-treatment victims of domestic violence variables. Both are measured retroactively. Bol Movie Watched is a dummy whether the teacher has previously seen the movie Bol. The *Fraction of Joint Treated Teachers* is the proportion of teachers treated with the joint *Visual Narrative & Self-Persuasion* treatment within schools. Dummy variables for *Utilitarian* and *Malleability* treatments are always added as controls in the regressions along with the interactions with Fraction of Joint Treated Teachers in the school. *Visual Narrative* represents the visual narrative treatment of the movie Bol augmented with the structured discussion. *Visual Narrative & Self-Persuasion* is the dummy that switches to one for participants who received the augmented visual narrative of the movie together with the gender studies curriculum. The teacher level controls include years of teaching experience, educational qualification, professional qualification, average teaching hours, class size, and marital status. *** p<0.01, ** p<0.05, * p<0.1.

Table B3: Joint Orthogonality Test - Balance over Characteristics

	(1)	(2)	(3)	(4)
	<i>Visual Narrative</i>	<i>Visual Narrative & Self-Persuasion</i>	<i>Utilitarian</i>	<i>Malleability</i>
<i>Pre-Treatment Stress</i>	0.00754 [0.0159]	0.0121 [0.0157]	-0.0192 [0.0159]	0.00382 [0.0159]
<i>Pre-Treatment Domestic Violence</i>	0.000479 [0.0171]	0.00912 [0.0169]	-0.0111 [0.0171]	0.00140 [0.0172]
<i>Pre-Treatment Gender Index</i>	-0.0173 [0.0296]	0.0292 [0.0292]	-0.0444 [0.0296]	0.00348 [0.0296]
<i>Married</i>	-0.0272 [0.0340]	-0.0112 [0.0336]	0.0415 [0.0340]	-0.0383 [0.0341]
<i>Bol Movie Watched</i>	0.0277 [0.0340]	0.0109 [0.0336]	-0.0229 [0.0340]	0.0238 [0.0341]
<i>Average Teaching Hours</i>	0.00803 [0.00645]	0.000586 [0.00637]	-0.00549 [0.00644]	-0.000110 [0.00646]
<i>Teaching Experience</i>	-0.00463 [0.00592]	0.00516 [0.00585]	-0.00325 [0.00591]	0.00666 [0.00593]
<i>Years of Education</i>	0.00910 [0.0115]	0.00773 [0.0114]	-0.00734 [0.0115]	-0.0103 [0.0115]
School Fixed Effects	Yes	Yes	Yes	Yes
Observations	607	607	607	607
R-squared	0.098	0.115	0.095	0.091
F-Statistics (Joint Significance)	0.55	0.49	0.98	0.44
Mean of Dep. Variable	0.201	0.199	0.199	0.199

Note: Robust standard errors appear in brackets (clustered at the teacher level). Dummy variables that turn on for our treatments are the dependent variables. *Visual Narrative* represents the visual narrative treatment of the movie Bol augmented with the structured discussion. *Visual Narrative & Self-Persuasion* is the dummy that switches to one for participants who received the augmented visual narrative of the movie together with the gender studies curriculum. *Utilitarian* and *Malleability* are dummy variables that switch on for the Utilitarian and Malleability treatments. The teacher-level controls include years of teaching experience, educational qualification, professional qualification, average teaching hours, class size, and marital status. F-Statistic in each column correspond to joint significance test for all available baseline teacher characteristics and pretreatment variables *** p<0.01, ** p<0.05, * p<0.1

Table B4.1: All Teacher Outcomes

	H1&H3	H1&H3	H1&H3	H1&H3	H9&H10	H9&H10	H9&H10	H9&H10
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	<i>Women's Rights Overall</i>	<i>Petition to Criminalize Dowry</i>	<i>Petition to Abolish Polygamy</i>	<i>IAT Score</i>	<i>Stress Likert</i>	<i>Stress Dummy</i>	<i>Cortisol</i>	<i>Standardized Cortisol</i>
<i>Visual Narrative & Self-Persuasion</i>	0.190*** [0.0509]	0.565*** [0.143]	0.516*** [0.146]	0.341** [0.162]	0.434*** [0.158]	0.229*** [0.0408]	1.156*** [0.424]	0.349*** [0.128]
<i>Visual Narrative</i>	0.143*** [0.0508]	0.357*** [0.129]	0.350** [0.140]	0.241* [0.136]	0.294* [0.152]	0.184*** [0.0341]	0.698* [0.417]	0.211* [0.126]
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	607	607	607	527	607	607	607	607
R-squared	0.137	0.139	0.199	0.129	0.153	0.283	0.142	0.142

Note: Robust standard errors appear in brackets (clustered at the teacher level). The dependent variable in Column (1) is an index consisting of 16 gender rights statements fielded concerning Women’s Economic, Social, Legal and Political Rights. The statements can be found in online Appendix of [Mehmood et al., \(2024\)](#). In Column 2, the dependent variable is a signed petition sent to the Pakistani parliament to criminalize dowry, while the dependent variable in Column (3) is similarly standardized –to mean zero and standard deviation one – petition to abolish the law that allows polygamy for men in Pakistan. Column 4 has the dependent variable on gender Implicit Association Test (IAT). All dependent variables are standardized to mean zero and standard deviation one. In Column 5, the dependent variable is the dummy, which is 1 answer to the question “Overall, how stressed are you?” is not zero and 0 otherwise. In Column 6, the dependent variable is the answer to the question “Overall, how stressed are you?”, on a 5-point Likert scale, with one being not stressed at all and 5 being very stressed. In Column 7, the dependent variable is the cortisol concentration in blood, measured in micrograms per deciliter, using the Chemiluminescence Immunoassay (CLIA) technique, while in Column 8, we standardize the cortisol concentration in blood to mean zero and standard deviation one. The teacher level controls include years of teaching experience, educational qualification, professional qualification, average teaching hours, class size, and marital status. School fixed effects are also included *** p<0.01, ** p<0.05, * p<0.1

