



Domestic violence perception and gender stereotypes

Veronica Grembi¹ · Anna C. Rosso²  · Emilia Barili³

Received: 25 October 2022 / Accepted: 7 January 2024

© The Author(s) 2024

Abstract

Using a survey of more than 4500 Italian women, we address the link between gender stereotypes and perceptions of domestic violence. We define a new measure of stereotypes at the individual level and show that women with stronger stereotypes are less likely to state that violence is common in their area of residence and are more likely to classify physical violence as less severe than privacy breaches. This ranking is associated with a victim-blaming mindset among respondents with stronger stereotypes, who are also more likely to attribute violent behaviors to event-specific circumstances (e.g., economic distress) than to personal characteristics of abusers (e.g., psychological issues) and to advise a hypothetical victim not to react to violence.

Keywords Gender stereotypes · Perceptions of domestic violence · Intimate partner violence · Justifications for violent behaviors

JEL Classification J12 · I18

1 Introduction

Perceptions regarding domestic violence (DV) can play an important role in driving legislation to support and protect victims, also by increasing its reporting. If DV is

Responsible editor: Milena Nikolova

✉ Anna C. Rosso
annacecilia.rosso@uninsubria.it

Veronica Grembi
veronica.grembi@uniroma1.it

Emilia Barili
emilia.barili@mondino.it

¹ Sapienza, Università di Roma, Rome, Italy

² Università dell'Insubria, LdA and CEP, Varese, Italy

³ IRCSS Mondino, Pavia, Italy

believed to be uncommon, reporting is also expected to be lower when victims do experience DV, as violence is perceived to be a private matter rather than a social problem (Aizer and Dal Bó 2009).¹ The proper recording of DV episodes, which can be considered consequential to societal awareness of the problem, faces several challenges related to either objective factors (e.g., lack of economic independence and job skills among women, presence of children) or subjective factors (e.g., personal or religious beliefs, underestimation of the severity of the episode, lack of knowledge of the services available to help victims), as mentioned by Bates et al. (2019). Using a unique survey of more than 4500 female respondents in Italy in July 2020, we investigate to what extent subjective factors, as proxied by gender stereotypes, affect perceptions of DV.

We define DV as any type of violence (psychological or physical) against women, which mainly includes, in our framework, partners, although the respondents were free to think of other domestic situations in terms of both the relationship with the abuser (e.g., relatives) and the characteristics of the victim (e.g., adults), and we focus on four dimensions of DV perceptions: *Diffusion*, defined as the perception of the prevalence of different categories of violence (overall, physical, and psychological) in the respondent's province of residence; *Severity*, as the severity of violent behavior and measured with the rank assigned by the respondent to 14 hypothetical behaviors, grouped into five macro categories of social, economic, psychological, physical, and sexual violence; *Causes*, capturing potential alibis or ex post rationalizations that the respondent chooses in the face of hypothetical DV behavior, referring either to situations deemed transitory, such as economic crisis or COVID-19-related justifications, or to personality features of the abuser, such as childhood trauma or substance abuse; and finally, *Appropriate advice*, capturing the advice a respondent would give to a hypothetical DV victim, including reporting DV to the police, using services (e.g., helplines, shelters, antiviolence centers), dealing directly with the partner, or doing nothing.

Italy is an ideal case study since there is regional heterogeneity in both gender stereotypes and DV underreporting. According to a 2014 National Institute of Statistics (ISTAT) survey on DV, almost 30% of victims do not disclose their experience to anyone, 12% report the event to the police, and only 4% report it to an antiviolence center.² Consistent with what is also reported in Aizer and Dal Bó (2009), when asked about their reasons for not reporting DV, respondents cite personal beliefs as key factors; these beliefs include the belief that the respondent can solve the issue by herself (40%), the belief that DV is not a serious issue (31%), or the fear of not being believed or being shamed for the violence (7%). The large role played by personal

¹ The mechanisms affecting DV have recently received a remarkable level of interest due to the disruptive impact of the COVID-19 pandemic on households' everyday lives. However, the issue of underreporting limits conclusive estimates, and the effect of the COVID-19 pandemic on DV is still an open question (e.g., Leslie and Wilson (2020); Bullinger et al. (2020); Bhalotra et al. (2020); Sanga and McCrary (2021); Miller et al. (2020); Agüero (2021)).

² DV is a structurally underreported phenomenon (see, e.g., Morgan and Truman (2018); Morgan and Oudekerk (2019); UN Women (2020)). According to the EU Agency for Fundamental Rights (FRA 2015), while DV is widespread (20% of women reported having experienced physical/sexual violence and 43% psychological violence), only a minority of the most serious events are reported to the police (14% in the case of partner violence, 13% in the case of nonpartner violence).

beliefs calls for a deeper understanding of their roots, such as gender stereotypes. As reported by ISTAT, gender stereotypes are widely spread throughout the country; they tend to increase among older and less educated individuals, and there exists a north-south gradient, with southern regions reporting, on average, the highest levels of gender stereotypes (Istat 2019).

Although we do not claim to provide evidence of a causal relationship between stereotypes and perceptions of DV, we are able to construct an individual-level stereotype index and to isolate the link between such individual measures and the average level of stereotypes and cultural attitudes in the area of residence of the respondents, proxied by area fixed effects. In doing so, we can control for time-invariant local historical determinants of gender stereotypes that could otherwise bias our results (Alesina et al. 2013).

According to our results, stereotypes play a minor role in explaining variations in the perceived diffusion of DV, with the exception of psychological violence, a more discretionary category, yet they are strongly correlated with systematically ranking social and economic violence as more severe than physical and sexual violence. By the same token, respondents holding stronger stereotypes are more likely to justify DV as resulting from temporary situations, such as the stress associated with an economic crisis, and less likely to link violent behaviors to long-term problems of the abuser. When questioned about the type of advice that they would give a hypothetical DV victim, respondents holding stronger stereotypes are less likely to recommend formally reporting the event or using available support services and more likely to recommend addressing the problem by talking to the partner. The correlation between the availability of help services across the country and our measure of stereotypes is positive, although not significant, such that the type of advice cannot be driven by services' actual availability.³ Comparing our results to the literature is difficult, as we do not focus on the frequency of reporting of DV episodes but rather on perceptions of severity and appropriate behavior. However, we provide additional context on how to interpret our results in light of the existing evidence, in addition to some simulation exercises to better appreciate the impact of respondents being on the lower or upper part of the stereotype distribution on the outcomes of interest.⁴

Although it is difficult to completely rule out the role of unobservable factors that contemporaneously affect individual stereotypes and perceptions of DV, we address this potential limitation in different ways through a rich set of robustness checks. Among these, we control for regional fixed effects, netting out the time-invariant cultural environment and also notorious criminal cases that took place in the respondent's area of residence. Then, following the seminal work of Fernandez and Fogli (2009), by including region-of-birth fixed effects, we check whether the effect of individual stereotypes could be absorbed by the average stereotypes of the respondents' birth region. We also control for media exposure. The media propose role models that directly affect both stereotypes and perceptions (e.g., La Ferrara et al. (2012); La Ferrara (2016)), and at the same time, they are a relevant source of information on the diffusion of DV through the news they share daily. We use weekly data on exposure

³ See Table A5

⁴ Section 5.2

to television from the end of February 2020 to the end of June 2020 at the regional level by age group. We focus on this period because it overlaps with the period of stay-at-home orders issued due to the COVID-19 pandemic. Since the stay-at-home orders were particularly strict in Italy (Caselli et al., 2022), people were naturally more exposed to information via television than from other sources, although they could access other sources of information for which we do not have proper controls. Finally, in addition to adding further controls such as the time spent compiling the survey, we run both the test in Altonji et al. (2005) and the one suggested by Oster (2019). Overall, our baseline estimates remain robust.

In the additional robustness checks, we use alternative measures of individual stereotypes. Some of the questions we used to construct our preferred measure could be subject to social desirability bias. Since this implies, in our context, that respondents would tend to be less outspoken (underreport) about the gender stereotypes they truly hold, it is important to be aware of this problem to provide a correct interpretation of our results. We thus generate alternative individual measures on the basis of other questions that are less subject, in our view, to social desirability bias. While the baseline results are confirmed, these alternative measures, which do not completely rule out a social desirability effect, support the idea that our preferred measure captures a lower bound of the true link between individual stereotypes and DV perceptions. The use of principal component analysis to generate our stereotype indexes does not change the outcomes.

As a final step, we examine the mechanisms that explain why, in relation to perceptions of DV severity, respondents with stronger stereotypes classify social and economic violence as more severe than physical and sexual violence. We look at acceptable behaviors between a couple (the acceptability of an occasional slap in a relationship) and at potential mentions of victims' behavior in justifying violence (whether a rape victim can be blamed if they were drunk at the time). Stronger stereotypes are associated with greater perceived acceptability of a slap in a relationship and a higher level of blame placed on a drunk rape victim. The results, robust across specifications, appear to point to a *victim-blaming* mindset when stereotypes are stronger.

Overall, our findings point to the relevance of stereotypes as a driver of perceptions of the acceptability of violent behaviors. Our results imply that providing role models, sharing information, and changing cultural references, for instance, through targeted media campaigns, might improve awareness of the problem of DV and increase its report rate. However, our contribution does not address either the link between stereotypes and crimes against women or the direct link between stereotypes and the level of reporting of DV. Rather, we argue that perceiving a behavior as deviating from what is acceptable in a romantic relationship is the first step in dealing with DV, both formally and informally.

The paper is organized as follows. Expectations regarding the drivers of DV perceptions are listed in Section 2, while Section 4 provides detailed accounts of the outcomes of interest and the measure of stereotypes and their descriptive statistics. Section 3 provides information on our survey, and that on Section 5 introduces the empirical strategy and discusses the results and robustness checks. In Section 6, we address the victim-blaming mindset to explain some of our results. Section 7 concludes the paper.