Table B4.2: All Student Outcomes

	H2&H4	H5&H6	H7&H8
	(1)	(2)	(3)
	<i>Attitudinal Survey</i>	<i>Math test score</i>	<i>Math test score</i>
<i>Visual Narrative & Self-Persuasion</i>	0.240*** [0.0499]	0.118*** [0.0259]	
<i>Visual Narrative</i>	0.132** [0.0526]	-0.0118 [0.0253]	
<i>i. Visual Narrative & Self-Persuasion</i>			0.0591* [0.0358]
<i>i. Visual Narrative</i>			0.116** [0.0492]
Individual Controls	Yes	Yes	Yes
School Fixed Effects	Yes	Yes	Yes
Observations	23,056	23,056	23,056
R-squared	0.046	0.598	0.599

Notes: Robust standard errors appear in brackets (clustered at the teacher level). The dependent variable in Column (1) is an index consisting of 5 gender rights statements fielded concerning gender attitudes. The statements can be found in online Appendix of [Mehmood et al. \(2024\)](#). The dependent variable in Column (2) and (3) are math test score in non mixed-gender study groups and mixed-gender study groups, respectively. All the regressions have Individual Fixed Effect and Controls, which are Gender, Class Teaching Experience, Years of Education, Average Teaching Hours, Class size, Married, Heard Bol Movie, Bol Movie Watched. In Column 3, following interactions are also included i.Mixed Class##i.U, i.Mixed Class##i.M, i.Mixed Class##i. Bol Movie Watched *** p<0.01, ** p<0.05, * p<0.1

Table B5: Is There Any Heterogeneity For Teachers That Previously Saw Bol?

	<i>Stress Likert</i>	<i>Stress Dummy</i>	<i>Victims of Domestic Violence</i>	<i>Standardized Cortisol</i>
<i>Previously Seen Bol</i>	-0.255	-0.137*	0.0297	-0.0300
<i>X Joint Treatment</i>	[0.278]	[0.0823]	[0.266]	[0.266]
<i>Previously Seen Bol</i>	-0.380	-0.176***	-0.0402	-0.175
<i>X Visual Narrative</i>	[0.274]	[0.0672]	[0.268]	[0.254]
<i>Visual Narrative & Self-Persuasion</i>	0.575**	0.302***	0.341*	0.359*
	[0.224]	[0.0645]	[0.197]	[0.192]
<i>Visual Narrative</i>	0.512**	0.284***	0.296	0.305*
	[0.209]	[0.0577]	[0.187]	[0.184]
Individual Controls and School FE	Yes	Yes	Yes	Yes
Observations	607	607	607	607
R-squared	0.160	0.309	0.102	0.146
Mean of Dep. Variable	2.269	0.091	0.000	0.000

Note: Robust standard errors appear in brackets (clustered at the teacher level). In Column (1), the dependent variable is the answer to the question “Overall, how stressed are you?”, on a 5-point Likert scale, with one being not stressed at all and 5 being very stressed. In Column (2), the dependent variable is a response to another question, but this time formulated as “Are you stressed?”, with one being yes and zero being no. In Column (3) the dependent variable is the answer to the question: “Over the past year, have you been a victim of domestic violence i.e., physical violence by your husband, father or/and brother? By physical abuse, we mean hurting or trying to hurt a partner by hitting, kicking, burning, grabbing, pinching, shoving, slapping, hair-pulling, biting, denying medical care or forcing alcohol and/or drug use, or using other physical force.”, while in Column (4) the dependent variable is the standardized cortisol concentration in blood to mean zero and standard deviation one. Dummy variables for *Utilitarian* and *Malleability* treatments are always added as controls in the regressions. The *Previously Seen Bol* is a dummy variable that switches on if the teacher is reported to have previously seen the Bol movie. *Visual narrative* represents the visual narrative treatment of the movie Bol augmented with the hour-long structured discussion of gender rights themes in the movie. *Visual Narrative & Self-Persuasion* is the dummy that switches one for participants who received the visual narrative of the movie (along with the structured discussion) together with the gender studies curriculum. The teacher level controls include years of teaching experience, educational qualification, professional qualification, average teaching hours, class size, and marital status. School fixed effects are also included. *** p<0.01, ** p<0.05, * p<0.1.

Table B6: Bandwagoning does not differentially impact Gender Attitudes or Domestic Violence

	(1)	(2)	(3)	(4)	(5)	(6)
	<i>Gender Rights Index</i>	<i>Petition to Criminalize Dowry</i>	<i>Petition to Abolish Polygamy</i>	<i>Standardized Gender IAT Score</i>	<i>Victim of Domestic Violence</i>	<i>Beliefs about Domestic Violence</i>
<i>Fraction of Joint Treated Teachers X Joint Treatment</i>	-0.204 [0.332]	0.524 [1.012]	-0.194 [0.997]	-1.355 [1.232]	0.649 [0.996]	-1.240 [0.963]
<i>Visual Narrative & Self-Persuasion</i>	0.242** [0.106]	0.426 [0.300]	0.564* [0.332]	0.693* [0.374]	0.185 [0.307]	0.443 [0.302]
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes
School Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	607	607	607	527	607	607
R-squared	0.139	0.141	0.201	0.134	0.102	0.133

Note: Robust standard errors appear in brackets (clustered at the school level). The dependent variable in Column (1) is an index consisting of 16 gender rights statements fielded concerning Women’s Economic, Social, Legal and Political Rights. The statements can be found in online Appendix of [Mehmood et al., \(2024\)](#). In Column 2, the dependent variable is a signed petition sent to the Pakistani parliament to criminalize dowry, while the dependent variable in Column (3) is similarly standardized –to mean zero and standard deviation one – petition to abolish the law that allows polygamy for men in Pakistan. Column 4 has the dependent variable on gender Implicit Association Test (IAT). All dependent variables are standardized to mean zero and standard deviation one. The *Fraction of Joint Treated Teachers* is the proportion of teachers treated with the joint Visual Narrative & Self-Persuasion treatment within schools. Dummy variables for *Utilitarian* and *Malleability* treatments are always added as controls in the regressions. *Visual Narrative* represents the visual narrative treatment of the movie Bol augmented with the structured discussion. *Visual Narrative & Self-Persuasion* is the dummy that switches to one for participants who received the augmented visual narrative of the movie together with the gender studies curriculum. The teacher level controls include years of teaching experience, educational qualification, professional qualification, average teaching hours, class size, and marital status. School fixed effects are also included. Please also note that this specification also includes all treatment dummies as in the baseline specification but these coefficients are not shown in the interest of brevity. *** p<0.01, ** p<0.05, * p<0.1

Table B7: Impact on Stress and Domestic Violence by Below and Above Median Pre-Treatment Gender Index

	(1)	(2)	(3)	(4)
	<i>Standardized Cortisol</i>		<i>Victim of Domestic Violence</i>	
	<i>Below Median Pre-Treatment Gender Index</i>	<i>Above Median Pre-Treatment Gender Index</i>	<i>Below Median Pre-Treatment Gender Index</i>	<i>Above Median Pre-Treatment Gender Index</i>
<i>Visual Narrative</i>	0.167 [0.182]	0.148 [0.212]	0.262 [0.177]	0.378* [0.210]
<i>Visual Narrative & Self-Persuasion</i>	0.482** [0.217]	0.149 [0.196]	0.516** [0.251]	0.351 [0.217]
Individual Controls	Yes	Yes	Yes	Yes
School Fixed Effects	Yes	Yes	Yes	Yes
Observations	304	301	304	301
R-squared	0.246	0.234	0.174	0.189
Mean of Dep. Variable	0.000	0.000	0.000	0.000
p-value (VN = VN & SP)	0.119	0.998	0.304	0.916

Note: Robust standard errors appear in brackets (clustered at the teacher level). In Columns 1 and 2 the dependent variable is the cortisol concentration in blood standardized mean zero and standard deviation one, while in Columns 3 and 4, the similarly standardized variable to the answer: “Over the past year, have you been a victim of domestic violence i.e., physical violence by your husband, father or/and brother? By physical abuse, we mean hurting or trying to hurt a partner by hitting, kicking, burning, grabbing, pinching, shoving, slapping, hair-pulling, biting, denying medical care or forcing alcohol and/or drug use, or using other physical force”. Columns 1 and 3 present results on a sample of teachers with pre-treatment below median gender rights index, while Columns 2 and 4 report results on a sample of teachers with pre-treatment above median gender rights index. The teacher level controls include years of teaching experience, educational qualification, professional qualification, average teaching hours, class size, and marital status. *** p<0.01, ** p<0.05, * p<0.1.

Table B8: Impact of schools more intensely treated by the Visual Narrative on Domestic Violence and Stress

Panel A: Utilitarian Treated Group				
	(1)	(2)	(3)	(4)
	<i>Victim of Domestic Violence</i>	<i>Stress Likert</i>	<i>Cortisol Raw</i>	<i>Standardized Cortisol</i>
<i>Fraction Treated with the Visual Narrative</i>	0.0390 [0.0423]	-0.0614* [0.0311]	0.129 [0.130]	0.0390 [0.0393]
Individual Controls	Yes	Yes	Yes	Yes
Observations	121	121	121	121
R-squared	0.068	0.099	0.071	0.071
Mean of Dep. Variable	0.0	2.033	10.709	-0.133
Panel B: Malleability Treated Group				
<i>Fraction Treated with the Visual Narrative</i>	-0.0199 [0.0231]	0.0257 [0.0377]	-0.00673 [0.0940]	-0.00204 [0.0284]
Individual Controls	Yes	Yes	Yes	Yes
Observations	121	121	121	121
R-squared	0.022	0.036	0.115	0.115
Mean of Dep. Variable	0.0	2.247	11.004	-0.044
Panel C: Placebo Treated Group				
<i>Fraction Treated with the Visual Narrative</i>	0.00979 [0.0124]	0.0462 [0.0340]	0.0819 [0.136]	0.0247 [0.0410]
Individual Controls	Yes	Yes	Yes	Yes
Observations	122	122	122	122
R-squared	0.063	0.115	0.035	0.035
Mean of Dep. Variable	0.0	2.073	10.835	-0.095

Note: Robust standard errors appear in brackets (clustered at the teacher level). In Column 1, the dependent variable is the answer to the question: “Over the past year, have you been a victim of domestic violence i.e., physical violence by your husband, father or/and brother? By physical abuse, we mean hurting or trying to hurt a partner by hitting, kicking, burning, grabbing, pinching, shoving, slapping, hair-pulling, biting, denying medical care or forcing alcohol and/or drug use, or using other physical force.” In Column 2, the dependent variable is the answer to the question “Overall, how stressed are you?”, on a 5-point Likert scale, with one being not stressed at all and 5 being very stressed. In Column 3, the dependent variable is the cortisol concentration in blood, measured in micrograms per deciliter, using the Chemiluminescence Immunoassay (CLIA) technique, while in Column 4, we standardize the cortisol concentration in blood to mean zero and standard deviation one. Panel A presents the analysis on the Utilitarian treated teachers, Panel B the Malleability-treated teachers and Panel C the placebo-treated group. The *Fraction of Visual Narrative Treated Teachers* is the proportion of teachers treated with the *Visual Narrative* treatment within schools. *Visual narrative* represents the visual narrative treatment of the movie Bol with a structured discussion of gender rights themes in the movie. The teacher level controls include years of teaching experience, educational qualification, professional qualification, average teaching hours, class size, and marital status. *** p<0.01, ** p<0.05, * p<0.1

Table B9: Impact of schools more intensely treated by the Visual Narrative and Self Persuasion on Teachers' Stress and Domestic Violence

Panel A: Utilitarian Treated Group				
	(1)	(2)	(3)	(4)
	<i>Victim of Domestic Violence</i>	<i>Stress Likert</i>	<i>Cortisol Raw</i>	<i>Standardized Cortisol</i>
<i>Fraction Treated with the Visual Narrative & Self-Persuasion</i>	0.265 [0.331]	-0.526 [0.752]	-2.423 [2.682]	-0.732 [0.810]
Individual Controls	Yes	Yes	Yes	Yes
Observations	121	121	121	121
R-squared	0.049	0.081	0.071	0.071
Mean of Dep. Variable	0.0	2.033	10.709	-0.133
Panel B: Malleability Treated Group				
<i>Fraction Treated with the Visual Narrative & Self-Persuasion</i>	0.154 [0.352]	0.284 [0.739]	0.310 [1.998]	0.0937 [0.604]
Individual Controls	Yes	Yes	Yes	Yes
Observations	121	121	121	121
R-squared	0.017	0.032	0.115	0.115
Mean of Dep. Variable	0.0	2.247	11.004	-0.044
Panel C: Placebo Treated Group				
<i>Fraction Treated with the Visual Narrative & Self-Persuasion</i>	-0.154 [0.214]	-1.060* [0.574]	-2.108 [1.898]	-0.637 [0.573]
Individual Controls	Yes	Yes	Yes	Yes
Observations	122	122	122	122
R-squared	0.063	0.123	0.040	0.040
Mean of Dep. Variable	0.0	2.073	10.835	-0.095