2 Expectations on DV perceptions and survey design

Knowledge of the drivers of DV perceptions is indeed of primary importance for defining appropriate policies to facilitate broader awareness at the societal level: we could expect that a higher awareness, among other things, would be correlated with a lower incidence of the most severe cases, either because victims might look for help at the first sign of an abusive situation and when they do they are also taken seriously or because abusers might more often look for help. The individual perception of DV might be affected by multiple factors, and in what follows, we focus on 3 groups of factors: personal experience, stereotypes, and observable (individual or contextual) characteristics.

2.1 Personal experience: the victim and the abuser

Although it has been shown that the acceptability of DV is positively correlated with knowing a perpetrator but negatively correlated with knowing a victim (Gracia and Herrero 2006), controlling for a respondent's personal experiences with respect to DV is of utmost importance. However, defining personal experiences presents several challenges. In principle, personal experience could refer to either the respondent herself or her own network (i.e., friends or relatives) and to either the past or the present. For instance, the sociology literature identifies experiences in the family of origin during childhood as one of the most relevant elements affecting DV perceptions, contributing to the intergenerational transmission of violence, the so-called theory of *social learning* (Mihalic and Elliot 1997; Boyd and Richerson 1995, 2005). The logic is that since learning occurs via observation, exposure to DV in childhood is reflected in a perceived lower gravity of DV behaviors at older ages (Kernsmith 2006). However, even personal experience later in life contributes to shaping individual perceptions (Kernsmith 2006; Copp et al. 2019).

To capture the dimension of personal experience, in our survey, we proceed in two ways. In the first part of the survey, we were straightforward in asking questions that define the nature of DV according to gender (i.e., questions 15–16, specifically referred to violence against women) and that required assessing the phenomenon at the territorial level (province of residence).⁵ Then, we proxy personal experiences of DV by the variable *Know a victim*. The variable is a dummy equal to 1 if the respondent declares that she knows a victim of DV. The question was intentionally asked indirectly so that respondents did not have to identify themselves as DV victims even if they had previously been victims or were victims at the time of the survey. The rationale behind this choice was to reduce concerns about underreporting that emerge when similar questions are asked directly. DV could be underreported even in survey data, as there is a large emotional cost associated with it (Overstreet and Quinn, 2013).⁶ Consistent

⁵ Given the framework of questions 14, 15, and 16, although we cannot rule it out, we do find it quite unlikely that a respondent would identify herself as being in a child abuse situation when providing advice, and it is much more likely that she would identify as being in an adult relationship. As such, she might think of advising the parent of an abused child.

⁶ We check the level and direction of the correlation between the variable *Know a victim* and stereotypes. The correlation is positive but equal to 0.021.

with this approach, for the rest of the survey, we did not give any specific benchmark to the respondents in terms of, for instance, the identity of the potential victim they would give their advice to.⁷

By the same token, although referring to the partner in some questions (i.e., 54, 55, 62, 63), we did not provide specific attributes of the potential abuser. The identity of the potential abuser might be relevant under two dimensions: their gender and their relationship with the victim.⁸ As reported by the National Institute of Statistics (ISTAT) on the basis of official data from the *Dipartimento per le Pari Opportunità*, more than 90% of victims who call the helpline for gender violence (1522) are women (with the share rising from 93% in 2013 to 98% in 2020).⁹ Descriptive evidence based on ISTAT data on crimes according to the gender of the offender and the victim show that, in 2016, 97% of the offenders investigated for rape, 93.5% for gang rape, and 81.9% for stalking were male. If we look at the data released by the Ministry of the Interior (in charge of public safety and law enforcement in Italy) based on the 2019 reported crimes on rape and stalking, the number of reported rapes in which a female was a victim was 13 times that of crimes with a male victim, while in the case of stalking, the ratio was 68 to 1. Because murders and rapes receive much media attention and they tend to involve female victims, it is quite likely that respondents would imagine their advice going to another woman, even when the self-reference mechanism is not in place. Overall, we do not have strong enough evidence to state that the type of advice, for instance, would differ based on the gender of the abuser. It would be reasonable to think that a respondent with higher levels of stereotypes might be much less inclined to accept that a male could be a victim of DV perpetuated by a woman, exactly because of those stereotypes.

Finally, there are no public and systematic data on the relation between the abuser and the victim involved in DV. However, based on 2018 aggregated data on the distribution of murder victims according to the relationship between the victim and killer and

⁷ Our approach is consistent with one based on self-reference questions, which ask how the information personally relates to the respondent in terms of, for example, their previous experience, preferences, or identity. Previous psychological literature has shown that such questions are more likely to lead to higher engagement with (and deeper processing of) the information than with other, more generic questions that do not activate memories of personal experiences (Rogers et al. 1977; Van Raaij 1988). This evidence has also been confirmed by recent literature showing that respondents reliably encode information more effectively when it might relate to themselves (Bentley et al. 2017). Additionally, self-reference increases identification with the person mentioned in the hypothetical situation, in this case, a female victim of violence, by triggering memories from the respondent's own lives (McDonald et al. 2015). For this reason, we believe that the respondents are more likely to anchor their answers to actual behaviors and experiences.

⁸ Another element that could be taken into account in future analyses is the gender identity of the victim and the abuser and whether the relationship between the two is heterosexual or homosexual. While identifying the difference in DV perceptions between homosexual and heterosexual respondents is beyond the scope of our survey, we could expect that, in a very stereotyped context, a homosexual respondent with very weak levels of stereotypes might not trust police and official channels and would less frequently recommend the use of formal tools to fight back against DV due to distrust in institutions. In this case, if we had many respondents like this, it would mean that a correlation between higher levels of stereotypes and lower referral to formal help should be considered a lower bound of the true effect.

⁹ Data source: <https://www.istat.it/it/archivio/246557>. Statistics by the Ministry of Interior show that for the criminal offence "Abuses against family members or cohabitants," in 2020, there were a total 21,709 victims and, among others, 17,584 were female while 2,377 were underage.

to the gender of the victim, it appears that while the majority of murders involving male victims were committed by unrelated killers, the overwhelming majority of murders involving female victims were committed by partners/ex-partners, or relatives, with the majority committed by the partner/ex-partner (Parliamentary-Committee, 2020). If this is true for the entire population of murders, we could expect an even stronger incidence of these relationships in DV.

We might expect that motivation for a lower/higher perception of DV would depend on the type of relationship between the abuser and the victim, but we do not have strong expectations on the direction of the perception (less or more) based only on this relationship.

To leave the door open to different types of relationships, in questions on hypothetical advice, we mention the partner, and we also mention *personal family matters*.

2.2 Stereotypes

By implicitly defining which actions are considered appropriate by society, stereotypes may play a significant role in individual perceptions of DV (Mulla et al. 2019), both because they affect what is acceptable and because they can affect the frequency of certain behaviors. For instance, according to a bargaining model approach, when traditional gender norms are stronger, women are less empowered, which means that they have fewer credible outside options to stop violence (Aizer 2010; Tauchen et al. 1991; Farmer and Tiefenthaler 1997). Hence, stronger stereotypes lead to both more violence and more victim blaming (e.g., Anderberg et al. 2016; Tur-Prats 2019b). However, according to male backlash theory (MacMillan and Gartner 1999), women's empowerment may have the opposite effect by challenging stereotypes, causing men to feel that their breadwinner role is threatened (Dobash and Dobash 1979; Bloch and Rao 2002; Zhang and Breunig 2023). Hence, a lower degree of emancipation would lead to less violence and lower perception (Agüero 2019).

However, ex ante, the direction of the correlation between individual stereotypes and individual perceptions of DV is not obvious. In discussing interactions between women and men, Glick and Fiske (1996) classified two types of sexism: *benevolent* sexism, according to which women are fragile creatures who need men's protection, and *hostile* sexism, based on an antagonistic attitude classifying women as manipulators who seek control through seduction. Glick et al. (2000) provided evidence that in certain countries, hostile sexism is stronger among men than among women, while women are more accepting of benevolent sexism, especially in contexts in which there is more gender inequality. The more prevalent benevolent sexism is, the higher women's belief that society is fair to them. According to the logic of benevolent sexism, because men should take care of women, women are to blame for any type of violent male behavior, since violence against women should not be accepted. At the same time, stronger stereotypes (benevolent sexism) might favor a victim-blaming attitude when facing episodes of gender violence. Women who become victims of violence might be perceived as deserving it since society is believed to be fair to them and men are supposed to protect them.

González and Rodríguez-Planas (2020) recently found a strong and robust correlation between stereotypes and DV incidence and intensity, relying on the assumption that stereotypes among first- and second-generation immigrants can be effectively proxied by measures of gender equality in their countries of origin.¹⁰ The expectation is that stronger gender stereotypes might induce individuals to have lower perceptions of the severity of certain behaviors or, conditional on perceiving the behavior as inappropriate, to resort to patience and commitment to try to overcome the problem. The role of stereotypes in the reporting of violence has been the object of several studies, with ambiguous findings.

To address the link between stereotypes and the perception of DV, we use a female-only sample, as in Arenas-Arroyo et al. (2021). The approach is similar to that of Alesina et al. (2021), who investigated the perception of racial gaps in US society and decided to oversample African American respondents. While we acknowledge that the male perspective is very important, to define the societal perception of DV, focusing only on females presents several advantages, mainly due to the framing of questions. As stated in Section 2.1, self-reference questions are easier to interpret when restricting the analysis to the gender most likely to be victim of DV. At the societal level, the levels of stereotype of female and male are highly correlated (see Section 4.1), so we might expect a similar effect on the perception of DV, while a difference, if any, could have emerged regarding the effect on potential advice.

2.3 Individual and neighborhood characteristics

The last two groups of causes affecting perceptions of DV are individual and neighborhood characteristics. Among the former, education, employment status, marital status, and the presence of children are considered factors that influence the acceptability of violence but also knowledge of the options available to victims. Similarly, neighborhood characteristics influence individual learning processes: authorities in areas with higher poverty rates have a lower ability to maintain basic social control, and thus, residents in these areas are more likely to justify violent acts as a way to resolve conflicts (e.g., Button (2008)). Although, on average, neighborhood characteristics are generally found to play a minor role relative to that of individual characteristics, they still matter in our context, and we proxy them with unemployment and the teen-pregnancy rate at the province-of-residence level.