Note: Robust standard errors appear in brackets (clustered at the teacher level). In Column 1, the dependent variable is answer to the question: “Over the past year, have you been a victim of domestic violence i.e., physical violence by your husband, father or/and brother? By physical abuse, we mean hurting or trying to hurt a partner by hitting, kicking, burning, grabbing, pinching, shoving, slapping, hair-pulling, biting, denying medical care or forcing alcohol and/or drug use, or using other physical force.” In Column 2, the dependent variable is the answer to the question “Overall, how stressed are you?”, on a 5-point Likert scale, with one being not stressed at all and 5 being very stressed. In Column 3, the dependent variable is the cortisol concentration in blood, measured in micrograms per deciliter, using the Chemiluminescence Immunoassay (CLIA) technique, while in Column 4, we standardize the cortisol concentration in blood to mean zero and standard deviation one. Panel A presents the analysis on the Utilitarian treated teachers, Panel B the Malleability treated teachers and Panel C the Placebo treated group. The *Fraction of Joint Treated Teachers* is the proportion of teachers treated with the joint *Visual Narrative & Self-Persuasion* treatment within schools. *Visual Narrative & Self-Persuasion* or *Joint Treatment* is the dummy that switches to one for participants who received the visual narrative of the movie together with the gender studies curriculum. The teacher level controls include years of teaching experience, educational qualification, professional qualification, average teaching hours, class size, and marital status. *** p<0.01, ** p<0.05, * p<0.1

Table B10: Dropping the likely misreporters - Marlowe Crowne

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	<i>Victims of Domestic Violence</i>	<i>Gender Rights Index</i>	<i>Petition to Criminalize Dowry</i>	<i>Petition to Abolish Polygamy</i>	<i>Standardized Gender IAT Score</i>	<i>Stress Likert</i>	<i>Stress Dummy</i>	<i>Cortisol</i>	<i>Standardized Cortisol</i>
<i>Visual Narrative</i>	0.267*	0.171***	0.298**	0.441***	0.245*	0.273	0.182* **	0.443	0.134
	[0.161]	[0.0577]	[0.146]	[0.170]	[0.136]	[0.176]	[0.0389]	[0.502]	[0.152]
<i>Visual Narrative & Self-Persuasion</i>	0.332*	0.225***	0.493***	0.548***	0.346**	0.430**	0.225* **	1.171**	0.354**
	[0.173]	[0.0587]	[0.157]	[0.165]	[0.162]	[0.177]	[0.0430]	[0.474]	[0.143]
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	526	526	526	526	526	526	526	526	526
R-squared	0.106	0.162	0.144	0.210	0.131	0.155	0.294	0.157	0.157

Note: Robust standard errors appear in brackets (clustered at the teacher level). The dependent and independent variables in this table are identical to those in main text but in this table, we apply Marlow Crowne and discard teachers who answer yes to the following statements: 1) I am never jealous of another person's good fortune. 2) I am always a good listener. 3) I am never angry. The teacher level controls include years of teaching experience, educational qualification, professional qualification, average teaching hours, class size, and marital status. School fixed effects are also included. The gender attitudes and behavior variables from Column (1) to (4) are described in [Mehmood et al., \(2024\)](#). *** p<0.01, ** p<0.05, * p<

Table B11: Impact of Any Movie Treatment Effect

Panel A	Stress Likert	Stress Dummy	Victim of Domestic	Cortisol
<i>AnyMovie</i>	0.397*** [0.105]	0.227*** [0.0261]	0.308*** [0.0921]	0.773*** [0.291]
Observations	607	607	607	607
R-squared	0.153	0.290	0.101	0.142

Note: Robust standard errors in brackets (clustered at the school level). In this table we use a group of teachers treated with Visual Narrative and Joint Visual Narrative & Self-Persuasion as a single treatment group and Utilitarian, Malleability and Placebo as a control group. * p<0.01, p<0.05, * p<0.1

Table B12: Married vs Unmarried teachers

	(1)	(2)	(3)	(4)
Panel A: Married	Stress Likert	Stress Dummy	Victim of Domestic	Cortisol
Visual Narrative & Self-Persuasion	0.757*** [0.275]	0.332*** [0.0784]	0.241 [0.210]	0.196 [0.222]
Visual Narrative	0.444* [0.229]	0.227** [0.0876]	0.300* [0.151]	0.0582 [0.173]
Utilitarian	-0.0626 [0.179]	0.00868 [0.0216]	0.133 [0.153]	-0.103 [0.202]
Malleability	0.0915 [0.196]	-0.0401 [0.0287]	0.129 [0.134]	-0.0587 [0.214]
Observations	285	285	285	285
R-squared	0.248	0.369	0.190	0.279
Panel B: Unmarried	Stress Likert	Stress Dummy	Victim of Domestic	Cortisol
Visual Narrative & Self-Persuasion	0.127 [0.295]	0.143** [0.0603]	0.556** [0.251]	0.373* [0.187]
Visual Narrative	0.181 [0.374]	0.179** [0.0877]	0.383 [0.236]	0.339** [0.146]
Utilitarian	-0.194 [0.191]	-0.0288 [0.0204]	-0.0638 [0.168]	0.0294 [0.237]
Malleability	-0.0437 [0.284]	-0.0361 [0.0247]	0.165 [0.171]	0.194 [0.163]
P-value (Joint Treatment Panel A = Panel B)	0.0396**	0.0146**	0.2394	0.4651
P-value (Visual Narrative Panel A = Panel B)	0.3847	0.4860	0.7379	0.2582
Observations	320	320	320	320
R-squared	0.210	0.355	0.213	0.176

Note: Robust standard errors in brackets (clustered at the teacher level). Here we look at the effect of treatment on the main outcome variables separately for married and unmarried teachers. *** p<0.01, ** p<0.05, * p<0.1

Table B13: Randomization Inference

<i>Impact on Stress and Domestic Violence</i>				
	<i>Stress Likert</i>	<i>Stress Dummy</i>	<i>Standardized Cortisol</i>	<i>Victim of Domestic Violence</i>
	(1)	(2)	(3)	(4)
<i>Visual Narrative</i>	0.189 (0.001) *** {0.001} ***	0.306 (0.046) ** {0.049} **	0.210 (0.098) * {0.123}	0.273 (0.033) ** {0.046} **
<i>Visual Narrative & Self-Persuasion</i>	0.231 (0.001) *** {0.001} ***	0.444 (0.005) *** {0.003} ***	0.344 (0.008) *** {0.013} **	0.357 (0.014) ** {0.004} ***
Observations	607	607	607	607

Note: p-values from our baseline regressions appear in parentheses for comparison, while p-values from randomization inference due to Heß (2017) are reported in curly brackets. The dependent and independent variables are identical to those used in the regressions in the main text.

C. Data Appendix

Appendix C1. Consent

For teachers:

I agreed to participate in the research study. I understand the purpose and nature of this study and I am participating voluntarily. I understand that I can withdraw from the study at any time, without any penalty or consequences.

Yes ☐ No ☐

I grant permission for the data generated from this survey to be used in the researcher's publications on this topic.

Yes ☐ No ☐

I grant permission to researchers to use my anonymized information for research purposes and this includes my personal data with PEN.

Yes ☐ No ☐

For parents:

I grant permission to researchers to use my son or daughter's anonymized information for research purposes and this includes the personal data with PEN.

Yes ☐ No ☐

Statements on Stress

S17) Overall, how stressed are you? Rate from 1 to 5, with 1 not stressed at all and 5 being very stressed

S18) Are you stressed?

Yes ☐ No ☐

Statements on Domestic Violence

S19) Over the past year, have you been a victim of domestic violence i.e., physical violence by your husband, father or/and brother? By physical abuse, we mean hurting or trying to hurt a partner by hitting, kicking, burning, grabbing, pinching, shoving, slapping, hair-pulling, biting, denying medical care or forcing alcohol and/or drug use, or using other physical force.

Yes ☐ No ☐

S19') Domestic violence by husbands cannot be justified (rate from 1 to 5)

Appendix C2. Ethical considerations

This study builds upon a considerable body of evidence indicating that progressive gender norms have a positive impact on society and promote economic development (see e.g. [Duflo, 2012](#) for review). Nevertheless, we use existing data on the experiment to assess potential unintended consequences and costs associated with fostering progressive gender attitudes. This merits a discussion of ethical considerations. Although, we obtained approval from the Institutional Review Boards (IRB) from non-Western local university, nonetheless, we recognize that there are concerns that extend beyond the scope of IRB, which we will address in this note, drawing upon the framework presented in Asiedu et al. (2021).

Policy Equipoise: Theoretically it is unclear whether fostering progressive gender norms will only have positive impacts. Although, the majority of the body of literature documents positive effects ([Bertrand and Duflo, 2017](#); [Rao, 2019](#); [Alan et al., 2020](#); [Jayachandran, 2021](#)) but a small yet non-negligible research has hypothesized potential backlash effects ([Aizer, 2010](#); [Anderson, 2021](#)). Experiments have seldom attempted to measure costs, rightly due to ethical considerations, therefore, we use existing data and novel measurements to assess these backlash effects. As a result of the conflicting predictions, we consider the treatments in a state of policy equipoise.

Role of researchers with respect to the implementation: In regards to the implementation of the project, the researchers actively participated, and this information was conveyed to the subjects through verbal consent that participants signed prior to treatments. However, there may be concerns regarding the possibility of Western researchers asking the subjects to undertake activities that could harm them or the Pakistani state. To alleviate such concerns, we want to highlight that two out of three researchers involved in this project are from Pakistan. Both are natives to the province where the project was conducted. Additionally, during the implementation of the project, one member of the research team was not affiliated with a Western institution and was working locally in Pakistan.

Potential harms to participants or nonparticipants from the interventions: Before conducting this experiment, we carefully reviewed existing literature and found that the intervention, such as movies emphasizing progressive gender ideas was expected to have a positive impact on the participants (see e.g. [Ashraf et al., 2020](#); [Dhar et al., 2022](#); [Riley, 2022](#)). Therefore, we had no strong reason to anticipate that the experiment would cause harm to the participants. Furthermore, we want to highlight that the intervention, in this case, the movie, has already premiered over 10 years ago, and its impact on audiences was largely deemed positive and its critical reception was overwhelmingly positive. Additionally, there were no contemporaneous reports of heightened stress or domestic violence with the movie, which was designed by an international consortium including Johns Hopkins University. Therefore, we had no strong reason to anticipate that the experiment would cause harm to the participants. That

said, PEN suggested looking at these outcomes, which the economics literature so far has not, when it comes to studying the shaping of gender attitudes. We found that our treatment design allows examining these PEN-hypothesized harms and also a policy to address these harms through the moral bandwagoning effect.

Financial and reputational conflicts of interest: Regarding ethics issues, we want to clarify that the researchers had no financial or reputational conflicts of interest.

Intellectual freedom: We have had full intellectual freedom to report the study's results without any constraints.

Feedback to participants or communities: We provided feedback to all 607 teachers and connected them with mental health support groups and domestic violence helplines.

Foreseeable misuse of research results: We do not anticipate any misuse of the research results.

Other ethics issues to discuss: to the best of our knowledge, there are no other ethics issues that require discussion.