3 Our survey

Our survey was conducted by a survey company (Demetra), which administered the questionnaire and guaranteed that the sample of female respondents was representative of women aged 20–65 in the northern regions and the macroareas of central and southern Italy. We focus on the working age bracket so that we can additionally control for employment status once estimating the effects of stereotypes.

¹⁰ Beyond the economic literature, Seff (2021) presents a comprehensive discussion regarding the relation between stereotypes and DV.

The survey includes 7 sections covering the socioeconomic profile of the respondent (e.g., marital status, education level), the respondent's and her partner's (if any) working status, her beliefs regarding gender roles, her working conditions during the stay-at-home orders, her perceptions of violence, her advice to hypothetical DV victims, issues related to couple relationships and the stay-at-home orders, and her perceived severity of different DV behaviors.¹¹ Alongside the different parts of the survey, we assessed the changes due to the first wave of the COVID-19 pandemic (e.g., change in occupational status, financial distress). The first wave of the pandemic started at the end of February 2020 and ended by the beginning of June 2020, and the survey was administered from July 15 to July 31, 2020. During an overall period of ten weeks from the official start of the pandemic, very strict measures were introduced to contain the emergency, such as quarantining municipal clusters, imposing strict restrictions on people's movements, closing schools and shops, and halting industrial activities for an overall period of ten weeks.¹² This contextual situation was obviously correlated with higher levels of distress and more uncertainty about the future among households, which might have affected individual perceptions of DV. Northern regions were particularly affected by the first wave.

The questionnaire was administered online through email invitations, and respondents were invited to participate via computer-assisted web interviews (CAWIs). On average, it took 20 min to complete the 66-question survey. Each respondent received an economic incentive to answer the questionnaire—0.1 euro per minute, so for a 20-minute interview, they received approximately 2 euro. All details on the sampling strategy and response rates are explained in Online Appendix B. The final analysis was performed with a sample of 4574 women. Following the spirit of Alesina et al. (2021), who investigate perceptions of racial gaps in the US, underrepresenting the southern part of the US where the racial gaps are stronger than elsewhere, we were concerned that the representativeness of our sample was particularly accurate for the northern part of the country since such regions are traditionally considered more emancipated in terms of gender norms, with higher female involvement in the labor market. If any, we slightly oversampled the population of some regions in the north. In Table B1, we show the distribution of the female Italian population in the seven regions in the north and the macroareas of the center and the south, which is consistent with the distribution of the same population in our survey. The average age appears balanced, but for the subsample in the south, the survey population is slightly younger than the actual population. This is consistent with the method used to sample (CAWI). The daily use of the internet is more common in younger cohorts, and since in the south, the level of daily use of the internet is lower than that in the rest of the country, although levels of internet access are equal, older cohorts are slightly less represented. However, if any, this element only waters down our results, since younger cohorts hold fewer stereotypes than older cohorts (see Table B2).

We then check the representativeness of our sample using another survey run by ISTAT along a subsample of control variables that, in principle, play an important role in shaping both stereotypes and the equilibrium in adult relationships, such as

¹¹ The full survey questionnaire is available in Appendix B – Section B2.

¹² Pupils returned to school only in September 2020.

employment status, level of education and religiosity, marital status, number of children, and the dimensions of residence (i.e., indirect proxy for wealth). In Table A1, we also provide the main correlation coefficients for the controls used in the empirical analysis. As it appears, our measure of stereotypes is positively correlated with having children, being married, defining oneself as a religious person and living in larger accommodations, while it is negatively correlated with a higher level of education and being younger. In Appendix B, we provide further explanations, and in Table B3, we report the distribution of the selected variables in our sample and in the ISTAT samples. Our respondents are slightly more likely to be employed and slightly less likely to be religious, although the religiosity is differently defined in the ISTAT questions.

4 Stereotypes, outcomes, and descriptive statistics

4.1 Stereotypes

We provide a new measure of stereotypes at the individual level based on the extent to which respondents agreed or strongly agreed with 8 statements, as reported in Table A2 and in Appendix B2, question 30. The statements either imply clearly stereotypical assessments (for example, that men are better in leadership roles), borrowed from the World Values Survey (WVS), or simply capture the burden of some situations experienced by women, such as the distress faced by women who earn more than their male partner. We construct an index based on the z score of the responses to the statements; for each respondent, the z score is equal to the average of all the standardized replies. This index has a zero average by construction and a standard deviation of 0.532.

The advantage of our index of stereotypes over others available only at the area-of-residence level (e.g., the regional level) is that it allows more variation within and between geographical areas. We externally validate our index with two measures of stereotypes. The first is based on a 2018 survey on gender violence and stereotypes conducted by ISTAT.¹³ ISTAT released aggregated regional data; we exploit the regional variation in the share of respondents who agreed or strongly agreed with three statements, as explained in Table A2. As we do for our main index, we create two z score indexes based on the standardized regional shares, one including the share of female respondents only (*ISTAT females*) and one including both female and male respondents (*ISTAT all*). As a second measure of stereotypes, we use regional-level data from the 2018 Labor Force Survey, run by ISTAT, and construct a ratio of the *employment rate* of women relative to that of men, with both rates measured for the age group 54 and older. Women generally have higher employment rates at the beginning of their working careers but tend to leave the labor force after having children. Higher values for this ratio represent higher levels of gender equality in labor market participation, which we use as a proxy for lower stereotypes.

¹³ This national survey also borrowed several questions from the WVS. The correlation between the stereotype index of female respondents and that of male respondents is equal to 0.85.

We aggregate our measure of individual-level stereotypes at the regional level to increase the comparability of the different indexes. The correlations among them are strong and in the expected direction: as shown in Table A3, there is a positive correlation with the indexes created by ISTAT and a negative correlation with the employment ratio.¹⁴ By decomposing the correlations by macroareas (north vs. center and south), we can observe that the correlations between the stereotypes in our sample and the two indexes from the ISTAT surveys are even stronger, especially in the north.¹⁵

Finally, we use the different available measures of stereotypes to check whether they are negatively correlated with the presence of antiviolence services. This is an important check since both a low perception of DV and a potential low propensity to advise for the use of formal help channels might be a function of a low presence of services. The results of a simple regression in Table A5 show that the direction of the correlation between the number of antiviolence centers and stereotypes indexes is positive and not negative. However, as shown in Table A6, respondents with higher stereotypes are less likely to be aware of the existence of support services, which, however, are not less likely to be present in their residence area. In the survey, we include questions on the knowledge of the helpline number, which was highly advertised during the stay-at-home periods. We then estimate the impact of stereotype on an outcome, which is 1 when the respondent is aware of the existence of any of the following three services: helplines, antiviolence centers, and shelters. The results, cleaned for the regional fixed effects, show that women with above-average levels of stereotypes are 4 percentage points less aware of any types of antiviolence services (a 4% of the baseline average awareness of 86%).

4.2 Outcomes

We define the variable on the perceived diffusion of DV in the respondent's province of residence with three dummies: *Overall Violence*, *Physical violence*, and *Psychological violence*. Each dummy is equal to 1 when the respondent reports that she considers DV (overall, physical, or psychological) to be widespread in her province of residence and is equal to 0 otherwise. The questions specifically refer to violence against women.

The measure of the perceived severity of violent behaviors is the result of an exercise that respondents had to perform: ranking a list of 14 given behaviors from the least (1) to the most (14) severe.¹⁶ Given the design of the question (*i.e.*, respondents were asked to rank behaviors by degree of severity), we expect that choosing a lower rank for a specific behavior mechanically implies imposing a higher rank on another behavior. Overall, the listed behaviors reflect 5 dimensions of DV: *Social* (*i.e.*, control over social media, control over friends), *Economic* (*i.e.*, control over work, control over

¹⁴ The same holds at the individual level, but the magnitude of the correlations is, as expected, much smaller because of the large within-region variation in the individual-level measure of stereotypes.

¹⁵ We also regress the individual-level stereotypes on the three measures. In the regression, we control for the same variables we use in our baseline model, specifically personal experience, individual characteristics, and neighborhood characteristics. The results are available in Table A4, confirming the correlations.

¹⁶ The suggested behaviors that are then grouped for the construction of this outcome are reported in Appendix B2, Question 64.

expenses), *Psychological* (i.e., quarrels, threats, insults, humiliation, recrimination, spite), *Physical* (i.e., throwing objects, battering), and *Sexual* (i.e., sex requests, sexual violence). At the individual level i , the score assigned to each dimension ($Dimension_i$) is defined as the average individual rank ($rank_{ni}$) assigned to each of the N relevant behaviors associated with that dimension, as in Eq. 1.

$$Dimension_i = \frac{1}{N} \sum_{n=1}^N rank_{ni} \quad (1)$$

By construction, a respondent's rank assigned to each behavior, $rank_{ni}$, is a discrete measure that ranges from 1 to 14. For instance, if respondent A ranks *throwing objects* as 6 and *battering* as 10, the score assigned to physical violence ($Physical_i$), which by definition combines the two, is equal to 8 (the average of the individual ranks, $rank_{ni}$). Figure A1 graphically presents the individual- and provincial-level ranks for each dimension, while Fig. A2 reports the specific rank of each of the 14 behaviors. There is a marked difference between the scores at the individual level and the scores at the provincial level. Individual-level scores exhibit only marginal differences in the different dimensions of violence, while the provincial-level scores exhibit a clear expected gradient in the perceived severity of violent behaviors, with the social and economic dimensions placed at the bottom of the distribution and psychological, physical, and sexual violence at the top. These graphs also show a large dispersion of the perceptions of severity at the individual level, which is averaged out at the provincial level.

Causes of violent behaviors are defined as potential justifications or rationales used to explain hypothetical DV episodes. With respect to the 14 previously listed violent behaviors, respondents are asked to choose the most likely explanations among a preset list of potential causes. The causes are grouped into three categories: COVID-19-related factors (i.e., limitations on social life, inadequate living space, inadequate access to IT devices), economic distress not necessarily due to the COVID-19 crisis (i.e., stress due to work uncertainty, financial problems within the relationship), and factors associated with less temporary situations and mostly related to the personality of the abuser (i.e., issues balancing family and work, addiction to or abuse of toxic substances, a need to feel superior to one's partner, and negative experiences as a child). The underlying purpose of asking this question is to differentiate between justifications associated with temporary circumstances and those associated with long-term problems that should not be considered situation-specific. To facilitate interpretation, both of the outcome variables on the rankings of behaviors and justification scores are defined as dummies that take a value of 1 when the index is above the sample median of each distribution.