Appendix D. Transmitting Rights Appendix

Table D1: Impact on Teachers' Gender Attitudes

Overall	Gender Rights	Economic Rights	Political Rights	Social Rights	Legal Rights
	(1)	(2)	(3)	(4)	(5)
<u>Visual narrative (movie)</u>	0.140	0.165	0.173	0.0687	0.201
p-value	(0.0062)***	(0.0082)***	(0.1102)	(0.3226)	(0.0838)*
Sharpened q-value	[0.022]**	[0.026]**	[0.166]	[0.303]	[0.141]
Romano-Wolf corrected p-value	{0.030}**	{0.026}**	{0.5864}	{0.9880}	{0.061}*
<u>Joint Movie-Curriculum</u>	0.187	0.234	0.258	0.0434	0.315
p-value	(0.0003)***	(0.0005)***	(0.0171)**	(0.5472)	(0.0051)***
Sharpened q-value	[0.004]***	[0.004]***	[0.033]**	[0.460]	[0.022]**
Romano-Wolf corrected p-value	{0.012}**	{0.012}**	{0.054}*	{0.9770}	{0.018}**
<u>Utilitarian</u>	0.0607	0.0805	0.0783	0.0365	0.0345
p-value	(0.1725)	(0.1760)	(0.4280)	(0.5743)	(0.7113)
Sharpened q-value	[0.239]	[0.239]	[0.422]	[0.460]	[0.477]
Romano-Wolf corrected p-value	{0.4345}	{0.4166}	{0.8472}	{0.9880}	{0.9880}
<u>Malleability</u>	0.0897	0.102	0.155	0.0290	0.132
p-value	(0.0916)*	(0.1162)	(0.1836)	(0.6739)	(0.2758)
Sharpened q-value	[0.145]	[0.166]	[0.239]	[0.477]	[0.299]
Romano-Wolf corrected p-value	{0.1209}	{0.1529}	{0.7502}	{0.9880}	{0.3147}
Controls and School FEs	Yes	Yes	Yes	Yes	Yes
Observations	607	607	607	607	607
R-squared	0.138	0.118	0.097	0.102	0.125
P-value (BM = BMC)	0.388	0.318	0.445	0.725	0.360

Note: The figure summarizes our main results – effect of the treatments on the different gender attitude indices summarizing attitudes towards women. Standardization to mean zero and standard deviation are performed to each dependent variable. The treatments are compared relative to the placebo treated control group. The outcomes are recorded 12 months after the treatment. Women's Rights Overall is an average of all the statements concerning women's economic, social, legal, and political rights. Women's Economic Rights is an index combining women's rights to education and work outside home, based on reactions to statements "Women should be allowed to work outside the home". "Women and men should have equal rights to jobs". "I have no problem with my sister or female cousin working outside the home". "Daughters should have the same right to inherit property as sons". "Women and men should have equal rights to get an education". "Wives should not be less educated than their husbands". "Boys should not have more opportunities and resources for education than girls." Women's Political Rights is based on statements "It would be a good idea to elect a woman as the village Sarpanch (local politician)." "Women and men have equal rights to be President or Prime Minister." Women's Social Rights is based on statements "Domestic violence by husbands cannot be justified" "Parents should seek their daughter's consent before fixing her marriage". "A woman should not necessarily get married before her 25th Birthday". "Women who give birth to a son need not be honored in the family". "A woman with five daughters should not be under social pressure to bear a son." Finally, the Women's Legal Rights index is based on statements "Laws should be passed to ban dowry." "Under Article 35 of the Constitution of Pakistan & Judgment of Federal Shariat Court, the consent of 'Wali' is not required and a sui juris Muslim female can enter into a valid Nikah / Marriage under her own free will without the consent of Wali. To what extent do you approve of this legal right of women to enter marriage under their own free will." Equation (1) is estimated with all controls. 'Utilitarian' variable is a binary indicator that assumes a value of one upon the teacher's receipt of the corresponding treatment, analogous to the 'Malleability' treatment indicator. 'Visual Narrative (Movie)' is similarly a dummy turning on for subjects assigned the Bol Movie. The 'Joint Movie and Curriculum' indicator turning on for teachers assigned the joint Bol Movie and the Gender-rights Curriculum treatment. P-values computed using the Newey-West estimator are reported in parentheses, along with the multiple hypothesis-adjusted FDR q-values in square brackets and FWER-adjusted p-values in curly braces. Further details on this are provided in Appendix D8 (considering 36 hypotheses with 4 treatments X 9 outcomes). The teacher-level controls include years of teaching experience, educational qualification, professional qualification, average teaching hours, class size, and marital status. The student-level controls include dummies for student grade (i.e., KG, Nursery, Prep, one, two, three, four, five and six class) and pre-treatment math scores. *** p<0.01, ** p<0.05, * p<0.1. Source: [Mehmood et al., \(2024\)](#).

Table D2: Impact of Teachers' Training on Petitions to Parliament

	<i>Petition to Criminalize Dowry</i>		<i>Petition to Abolish Polygamy</i>	
	(1)	(2)	(3)	(4)
<u>Visual narrative (movie)</u>	0.104	0.115	0.0598	0.0593
p-value	(0.0113)**	(0.0055)***	(0.0134)**	(0.013)**
Sharpened q-value	[0.029]**	[0.022]**	[0.029]**	[0.029]**
Romano-Wolf corrected p-value	{0.036}**	{0.023}**	{0.036}**	{0.036}**
<u>Joint Movie-Curriculum</u>	0.173	0.180	0.0866	0.0870
p-value	(0.0002)***	(0.0001)***	(0.0004)***	(0.0005)***
Sharpened q-value	[0.004]***	[0.004]***	[0.004]***	[0.004]***
Romano-Wolf corrected p-value	{0.01}***	{0.01}***	{0.013}**	{0.013}**
<u>Utilitarian</u>	0.0083	0.0070	-0.0117	-0.0106
p-value	(0.8038)	(0.832)	(0.1972)	(0.2622)
Sharpened q-value	[0.477]	[0.479]	[0.246]	[0.296]
Romano-Wolf corrected p-value	{0.9880}	{0.988}	{0.586}	{0.724}
<u>Malleability</u>	0.0089	0.0189	-0.0055	-0.0033
p-value	(0.7959)	(0.5844)	(0.567)	(0.7514)
Sharpened q-value	[0.477]	[0.46]	[0.46]	[0.477]
Romano-Wolf corrected p-value	{0.988}	{0.987}	{0.987}	{0.988}
Individual controls	No	Yes	No	Yes
School Fixed Effects	Yes	Yes	Yes	Yes
Observations	607	607	607	607
R-Squared	0.124	0.140	0.188	0.200
Mean of dependent variable	0.114	0.114	0.030	0.030
P-value (BM = BMC)	0.185	0.206	0.433	0.409

Note: The dependent variable in Columns (1) and (2) is a dummy variable that switches on if the teacher signed a petition seeking criminalization of dowry while the dependent variable in Columns (3) and (4) is a similar dummy variable turning on for a petition seeking to abolish laws allowing polygamy in Pakistan. The outcomes are recorded 12 months after the treatment. 'Utilitarian' variable is a binary indicator that assumes a value of one upon the teacher's receipt of the corresponding treatment, analogous to the 'Malleability' treatment indicator. 'Visual Narrative (Movie)' is similarly a dummy turning on for subjects assigned the Bol Movie. The 'Joint Movie and Curriculum' indicator turning on for teachers assigned the joint Bol Movie and the Gender-rights Curriculum treatment. Each treatment is followed by a 30-minute structured discussion, the particulars of which are delineated in Table A3. The treatments are compared relative to the placebo treated control group. P-values computed using the Newey-West estimator are reported in parentheses, along with the multiple hypothesis-adjusted FDR q-values in square brackets and FWER-adjusted p-values in curly braces. Further details on this are provided in Appendix D8 (considering 36 hypotheses with 4 treatments X 9 outcomes). The teacher-level controls include years of teaching experience, educational qualification, professional qualification, average teaching hours, class size, and marital status. The student-level controls include dummies for student grade (i.e., KG, Nursery, Prep, one, two, three, four, five and six class) and pre-treatment math scores.*** p<0.01, ** p<0.05, * p<0.1. Source: [Mehmood et al., \(2024\)](#).

Table D3: Impact of Teachers' Training on Students' Attitudes and Math Test Scores

	<i>Student Attitudinal Survey</i>		<i>Maths</i>	
	(1)	(2)	(3)	(4)
<u>Visual narrative (movie)</u>	0.145	0.137	-0.00811	-0.00482
p-value	(0.0057)***	(0.0093)***	(0.8949)	(0.8433)
Sharpened q-value	[0.018]**	[0.021]**	[0.687]	[0.687]
Romano-Wolf corrected p-value	{0.001}***	{0.001}***	{0.992}	{0.992}
<u>Joint Movie-Curriculum</u>	0.254	0.245	0.160	0.119
p-value	p < 0.01***	p < 0.01***	p < 0.01***	p < 0.01***
Sharpened q-value	[0.001]***	[0.001]***	[0.018]***	[0.001]***
Romano-Wolf corrected p-value	{0.001}***	{0.001}***	{0.001}***	{0.001}***
<u>Utilitarian</u>	0.0713	0.0723	0.105	0.0293
p-value	(0.1767)	(0.1722)	(0.1055)	(0.2544)
Sharpened q-value	[0.245]	[0.245]	[0.178]	[0.318]
Romano-Wolf corrected p-value	{0.0709}*	{0.0599}*	{0.027}**	{0.1429}
<u>Malleability</u>	0.00171	-0.00252	0.00324	0.00419
p-value	(0.9776)	(0.9666)	(0.9584)	(0.8669)
Sharpened q-value	[0.687]	[0.687]	[0.687]	[0.687]
Romano-Wolf corrected p-value	{0.992}	{0.992}	{0.992}	{0.992}
Individual controls	No	Yes	No	Yes
School Fixed Effects	Yes	Yes	Yes	Yes
Observations	13,911	13,911	13,911	13,911
R-squared	0.038	0.044	0.090	0.596
P-value (BM = BMC)	0.020**	0.023**	0.007***	p < 0.01***

Note: The dependent variables are standardized to mean zero and standard deviation for mathematics test scores and student attitudinal survey. The outcomes are recorded 12 months after the treatment. The corresponding survey statements from students are reported in Appendix D4. 'Utilitarian' variable is a binary indicator that assumes a value of one upon the teacher's receipt of the corresponding treatment, analogous to the 'Malleability' treatment indicator. 'Visual Narrative (Movie)' is similarly a dummy turning on for subjects assigned the Bol Movie. The 'Joint Movie and Curriculum' indicator turning on for teachers assigned the joint Bol Movie and the Gender-rights Curriculum treatment. Each treatment is followed by a 30-minute structured discussion, the particulars of which are delineated in Table A3. The treatments are compared relative to the placebo treated control group. P-values computed using the Newey-West estimator are reported in parentheses, along with the multiple hypothesis-adjusted FDR q-values in square brackets and FWER-adjusted p-values in curly braces. Further details on this are provided in Appendix D8 (considering 16 hypotheses with 4 treatments X 4 outcomes). The teacher-level controls include years of teaching experience, educational qualification, professional qualification, average teaching hours, class size, and marital status. The student-level controls include dummies for student grade (i.e., KG, Nursery, Prep, one, two, three, four, five and six class) and pre-treatment math scores. *** p<0.01, ** p<0.05, * p<0.1. Source: [Mehmood et al., \(2024\)](#).

Table D4: Impact of Teachers' Training on Standardized Student Math Test Scores

	<i>Math Test Scores</i>				
	(1)	(2)	(3)	(4)	(5)
<i>Visual narrative (movie)</i>	-0.0384	0.0194	-0.0116	0.0106	0.0202
p-value	(0.3016)	(0.5803)	(0.7111)	(0.7533)	(0.5431)
Sharpened q-value	[0.999]	[0.999]	[0.999]	[0.999]	[0.999]
Romano-Wolf corrected p-value	{0.6084}	{0.9271}	{0.9730}	{0.9730}	{0.9041}
<i>Joint Movie-Curriculum</i>	0.164	0.0463	0.137	0.110	0.0548
p-value	p < 0.01***	(0.2258)	p < 0.01***	(0.0015)***	(0.1159)
Sharpened q-value	[0.001]***	[0.999]	[0.001]***	[0.01]***	[0.971]
Romano-Wolf corrected p-value	{0.001}***	{0.4635}	{0.001}***	{0.001}***	{0.1578}
<i>Utilitarian</i>	0.0379	0.0388	0.0256	0.0354	0.0375
p-value	(0.2782)	(0.2834)	(0.4647)	(0.3045)	(0.2972)
Sharpened q-value	[0.999]	[0.999]	[0.999]	[0.999]	[0.999]
Romano-Wolf corrected p-value	{0.5894}	{0.5894}	{0.8422}	{0.6084}	{0.6084}
<i>Malleability</i>	0.00168	-0.0204	0.0203	0.00591	0.0095
p-value	(0.9592)	(0.5847)	(0.5333)	(0.8618)	(0.7926)
Sharpened q-value	[0.999]	[0.999]	[0.999]	[0.999]	[0.999]
Romano-Wolf corrected p-value	{0.9730}	{0.9271}	{0.9041}	{0.9730}	{0.9730}
<i>Movie-Curriculum X Mixed Study</i>					0.126*** (0.0476)
<i>Mixed Study Group</i>					0.0191 (0.0310)
<i>U X Mixed Study Group</i>					-0.0147 (0.0489)
<i>M X Mixed Study Group</i>					-0.00969 (0.0475)
<i>Movie X Mixed Study Group</i>					-0.0535 (0.0487)
Students are Girls			No	Yes	
Mixed-gender Study Group Sample	Yes	No			
Individual Controls & School FE	Yes	Yes	Yes	Yes	Yes
Observations	6,959	6,952	7,107	6,804	13,911
P-value (BM = BMC)	p < 0.01***	0.485	p < 0.01***	0.005***	0.346