Finally, we analyze the respondent's perceptions of what constitutes appropriate advice for a DV victim.¹⁷ The responses are grouped into four categories: *Reporting* (reporting the incident to the police), *Use of Services* (referring the victim to an anti-

¹⁷ Consistent with what we have already stated, these could be actions the respondent herself took as a victim of DV if she refers to herself as a DV victim. We do not refer to a single episode of violence, which enables respondents to consider their entire experience.

iolence center, calling the 1522 helpline, or seeking other services), *Partner* (talking with or leaving the partner) and *No reaction* (offering no advice, offering no interference, or dismissing the concern). Importantly, we are not evaluating the effectiveness of alternative types of advice but comparing the drivers of the respondents' preferences for each of the suggested options. We consider hypothetical advice a potential behavior based on the perception of DV.

Detailed outcome definitions descriptive statistics are available in Table A7, while in Table 1, we compare the mean outcomes for individuals with weak levels of stereotypes (*Weak Stereotypes*), and strong levels of stereotypes (*Strong Stereotypes*) and we calculate the p -value of the difference between the former and the latter.¹⁸ The direction of the difference is positive for the measures of perception of diffusion, which means that respondents with weak levels of stereotypes perceive the level of violence to be higher, especially psychological violence. By the same token, respondents with weak levels of stereotypes perceive the severity of physical and sexual violence to be higher (positive difference-see column 3), and they tend to see psychological problems of the abuser as explanations for their violent behaviors (*Alibi abuser*). Finally, respondents with weak levels of stereotypes recommend more formal tools to deal with violence (i.e., *Reporting* and *Use of Services*).

4.3 Controls

Table A8 reports a detailed definition of each variable used in our analysis and presents the descriptive statistics of the sample. Consistent with the expectations set in Section 2, controls are grouped into 4 categories: personal experience, socioeconomic controls, COVID-19-related controls, and neighborhood characteristics.

Average statistics show that 38.8% of respondents declared that they know a victim, a percentage in line with the estimated incidence of domestic violence based on other surveys (FRA 2015). When asked about knowing a victim experiencing DV during the health emergency, 10% of the sample answered affirmatively. Respondents are mainly Italian nationals, 47% were married, 60% had at least one child, and 61% self-identified as religious. They are almost uniformly distributed across 4 age groups, with 31% younger than 35 and 19% older than 55. A total of 37% live in homes of more than 100 sq. m. (which captures both the level of wealth and the severity of constraints imposed during the strict stay-at-home orders). A total of 78% of respondents have at least a high school degree, and 35% declared that they were not employed in February 2020 (before the first wave of the COVID-19 pandemic). A total of 76% of the respondents considered it highly important to comply with the stay-at-home orders. Almost 30% had struggled with serious financial problems due to the pandemic, 34% feared that they would lose their job because of the pandemic, and in 17% of the cases, the respondent experienced distress due to cohabitation because of the restrictions.

Finally, since considering DV a private matter (e.g., refusing the involvement of other people in a relationship) is among the reasons a victim might not report it to the authorities, we also use a question about respondents' networks. We asked about

¹⁸ Individuals with weak level of stereotypes are those below the sample median of our index, while strong levels are above the sample median.

Table 1 Outcomes averages by level of stereotypes

Outcome	Weak stereotypes (1)	Strong stereotypes (2)	Difference (3)
<i>Perception of diffusion:</i>			
Overall Violence	0.259	0.219	0.040***
Physical Violence	0.098	0.101	-0.002
Psychological Violence	0.245	0.199	0.045***
<i>Perception of severity (behaviors):</i>			
Social	0.434	0.496	-0.062***
Economic	0.475	0.521	-0.046***
Psychological	0.460	0.487	-0.026*
Physical	0.511	0.423	0.088***
Sexual	0.504	0.409	0.095***
<i>Perception of causes (alibis):</i>			
Alibi COVID19	0.456	0.518	-0.062***
Alibi crisis	0.471	0.504	-0.033**
Alibi abuser	0.528	0.405	0.123***
<i>Perception of appropriate advice:</i>			
Reporting	0.544	0.495	0.049***
Use of Services	0.822	0.788	0.034***
Partner	0.284	0.324	-0.040***
No reaction	0.036	0.077	-0.041***

Notes: T-test comparing the outcome averages by level of stereotypes. Stereotypes are classified into two categories based on the sample median: *Weak stereotypes* (below the sample median) and *Strong stereotypes* (above the sample median). Significance levels: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

the type of network that respondents rely on when they have to share positive (e.g., promotions or good news from their personal life) or negative (e.g., personal or work-related struggles) events in their lives. Based on their answers, we identify three types of networks: a family network composed of the partner and the family of origin, a formal network that includes professional figures such as general practitioners (GPs) or therapists, and an informal network that includes friends and coworkers. Note that respondents could choose up to 3 options for each type of event, including the option “no one.” Therefore, the indexes are not mutually exclusive. Additionally, in this case, to aid in the interpretation of the results, the network indexes are defined with dummies: each dummy is equal to 1 if the respondent has an above-median level of the specific network support.

5 Empirical analysis

5.1 Empirical strategy

As described in Section 2, we aim to analyze how individual stereotypes (S_i) correlate with DV perceptions (P_i), controlling for other potential confounders such

as personal experience (E_i), individual characteristics (L_i), and neighborhood characteristics (N_p). We estimate Eq. 2 using a linear probability model. The additional inclusion of area-level fixed effects (τ_a) allows us to control for broader time-invariant, spatial-specific characteristics (e.g., cultural context, trust in institutions, provision of services). We alternatively consider the 5 Italian macroareas (i.e., northeast, northwest, center, south, and islands) that group together more regions defined using ISTAT's official definition or the 21 regions (i.e., 19 regions and 2 autonomous provinces) as geographic units. Standard errors are clustered at the corresponding area level.¹⁹ Conditional on the fixed effects (τ_a), the estimation compares individuals within homogeneous geographic areas, leveraging on differences across individuals.

$$P_i = \alpha + \beta_1 S_i + \beta_2 E_i + \beta_3 L_i + \beta_4 N_p + \tau_a + \epsilon_i \quad (2)$$

We sharpen our empirical analysis through robustness checks (Section 5.3).

5.2 Results

The baseline results are presented in Table 2. In *Panel A*, we report the results using macroarea fixed effects, while in *Panel B*, we use a robust specification incorporating regional fixed effects. The results are robust across panels, meaning that the effect of stereotypes does not depend on the differences in the average level of gender stereotypes across regions but is driven by within-region variations in individual beliefs about stereotypes. This result represents an improvement on the literature, which has thus far considered only aggregate measures of stereotypes, without exploiting individual-level variation *within* regions. Since *Panel B* provides the robust version of the baseline model, our preferred specification is the one including region-of-residence fixed effects.

Stereotypes have a minor association with *Diffusion*, significantly affecting the perceived diffusion only of psychological violence (-6.5% at the mean of the outcome when stereotypes increase by 1 standard deviation). The effect goes in the same direction for perceived diffusion overall, even though it is not significantly different from zero. Such a path seems to indicate that subscribing to strong stereotypes reduces awareness regarding the diffusion of DV, particularly psychological violence, the definition of which can be highly subjective.

Stereotypes are strongly correlated with *Severity*, reducing the perceived severity of physical and sexual violence (-4.4% and -6.5% when stereotypes increase by 1 SD) and increasing that of social and economic violence ($+5.6\%$ and $+2.6\%$). Although these effects are partially mechanical, the results systematically highlight that individuals with above-average levels of stereotypes assign a lower level of severity to physical and sexual violence and a higher level of severity to social and economic violence than individuals with lower levels of stereotypes. This means that stronger stereotypes are correlated with a higher perceived acceptability of physical and sexual violence.

¹⁹ Notably, macroareas are statistical geographic units, while regions are administrative units with their own political administration responsible for the provision of several public services, including social and health services.

Table 2 Baseline estimation

Diffusion		Severity		Causes		Appropriate advice									
Overall	Physical Psych	Social	Economic Psych	Physical	Sexual	COVID-19 Crisis	Abuser	Reporting Service	Partner	No reaction					
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	
<i>Panel A: Macroarea fixed effects:</i>															
Stereotypes	-0.024	0.005	-0.026	0.050***	0.024*	0.011	-0.038**	-0.056***	0.057***	0.046***	-0.072***	-0.061***	-0.027*	0.044***	0.037**
	(0.016)	(0.017)	(0.018)	(0.010)	(0.009)	(0.020)	(0.010)	(0.006)	(0.012)	(0.007)	(0.015)	(0.012)	(0.011)	(0.009)	(0.009)
<i>Panel B: Region fixed effects:</i>															
Stereotypes	-0.025	0.005	-0.027*	0.049***	0.024*	0.013	-0.039***	-0.056***	0.057***	0.047***	-0.072***	-0.061***	-0.027***	0.043***	0.037***
	(0.016)	(0.014)	(0.015)	(0.015)	(0.014)	(0.015)	(0.008)	(0.013)	(0.015)	(0.011)	(0.011)	(0.017)	(0.009)	(0.009)	(0.009)
Mean Out	0.239	0.099	0.222	0.465	0.498	0.474	0.467	0.457	0.487	0.488	0.467	0.320	0.805	0.304	0.056
N obs	4573	4573	4573	4573	4573	4573	4573	4573	4573	4573	4573	4573	4573	4573	4573

Notes: Standard errors are reported in parentheses and clustered by macroarea (Panel A) or region (Panel B–C). Regressions in Panel A include macroarea fixed effects, and regressions in Panel B include region-fixed effects. All regressions include controls for experience, life status, and neighborhood. Significance levels: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Similarly, regarding *Causes*, stronger stereotypes are negatively correlated with attribution of the causes of DV to longstanding characteristics of the abuser (e.g., childhood trauma) and positively correlated with citing context-specific causes. Respondents with stronger levels of stereotypes are more likely to justify violent behaviors during transitory shocks, such as the COVID-19 emergency. A 1 SD increase in the stereotype index accounts for a +6.2%, +5.1%, and -8.2% change in perceptions of DV as being caused by COVID-19-, economic crisis- and abuser personality-related factors, respectively.