Note: Dependent variable is standardized to mean zero and standard deviation for math test scores. The outcomes are recorded 12 months after the treatment. The outcomes are recorded 12 months after the treatment. 'Utilitarian' variable is a binary indicator that assumes a value of one upon the teacher's receipt of the corresponding treatment, analogous to the 'Malleability' treatment indicator. 'Visual Narrative (Movie)' is similarly a dummy turning on for subjects assigned the Bol Movie. The 'Joint Movie and Curriculum' indicator turning on for teachers assigned the joint Bol Movie and the Gender-rights Curriculum treatment. Each treatment is followed by a 30-minute structured discussion, the particulars of which are delineated in Table A3. P-values computed using the Newey-West estimator are reported in parentheses, along with the multiple hypothesis-adjusted FDR q-values in square brackets and FWER-adjusted p-values in curly braces. Further details on this are provided in Appendix D8 (considering 20 hypotheses with 4 treatments X 5 outcomes). Mixed Study Group is a dummy that switches on when the student group is mixed-gender. *U X Mixed Study Group*, *M X Mixed Study Group*, *Movie X Mixed Study Group*, and *Joint Movie-Curriculum X Mixed Study Group* are interaction terms of Mixed Study Group with U, M, movie, and joint movie-curriculum treatments, respectively. The treatments are compared relative to the placebo treated control group. The teacher-level controls include years of teaching experience, educational qualification, professional qualification, average teaching hours, class size, and marital status. The student-level controls include dummies for student grade (i.e., KG, Nursery, Prep, one, two, three, four, five and six class) and pre-treatment math scores. Table C2 reports the corresponding short-term results 6 months post-treatment. *** p<0.01, ** p<0.05, * p<0.1.

Table D5: Impact of Treatments on Social Behavior Games played against Opposite Gender

	<i>Redistribution</i>	<i>Competitiveness</i>	<i>Cooperation</i>	<i>Coordination</i>
	(1)	(2)	(3)	(4)
<u><i>Movie X Mixed Study Group</i></u>	-0.0375	-0.0666	0.171	0.184
p-value	(0.5956)	(0.3455)	p < 0.01***	p < 0.01***
Sharpened q-value	[0.999]	[0.939]	[0.001]***	[0.001]***
Romano-Wolf corrected p-value	{0.9820}	{0.8472}	{0.001}***	{0.0010}***
<u><i>Movie-Curriculum X Mixed Study Group</i></u>	-0.0406	-0.0358	0.299	0.333
p-value	(0.5457)	(0.6251)	p < 0.01***	p < 0.01***
Sharpened q-value	[0.999]	[0.999]	[0.001]***	[0.001]***
Romano-Wolf corrected p-value	{0.9820}	{0.9820}	{0.0010}***	{0.0010}***
<u><i>U X Mixed Study Group</i></u>	-0.0661	-0.0219	-0.00630	-0.0168
p-value	(0.3066)	(0.7428)	(0.8673)	(0.6270)
Sharpened q-value	[0.939]	[0.999]	[0.999]	[0.999]
Romano-Wolf corrected p-value	{0.8052}	{0.9820}	{0.9820}	{0.9820}
<u><i>M X Mixed Study Group</i></u>	-0.0812	-0.0961	-0.0230	0.0122
p-value	(0.2064)	(0.1513)	(0.5444)	(0.7247)
Sharpened q-value	[0.703]	[0.571]	[0.999]	[0.999]
Romano-Wolf corrected p-value	{0.5894}	{0.4266}	{0.9820}	{0.9820}
Playing with Opposite Gender	Yes	Yes	Yes	Yes
Individual Controls & School FE	Yes	Yes	Yes	Yes
Observations	9,145	9,145	9,145	9,145
R-squared	0.008	0.013	0.610	0.331

Note: Dependent variables are outcomes on redistribution, competitiveness, cooperation, and coordination games, respectively standardized to mean zero and standard deviation one. The outcomes are recorded 12 months after the treatment. The outcomes are recorded 12 months after the treatment. 'Utilitarian' variable is a binary indicator that assumes a value of one upon the teacher's receipt of the corresponding treatment, analogous to the 'Malleability' treatment indicator. 'Visual Narrative (Movie)' is similarly a dummy turning on for subjects assigned the Bol Movie. The 'Joint Movie and Curriculum' indicator turns on for teachers assigned the joint Bol Movie and the Gender-rights Curriculum treatment. Each treatment is followed by a 30-minute structured discussion, the particulars of which are delineated in Table A3. The treatments are compared relative to the placebo treated control group. P-values computed using the Newey-West estimator are reported in parentheses, along with the multiple hypothesis-adjusted FDR q-values in square brackets and FWER-adjusted p-values in curly braces. Further details on this are provided in Appendix D8 (considering 16 hypotheses with 4 treatments X 4 outcomes). Each student plays the game twice, with the same gender and with the opposite gender. The teacher-level controls include years of teaching experience, educational qualification, professional qualification, average teaching hours, class size, and marital status. The student-level controls include dummies for student grade (i.e., KG, Nursery, Prep, one, two, three, four, five and six class) and pre-treatment math scores. *** p<0.01, ** p<0.05, * p<0.1.