As far as *Appropriate advice* is concerned, stronger stereotypes are correlated with less frequently recommending formal reporting: a 1 SD increase in the holding of stereotypical norms decreases the likelihood of advising a victim to report DV to the police by -10.1% and to use support services by -1.8%. The results in the last two columns of the table show that respondents holding more stereotypical beliefs recommend behaviors that exacerbate the underreporting of DV: a 1 SD increase in stereotypes increases the probability that the respondent advises the victim to try to resolve DV with the partner by +7.5% and the probability of advising the victim not to react by +35.2%.

To further appreciate the magnitude of the estimated effects, we calculate the predicted outcome levels along the distribution of the index of stereotypes from our sample (the 25th, median (50th), and 75th), based on the estimated coefficient of stereotype in *Panel B* of Table 2. Column (4) reports the difference between the predicted outcome level at the 75th percentile of stereotypes and the outcome level at the 25th percentile. As individual stereotypes become stronger (from the bottom to the top), for instance, we record a 7.8% decrease in the perception of overall violence, a 9.6% decrease in the probability of perceiving sexual violence as severe, and a 12% decrease in the probability of considering the abuser's personal characteristics as a cause of their violent behaviors. The larger impact is on the recommendation to take no reaction, which increases by 38%. All results are available in Table A9.

Our estimates are smaller than those found in the literature, yet the comparability with the literature is not straightforward since unlike many studies on DV that address the trends in reported episodes (Leslie and Wilson 2020; Bullinger et al. 2021), we focus on the determinants of population perception. In this respect, we use different outcomes as well as a different way to proxy stereotypes. We measure stereotypes at the individual and not at the aggregate level: aggregated data are more likely to carry *aggregate biases*. In fact, the underlying assumption when using aggregate data is that the estimated relation between variables is homogeneous across all individuals for whom the aggregate index is used. As a consequence, the estimated impact of the changes in the explanatory variable on the dependent variable is much larger in the aggregate regression than in the less aggregated regression (Garrett 2003). Additionally, the outcomes found in previous literature are not always comparable since they refer to episode frequencies and, in some cases, mainly focused on intimate partner violence (IPV), specifically referring to physical and sexual violence. For example, González and Rodríguez-Planas (2020) find that a 1-SD increase in gender inequality increases the incidence of violence, intensity of violence, and number of cases of violence by 28%, 43%, and 33, respectively. Tur-Prats (2019a) focuses on a specific country, as in our case, and measures norms using the contemporaneous family

structure (stem vs nuclear) in Spain, claiming that women living in stem families experience a lower incidence of IPV by as much as 57%.²⁰ Alesina et al. (2021) find similar country-level results not only on the incidence of IPV but also on the attitudes toward acceptance of IPV, proxying for stereotypes with information on the ancestral ethnicity of respondents, including the mode of production, marriage status, and residential patterns. They find that respondents with less gender-equal backgrounds are 77% more likely to justify violent acts by the husband. Finally, recent work by Guarnieri and Tur-Prats (2023) found a positive effect between gender norms and sexual violence during civil conflicts. Additionally, in this case, gender norms are measured at the aggregated level of the armed actors. They find that a 1-SD increase in the index of male dominance increases violence by 0.29 SD.

Our results are consistent with previous evidence based on Italian data on the use of help-seeking services, such as antiviolence helplines for women victims of domestic abuse. For example, Colagrossi et al. (2022) studies the effect of a television campaign on the use of the helpline in case of domestic violence during the first wave of COVID-19 in Italy on the actual use of the service. They find an overall increase in calls of 30 to 40% after the campaign in areas with a 1-SD higher exposure to the campaign. However, this effect is much lower in areas with higher gender stereotypes, where the number of calls increased by only 3 to 10%.

5.3 Robustness checks

As discussed in Section 5.2, robustness checks are presented only for our preferred specification. We group the tests in five sets: 1) to address potential problems due to multiple hypothesis testing, 2) to address potential bias stemming from the definition of our outcomes, 3) to address potential unobservable factors, 4) to address potential bias from our definition of stereotypes, and 5) to address potential problems due to the slight imbalance in certain observable dimensions.

5.3.1 Alternative specifications

In a first set of robustness tests, we address potential problems due to multiple hypothesis testing and to the clusters of the standard errors. When performing simultaneous inference with a large number of outcomes, as in our case, there might be a problem of overrejecting the null hypothesis. This might undermine our results since significant coefficients may emerge simply by chance, even if there are no strong effects (Anderson, 2008). Hence, we adjust the p -values to limit the proportion of false rejection of the null hypothesis across all individual outcomes (Allcott et al., 2020) following the approach in Simes (1986), as implemented in Erten and Keskin (2018). Our main results are still robust, as shown in Panel (A) of Table 3.

²⁰ As an alternative estimation strategy, we also propose an instrumental variable strategy following Lewbel (2012); the results are consistent with our main specification and are available upon request.

Table 3 Standard error corrections

Diffusion		Severity		Causes					Appropriate advice						
Overall	PhysicalPsych	Social	EconomicPsych	Physical	Sexual	COVID-19 Crisis	Abuser	Reporting Service	Partner	No reaction					
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	
<i>Panel A: Regional clusters and standard error correction for multiple hypothesis testing:</i>															
Stereotypes	-0.025	0.005	-0.027*	0.049***	0.024*	0.013	-0.039***	-0.056***	0.057***	0.047***	-0.072***	-0.061***	-0.027***	0.043***	0.037***
	(0.016)	(0.014)	(0.015)	(0.015)	(0.014)	(0.015)	(0.008)	(0.013)	(0.015)	(0.011)	(0.011)	(0.017)	(0.009)	(0.009)	(0.009)
Adj. <i>p</i> -values	{0.151}	{0.718}	{0.106}	{0.005}***	{0.121}	{0.433}	{0.001}***	{0.001}***	{0.002}***	{0.001}***	{0.000}***	{0.004}***	{0.010}***	{0.001}***	{0.001}***
<i>Panel B: Wild cluster bootstrap:</i>															
Stereotypes	-0.025	0.005	-0.027*	0.049***	0.024*	0.013	-0.039***	-0.056***	0.057***	0.047***	-0.072***	-0.061***	-0.027***	0.043***	0.037***
	(0.016)	(0.014)	(0.015)	(0.015)	(0.014)	(0.015)	(0.008)	(0.013)	(0.015)	(0.011)	(0.011)	(0.017)	(0.009)	(0.009)	(0.009)
<i>Panel C: Heteroskedastic robust standard error:</i>															
Stereotypes	-0.025*0.005	-0.027**0.049***	-0.027**0.049***	0.024	0.013	-0.039***	-0.056***	0.057***	0.047***	0.047***	-0.072***	-0.061***	-0.027**	0.043***	0.037***
	(0.013)	(0.010)	(0.012)	(0.014)	(0.015)	(0.014)	(0.014)	(0.014)	(0.015)	(0.015)	(0.014)	(0.015)	(0.011)	(0.013)	(0.007)
Mean Out	0.239	0.099	0.222	0.465	0.498	0.474	0.467	0.457	0.487	0.488	0.467	0.320	0.805	0.304	0.056
N obs	4573	4573	4573	4573	4573	4573	4573	4573	4573	4573	4573	4573	4573	4573	4573

Notes: In Panel A, we report standard errors clustered at the regional level in parentheses, and in curly brackets, we report the *p*-values corrected for multiple-hypothesis testing using Simes adjustment, where †, ††, and ††† denote significance for *p* < 0.1, *p* < 0.05, *p* < 0.01, respectively. Standard errors are estimated using a wild cluster bootstrap procedure (1000 repetitions) in Panel B, correcting for heteroskedasticity in Panel C. All regressions include region-fixed effects and controls for experience, life status, and neighborhood. Significance levels: **p* < 0.1; ***p* < 0.05; ****p* < 0.01

Since in our preferred specification, we cluster errors at the regional level, we might have biases generated by the use of a low number of clusters. To address this problem, we exploit the wild cluster bootstrap standard error estimation (Cameron et al., 2008). In Panel B of Table 3, we show that our baseline results are robust to the use of a wild cluster bootstrap approach (Roodman et al., 2019). Our results are also robust to the use of a simpler heteroskedasticity robust standard error regression, as shown by the results in Panel C of Table 3.

5.3.2 Alternative outcomes

As explained in Section 4.2, we asked respondents to rank 14 different types of violence, and we calculated the individual-level average of the elements in each group. However, we did not spell out the value of the position *per se*. For this reason, we also run our baseline regression separately for each element of each group of behavior for a total of 14 regressions, and given the high number of outcomes, we report both unadjusted and adjusted standard errors for multiple hypothesis testing using a Bonferroni procedure (Erten and Keskin, 2018). The results shown in Table 4 are consistent with our baseline estimation.

Additionally, we use a principal component analysis (PCA) with the same dimensions to create an alternative measure of the 5 groups of violent behaviors. Hence, we generate a PCA index for *Social violence* using control over social media and control over friends. Similarly, we created an index for *Economic violence* by running the PCA on the position of control over work and control over expenses; for *Psychological violence*, we do so by running the PCA on the position of quarrels, threats, insults, humiliation, recrimination, and spite; for *Physical violence*, we do so by running the PCA on the position of throwing objects and battering; and for *Sexual violence*, we do so by running the PCA on the position of sex requests and sexual violence. We then use these indexes as outcomes in the regression analysis, as shown in Table 5, and the results do not substantially differ from our baseline.

5.3.3 Alternative controls for unobservables

The presence of other factors simultaneously affecting both stereotypes and DV perceptions (*i.e.*, omitted variable bias) may indeed undermine the robustness of our results. To address the role of unobservable confounding factors, we run several checks. First, we add the average stereotypes index in the birth region of the respondent to the baseline specification, building on the approach proposed by Fernandez and Fogli (2009). According to their work, the beliefs and preferences in the country of origin have significant effects on the realization of individual outcomes. We define a synthetic measure of birth-region stereotypes by assigning the average regional stereotypes index to individuals based on their region of birth. Since we include region-of-residence fixed effects, the effect is identified by movers (*i.e.*, 19% of the sample) whose birth region differs from their region of residence. The results in Panel A of Table 6 show that adding this control does not affect the estimate of the effect of stereotypes.

Table 4 Alternative outcome definitions: ranking, single items

	Social		Economic			Psychological			Physical			Sexual		
	Friends (1)	Social media (2)	Work-life (3)	Expenses (4)	Quarrels (5)	Threats (6)	Insults (7)	Humiliations (8)	Recriminat (9)	Spites (10)	Throw obj (11)	Battering (12)	Violence (13)	Requests (14)
<i>Region fixed effects:</i>														
Stereotypes	0.491*** (0.079)	0.299** (0.110)	0.303** (0.114)	0.218** (0.092)	0.283*** (0.071)	-0.293*** (0.091)	-0.036 (0.070)	-0.334*** (0.100)	0.261*** (0.054)	0.287*** (0.097)	-0.031 (0.084)	-0.512*** (0.101)	-0.637*** (0.108)	-0.299** (0.124)
Adj. <i>p</i> -values	{0.00}***	{0.02}##	{0.02}##	{0.03}***	{0.00}***	{0.01}***	{0.66}	{0.01}***	{0.00}***	{0.01}***	{0.72}	{0.00}***	{0.00}***	{0.03}##
N obs	4573	4573	4573	4573	4573	4573	4573	4573	4573	4573	4573	4573	4573	4573

Notes: We report standard errors clustered at the regional level in parentheses, and in curly brackets, we report the *p*-values corrected for multiple-hypothesis testing using Simes adjustment, where †, †† and ††† denote significance for *p* < 0.1, *p* < 0.05, *p* < 0.01, respectively. All regressions include region-fixed effects and controls for experience, life status, and neighborhood. Significance levels: **p* < 0.1; ***p* < 0.05; ****p* < 0.01

Table 5 Alternative outcome definitions: ranking computed using the principal component analysis

	Social (1)	Economic (2)	Psychological (3)	Physical (4)	Sexual (5)
Stereotypes	0.134*** (0.028)	0.101*** (0.035)	-0.187*** (0.026)	-0.100*** (0.023)	-0.173*** (0.030)
N obs	4573	4573	4573	4573	4573

Notes: Standard errors are robust. Significance levels: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Outcomes are constructed using the PCA instead of the average like in our baseline estimation. All regressions include region-fixed effects and controls for experience, life status, and neighborhood

A potential confounding factor affecting both perceptions of DV and the holding of stereotypes is media.²¹ While we do not have information on the main source of information used by respondents, we recover weekly data on the exposure (time) to television from the end of February to the beginning of June 2020, covering the entire period of the first wave of the pandemic. During that time, because of strict stay-at-home orders, people tended to be exposed to television information more than usual, with some territorial and age variations. Data on the *average time viewing* (ATV) were provided by an Italian company that collected this information (AUDITEL) for both 2019 and 2020. The data refer to two age groups, women aged 25–44 and 45–64. The reason for including these data in our model was to test two specifications. In *Panel B* of Table 6, we show the results for the first specification, where we control for the absolute number of minutes per week per region. In *Panel C*, we control for the rate of growth in ATV from 2019 to 2020 for the same months. Even if the impact of media is not limited to that of television, our baseline results are robust to this specific definition of access to information.

In *Panel D*, we control for an alternative definition of personal experience with DV, since, in the baseline results, we control for the variable *Know a victim*, based on whether the respondent knew a victim of domestic violence *before* the first wave of the COVID-19 pandemic. The idea is to check whether the main effect is driven by experience in the recent past, which might affect both beliefs regarding gender stereotypes and perceptions of violence. The results substantially confirm our baseline estimates.

Finally, to cope with potential measurement errors driven by the individual characteristics of the respondent (Elliott and Valliant 2017) due to the use of self-reported measures, we propose an alternative specification where we also control for *Length of interview*. This is an individual continuous measure reporting the number of minutes elapsed from the start of the survey to its submission, which serves as a proxy for the respondent's level of attention at the time of the interview. The results available

²¹ Following Rawlings and Siddique (2020), we run the tests proposed by Oster (2019) and Altonji et al. (2005). In Table A10, we report the values of both tests for our preferred estimates (Panel B Table 2). The value δ is consistently higher than 1, offering some reassurance regarding the stability of the estimated effect of stereotypes, and β_{lower} are consistent with the estimated coefficients of the baseline. Based on the AET ratio values, the outcomes that might be most sensitive to the selection on unobservables are the perceived diffusion of physical violence and the perceived severity of DV. For this reason, we run an alternative estimation where we also control for media exposure.

Table 6 Alternative controls for unobservables

Diffusion		Severity		Economic Psych		Physical	Sexual	Causes		Appropriate Advice		Partner	No reaction		
Overall	Physical Psych	Social	(4)	(5)	(6)	(7)	(8)	COVID-19	Crisis	Abuser	Reporting	Service	(15)		
(1)	(2)	(3)	(3)	(4)	(5)	(6)	(7)	(9)	(10)	(11)	(12)	(13)	(14)		
<i>Panel A: Stereotypes in region of birth:</i>															
Stereotypes	-0.020	0.007	-0.023	0.049***	0.026*	0.011	-0.038***	-0.055***	0.061***	0.045***	-0.071***	-0.059***	-0.029***	0.043***	0.038***
	(0.017)	(0.014)	(0.015)	(0.014)	(0.013)	(0.016)	(0.009)	(0.013)	(0.016)	(0.013)	(0.011)	(0.016)	(0.010)	(0.011)	(0.009)
<i>Panel B: Auditel 1:</i>															
Stereotypes	-0.024	0.006	-0.027*	0.049***	0.024	0.011	-0.039***	-0.055***	0.056***	0.045***	-0.071***	-0.060***	-0.027***	0.043***	0.037***
	(0.016)	(0.014)	(0.015)	(0.015)	(0.014)	(0.015)	(0.008)	(0.013)	(0.015)	(0.011)	(0.012)	(0.017)	(0.009)	(0.009)	(0.009)
<i>Panel C: Auditel 2:</i>															
Stereotypes	-0.025	0.006	-0.029*	0.048***	0.022	0.014	-0.039***	-0.052***	0.056***	0.045***	-0.072***	-0.060***	-0.028***	0.043***	0.036***
	(0.016)	(0.014)	(0.015)	(0.015)	(0.014)	(0.015)	(0.008)	(0.014)	(0.015)	(0.011)	(0.012)	(0.017)	(0.009)	(0.010)	(0.009)
<i>Panel D: Personal Experience (before first wave of COVID-19) - substitute experience:</i>															
Stereotypes	-0.020	0.007	-0.022	0.047***	0.022	0.014	-0.037***	-0.056***	0.054***	0.047***	-0.068***	-0.056***	-0.037***	0.055***	0.038***
	(0.016)	(0.014)	(0.015)	(0.014)	(0.013)	(0.015)	(0.008)	(0.013)	(0.014)	(0.011)	(0.011)	(0.016)	(0.009)	(0.009)	(0.009)
<i>Panel E: Length of interview:</i>															
Stereotypes	-0.026	0.004	-0.028*	0.047***	0.023	0.014	-0.038***	-0.055***	0.058***	0.049***	-0.069***	-0.061***	-0.027***	0.044***	0.037***
	(0.016)	(0.014)	(0.015)	(0.014)	(0.014)	(0.015)	(0.009)	(0.014)	(0.015)	(0.011)	(0.012)	(0.017)	(0.009)	(0.010)	(0.009)
Mean Out	0.239	0.099	0.222	0.465	0.498	0.474	0.467	0.457	0.487	0.488	0.467	0.320	0.805	0.304	0.056
N obs	4573	4573	4573	4573	4573	4573	4573	4573	4573	4573	4573	4573	4573	4573	4573

Notes: Standard errors are reported in parentheses and clustered by region. All regressions include region-fixed effects and controls for experience, life status, and neighborhood. Significance levels: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

in *Panel E* of Table 6 confirm that individual characteristics that can simultaneously affect multiple self-reported answers, such as the level of attention, do not drive the results.²²

5.3.4 Alternative stereotype indexes

We test for possible interpretation challenges related to our index. For instance, it might be that the type of questions on which the index relies affects the interpretation of the results. The concern behind this check is related to so-called social desirability bias. This bias arises if respondents tend to try to comply with or satisfy the expectations of whoever designed the survey. In our context, this would mean that respondents, when facing some clearly gender-stereotypical questions, such as whether men are better than women at performing a specific task, would tend to state that men and women are equally able to perform the task even if their true belief differs. If the bias is strong, it would mean that the estimates of the baseline specification are identified by respondents who — notwithstanding social desirability bias — are outspoken in affirming that men are better than women at performing specific tasks. As a consequence, the identified effect would represent the upper part of the distribution of stereotypical beliefs among women. The answers corresponding to those women who subscribe to stereotypes but are ashamed to be straightforward about it would not explain the results we found.

Hence, to account for the potential role of social desirability bias, we measure individual stereotypical beliefs with questions potentially less sensitive to this bias. The new indexes are again given z scores, as already discussed in Section 4.1, and the results are shown in Table 7). The first index (*Panel A*) relies on respondents' identification of gender-specific job positions (question 31). Respondents were asked to state whether each of 8 listed professions (from schoolteacher to public transport inspector) was more suitable for a man, a woman, or someone of either gender. The index on this question measures how frequently a respondent chose a specific gender instead of the gender-neutral "either." *Panel B* shows the results based on an index computed on answers to general statements about the balance between private and working life (question 32). Again, respondents faced 8 hypothetical situations (from "a woman works more than her partner" to "a boy does not have a driving license") and expressed how acceptable (from *very* to *not at all*) they considered the situations to be.

The baseline estimates are confirmed. In some cases, these alternative indexes have a higher correlation (i.e., larger magnitude) with the outcomes of interest. For instance, respondents with stronger stereotypes have a much lower perception of the diffusion of *all* types of violence (overall, psychological, and psychological violence). Since the

²² As far as measurement errors and spurious correlations are concerned, as an additional check, following the concern raised by Podsakoff and Organ (1986), that in a survey, there might be issues stemming from the fact that both the independent and dependent variables are captured by the same response method, producing spurious correlations by construction, we run a variance analysis by running a PCA on all 15 outcomes used in the paper, our measure of stereotypes and whether the individual knows a victim. We report these analyses in Table A11. Harman's test focuses on the proportion of variance of each single component in explaining the overall variance. As we can see, the variance of each component does not explain more than 14% of the overall variance.

Table 7 Alternative measures of stereotypes

	Diffusion			Severity			Causes			Appropriate advice					
	Overall (1)	Physical (2)	Psych (3)	Physical (4)	Psych (5)	EconomicPsych (6)	Physical (7)	Sexual (8)	COVID-19Crisis (9)	Abuser (10)	Reporting Service (11)	Partner (12)	No reaction (13)		
<i>Panel A: A31:</i>															
Stereotypes	-0.052*	-0.023**	-0.059**	0.054*	0.083*	-0.027	-0.057*	-0.095***	0.142***	0.114***	-0.103***	0.003	-0.039*	0.057**	0.039***
	(0.025)	(0.009)	(0.025)	(0.030)	(0.044)	(0.023)	(0.031)	(0.031)	(0.034)	(0.028)	(0.033)	(0.033)	(0.020)	(0.026)	(0.011)
<i>Panel B: A32:</i>															
Stereotypes	-0.063***	-0.026**	-0.067***	0.019	0.003	0.039***	-0.024*	-0.045***	0.030***	0.034***	-0.014	-0.003	-0.016	0.001	0.029***
	(0.009)	(0.010)	(0.008)	(0.013)	(0.014)	(0.010)	(0.013)	(0.014)	(0.010)	(0.011)	(0.010)	(0.014)	(0.013)	(0.007)	(0.008)
<i>Panel C: Principal component analysis for stereotypes:</i>															
Stereotypes	-0.007	0.003	-0.008	0.019***	0.011*	0.003	-0.015***	-0.022***	0.024***	0.020***	-0.027***	-0.025***	-0.005	0.014***	0.012***
	(0.006)	(0.005)	(0.005)	(0.006)	(0.005)	(0.006)	(0.003)	(0.005)	(0.005)	(0.005)	(0.005)	(0.007)	(0.003)	(0.004)	(0.004)
<i>Panel D: Stereotypes dummy:</i>															
Stereotypes	-0.029*	-0.000	-0.032**	0.025	0.017	0.012	-0.029**	-0.040**	0.039***	0.036**	-0.058***	-0.040**	-0.030***	0.043***	0.032***
	(0.014)	(0.014)	(0.014)	(0.015)	(0.017)	(0.018)	(0.010)	(0.015)	(0.012)	(0.013)	(0.014)	(0.014)	(0.010)	(0.012)	(0.007)
Mean Out	0.239	0.099	0.222	0.465	0.498	0.474	0.467	0.457	0.487	0.488	0.467	0.320	0.805	0.304	0.056
N obs	4573	4573	4573	4573	4573	4573	4573	4573	4573	4573	4573	4573	4573	4573	4573

Notes: Standard errors are reported in parentheses and clustered by region. All regressions include region-fixed effects and controls for experience, life status, and neighborhood. Significance levels: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. A31 is a question about the gender suitability of professions, while A32 is a question about the acceptability of a list of behaviors or conditions (see Appendix B for the full question). The principal component analysis is used to construct our measures of stereotypes using the same 8 questions used to construct our baseline measure of stereotypes, as reported in Table A2.

main results are not dramatically affected, we choose to stick with our initial index; given that we have no way to completely rule out social desirability bias, this index at least provides us with a clearer result once we keep in mind that it represents a lower bound of the true correlation between stereotypes and perceptions of DV.

Additionally, since our measure of stereotypes is in fact created using the average of standardized answers to 8 items of one question, it might be that the answers to these items are strongly correlated because they either capture similar dimensions of stereotypes or simply suffer from a common method bias. Hence, we create an alternative stereotype index using a (polychoric) principal component analysis. The polychoric PCA measures the correlation between two unobserved continuous variables, but above all, unlike normal PCA, it can address a larger proportion of variance. In fact, the polychoric PCA would allow us to maintain the ordinal nature of the Likert scale, specifically how respondents have answered those questions. Results, using this PCA transformation of the stereotype index, are confirmed as shown in Table 7, *Panel C*.

Finally, in *Panel D*, we also show the results for a different version of our index using a dummy variable rather than the continuous index. It distinguishes between *weak* and *strong* stereotypes based on the mean of the stereotype index distribution (mean=0). While simplifying the interpretation of the estimates, it also shows that the results are not substantially affected.

5.3.5 External validity

As discussed in Section 3, our sample is well representative along certain dimensions and is slightly less representative along other dimensions positively correlated with stereotypes. Hence, as a final check, we use an entropy balance procedure to match our sample to other sample statistics by region and consequently create weights to achieve this balance with respect to relevant dimensions of the population. We focus on regions since they define our territorial unit of interest, although the sample was not defined based on them, at least for the center and the south. We use the entropy balance methodology to create weights and run a weighted version of our baseline. The results are available in Table 8, both when we weight using only the age distribution *Panel A* and when we weight using education, flat size, and number of children distributions within regions *Panel B*.

6 Acceptability of violence

We go into greater depth on the beliefs expressed on the severity of violent behavior, with respondents holding stronger stereotypes ranking economic and social violence as more severe than physical violence. We proceed by looking at what is considered acceptable within a couple relationship and into the potential role of victims of violent behavior. We examine two questions: the first concerns the acceptability of an occasional slap in a relationship, while the second concerns blaming a rape victim for her own victimization if she was drunk at the time of the rape. These questions aim to check how likely respondents holding stronger stereotypes are to rationalize men's

Table 8 Weighted regression

	Diffusion		Severity		Causes				Appropriate advice						
	Overall	Physical Psych	Social	Economic Psych	Physical	Sexual	COVID-19	Crisis	Abuser	Reporting Service	No reaction				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
<i>Panel A: Age-Weighted regression:</i>															
Stereotypes	-0.020	0.005	-0.023	0.049***	0.025*	0.011	-0.037***	-0.054***	0.052***	0.045***	-0.075***	-0.059***	-0.031***	0.044***	0.038***
	(0.016)	(0.014)	(0.015)	(0.015)	(0.014)	(0.015)	(0.008)	(0.013)	(0.016)	(0.012)	(0.012)	(0.017)	(0.008)	(0.010)	(0.009)
<i>Panel B: Weighted regression using education, flat size and no. of children:</i>															
Stereotypes	-0.021	0.007	-0.025	0.051***	0.023	0.009	-0.041***	-0.043***	0.052**	0.040**	-0.080***	-0.054***	-0.029***	0.041***	0.038***
	(0.017)	(0.016)	(0.016)	(0.016)	(0.016)	(0.020)	(0.010)	(0.013)	(0.019)	(0.015)	(0.010)	(0.017)	(0.007)	(0.011)	(0.009)
Mean Out	0.239	0.099	0.222	0.465	0.498	0.474	0.467	0.457	0.487	0.488	0.467	0.320	0.805	0.304	0.056
N obs	4573	4573	4573	4573	4573	4573	4573	4573	4573	4573	4573	4573	4573	4573	4573

Notes: Standard errors are reported in parentheses and clustered by region. All regressions include region-fixed effects and controls for experience, life status, and neighborhood. Significance levels: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Weights in Panel A are created using the regional share of age groups as reported in Table B4. Weights in Panel B are created using regional population averages of education (high school and above), flat size and number of children from official statistics reported in Table B5

misbehavior as resulting from women's deviation from behavioral norms, consistent with a benevolent sexist approach. Having stronger stereotypes is associated with a higher perceived acceptability of a slap in a relationship and more blame attribution to a rape victim if she was drunk. The results are robust across specifications, as shown in Table 9.

7 Conclusion

An accurate understanding of the seriousness and potential consequences of certain types of behaviors is, however, the first step toward gaining awareness of and dealing with DV. If DV is believed to be a rare event, reporting is likely to be lower and the matter more likely to be relegated to the private realm. By the same token, the costs in

Table 9 Outcome: acceptability

	Drunk (1)	Slap Acceptability (2)
<i>Panel A: Baseline:</i>		
Stereotypes	0.185*** (0.013)	0.108*** (0.011)
R-squared	0.168	0.139
<i>Panel B: Stereotypes in region of origin:</i>		
Stereotypes	0.180*** (0.014)	0.111*** (0.012)
R-squared	0.166	0.144
<i>Panel C: Personal Experience (before first wave of COVID-19)</i>		
<i>Substitute experience:</i>		
Stereotypes	0.184*** (0.013)	0.108*** (0.012)
R-squared	0.169	0.140
<i>Panel D: Auditel 1:</i>		
Stereotypes	0.185*** (0.013)	0.108*** (0.012)
R-squared	0.172	0.142
<i>Panel E: Auditel 2:</i>		
Stereotypes	0.184*** (0.013)	0.107*** (0.012)
R-squared	0.170	0.141
Mean Out	0.187	0.088
N obs	4573	4573

Notes: Standard errors are reported in parentheses and clustered by region. All regressions include region-fixed effects and controls for experience, life status, and neighborhood. Significance levels: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

terms of psychical and mental health, the costs on working and career outcomes, and the costs on the cognitive and emotional development of involved children are also going to be relegated to the private realm. By changing how DV is perceived and how it affects public opinion, victims could become more empowered to deal with it.

Our analysis focuses on gender stereotypes among the main drivers of DV perception while controlling for a large set of factors that are expected to affect the perception of DV (e.g., socioeconomic factors). Our findings show that stereotypes are not necessarily just the product of the environment in which they are embedded, since we obtain significant results of stereotypes even when we absorb the territorial effects using respondents' region of residence fixed effects. This means that in principle, you might live in a very emancipated context and yet have strong stereotypes, and vice versa. We think that this is the main takeaway of our analysis, which provides a lower bound estimate of the true effects, and although it is common knowledge that cultural norms do not change overnight, slight changes are possible. For instance, stereotypes should be the target of ad hoc campaigns. As shown by La Ferrara et al. (2012); La Ferrara (2016), the media can influence and shape stereotypes, particularly gender roles, by holding up new role models. Providing more information on the availability of services to victims, such as helplines, can be equally useful. These campaigns could be easily implemented, and they could generate the benefit of increasing awareness of a problem that is all too common.

As personal backgrounds and experiences matter in defining both norms and perceptions of DV, future research should be devoted to measuring and recording personal experiences in depth, such as asking questions about task allocations in the family of origin and the working status of the female figures. In our survey, we addressed personal experiences, working on the framing of questions so as to allow for a self-reference approach. The use of a panel could further help in this respect.

Our results do not include male perceptions of DV since we rely on a survey of only females, and although the stereotypes of males and females are highly correlated at the territorial level, it would be interesting to analyze how a male respondent would perceive the severity of certain behaviors and what action he would suggest in the face of DV situations. Finally, we do not identify either the gender identity of the victim or the type of relationship between the victim and the abuser, which would be worthwhile to investigate in future works on the topic.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00148-024-00986-0>.

Acknowledgements We thank the anonymous referees for helpful comments and suggestions. We thank participants in the XXXIV Annual Conference of the Italian Society of Public Economics, the 2022 European Health Economics Seminar Series, and the 2023 Health Economics of Risky Behaviors Workshop for their comments on a previous version of this work. We also thank editor Milena Nikolova and two reviewers for helpful comments. We are grateful to Pamela Campa for giving us access to the 2005 World Value Survey. We gratefully acknowledge financial support from Fondazione Cariplo. All conclusions, however, are our own.

Funding Open access funding provided by Università degli Studi dell'Insubria within the CRUI-CARE Agreement.

Data availability The data that support the findings of this study are available on request from the corresponding author, Anna Cecilia Rosso. The data are not publicly available due to restrictions requested by the ethical committee given the sensitivity of the topic.

Declarations

Ethics approval The survey was approved by the Ethics Committee of the University of Milan (Prot. N.75/20 - 29-06-2020).

Consent to participate All participants gave their informed consent at the beginning of the survey.

Conflict of interest The authors declare no competing interests.

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

References

- Agüero JM (2019) Information and behavioral responses with more than one agent: the case of domestic violence awareness campaigns. Unpublished Manuscript
- Agüero JM (2021) COVID-19 and the rise of intimate partner violence. *World Dev* 137
- Aizer A (2010) The gender wage gap and domestic violence. *Am Econ Rev* 100(4):1847–1859
- Aizer A, Dal Bó P (2009) April. Love, hate and murder: commitment devices in violent relationships. *J Public Econ* 93:412–428
- Alesina A, Brioschi B, La Ferrara E (2021) Violence against women: a cross-cultural analysis for Africa. *Economica* 88(349):70–104
- Alesina A, Ferroni MF, Stantcheva S (2021) Perceptions of racial gaps, their causes, and ways to reduce them. NBER Working Paper No. 29245
- Alesina A, Giuliano P, Nunn N (2013) On the origins of gender roles: women and the plough. *Q J Econ* 128(2):469–530
- Allcott H, Braghieri L, Eichmeyer S, Gentzkow M (2020) The welfare effects of social media. *Am Econ Rev* 110:629–676
- Altonji JG, Elder TE, Taber CR (2005) Selection on observed and unobserved variables: assessing the effectiveness of Catholic schools. *J Polit Econ* 113(1):151–184
- Anderberg D, Rainer H, Wadsworth J, Wilson T (2016) Unemployment and domestic violence: theory and evidence. *Econ J* 126(597):1947–1979
- Anderson ML (2008) Multiple inference and gender differences in the effects of early intervention: a reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *J Am Stat Assoc* 103(484):1481–1495
- Arenas-Arroyo E, Fernandez-Kranz D, Nollenberger N (2021) Intimate partner violence under forced cohabitation and economic stress: evidence from the COVID-19 pandemic. *J Public Econ* 194:104350
- Bates EA, Klement KR, Kaye LK, Pennington CR (2019) The impact of gendered stereotypes on perceptions of violence: a commentary. *Sex Roles* 81(1):34–43
- Bentley SV, Greenaway KH, Haslam SA (2017) An online paradigm for exploring the self-reference effect. *PLoS ONE* 12(5):e0176611. <https://doi.org/10.1371/journal.pone.0176611>
- Bhalotra S, Britto DGC, Pinotti P, Sampaio B (2020) Job displacement, unemployment benefits and domestic violence. mimeo 40

- Bloch F, Rao V (2002) Terror as a bargaining instrument: a case study of dowry violence in rural India. *Am Econ Rev* 92(4):25
- Boyd R, Richerson PJ (1995) Why does culture increase human adaptability? *Ethol Sociobiol* 16(2):125–143
- Boyd R, Richerson PJ (2005) *The origin and evolution of cultures*. Oxford University Press
- Bullinger LR, Carr J, Packham A (2020) COVID-19 and crime: effects of stay-at-home orders on domestic violence. Technical Report w27667, National Bureau of Economic Research, Cambridge, MA
- Bullinger LR, Carr JB, Packham A (2021) COVID-19 and crime effects of stay-at-home orders on domestic violence. *Am J Health Econ* 7(3):249–280
- Button DM (2008) Social disadvantage and family violence: neighborhood effects on attitudes about intimate partner violence and corporal punishment. *Am J Crim Justice* 33(1):130–147
- Cameron AC, Gelbach JB, Miller DL (2008) Bootstrap-based improvements for inference with clustered errors. *Rev Econ Stat* 90(3):414–427
- Caselli M, Fracasso A, Scicchitano S (2022) From the lockdown to the new normal: individual mobility and local labor market characteristics following the COVID-19 pandemic in Italy. *J Popul Econ* 35(4):1517–1550
- Colagrossi M, Deiana C, Geraci A, Giua L (2022) Hang up on stereotypes: domestic violence and an anti-abuse helpline campaign. *Contemp Econ Policy* 40(4):585–611
- Copp JE, Giordano PC, Longmore MA, Manning WD (2019) The development of attitudes toward intimate partner violence: an examination of key correlates among a sample of young adults. *J Interpers Violence* 34(7):1357–1387
- Dobash RE, Dobash R (1979) *Violence against wives* (New York: F. ed.)
- Elliott MR, Valliant R (2017) Inference for nonprobability samples. *Stat Sci* 32(2):249–264
- Erten B, Keskin P (2018) For better or for worse?: education and the prevalence of domestic violence in Turkey. *Am Econ J Appl Econ* 10(1):64–105
- Farmer A, Tiefenthaler J (1997) An economic analysis of domestic violence. *Rev Soc Econ* 55(3):337–358
- Fernandez R, Fogli A (2009) Culture: an empirical investigation of beliefs, work, and fertility. *Am Econ J Macroecon* 1(1):146–177
- FRA (2015) *Violence against women: an EU-wide survey. Main Results, Report - EU Agency for Fundamental Rights*
- Garrett TA (2003) Aggregated versus disaggregated data in regression analysis: implications for inference. *Econ Lett* 81(1):61–65
- Glick P, Fiske ST (1996) The ambivalent sexism inventory: differentiating hostile and benevolent sexism. *J Pers Soc Psychol* 70(3):491–512
- Glick P, Fiske ST, Mladinic A, Saiz JL, Abrams D, Masser B, Adetoun B, Osagie JE, Akande A, Alao A, Annetje B, Willemsen TM, Chipeta K, Dardenne B, Dijksterhuis A, Wigboldus D, Eckes T, Six-Materna I, Exposito F, Moya M, Foddy M, Kim HJ, Lameiras M, Sotelo MJ, Mucchi-Faina A, Romani M, Sakalli N, Udegbe B, Yamamoto M, Ui M, Ferreira MC, Lopez WL (2000) Beyond prejudice as simple antipathy: hostile and benevolent sexism across cultures. *J Pers Soc Psychol* 5(79):763–775
- González L, Rodríguez-Planas N (2020) Gender norms and intimate partner violence. *J Econ Behav Organ* 178:223–248
- Gracia E, Herrero J (2006) Acceptability of domestic violence against women in the European Union: a multilevel analysis. *J Epidemiol Community Health* 60(2):123–129
- Guarnieri E, Tur-Prats A (2023) Cultural distance and conflict-related sexual violence. *Q J Econ* 138(3):1817–1861
- Istat (2019) *Stereotypes about gender roles and the social image of sexual violence*
- Kernsmith P (2006) Gender differences in the impact of family of origin violence on perpetrators of domestic violence. *J Fam Violence* 21(2):163–171
- La Ferrara E (2016) Mass media and social change: can we use television to fight poverty? *J Eur Econ Assoc* 14(4):791–827
- La Ferrara E, Chong A, Duryea S (2012) Soap operas and fertility: evidence from Brazil. *Am Econ J Appl Econ* 4(4):1–31
- Leslie E, Wilson R (2020) Sheltering in place and domestic violence: evidence from calls for service during COVID-19. *J Public Econ* 189:104241
- Lewbel A (2012) Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. *J Bus Econ Stat* 30(1):67–80
- MacMillan R, Gartner R (1999) When she brings home the bacon: labor-force participation and the risk of spousal violence against women. *J Marriage Fam* 61(4)

- McDonald DG, Sarge MA, Lin SF, Collier JG, Potocki B (2015) A role for the self: media content as triggers for involuntary autobiographical memories. *Commun Res* 42(1):3–29. <https://doi.org/10.1177/0093650212464771>
- Mihalic SW, Elliot D (1997) A social learning theory model of marital violence. *J Fam Violence* 12(1):21–47
- Miller AR, Segal C, Spencer MK (2020) Effects of the COVID-19 pandemic on domestic violence in Los Angeles. Technical report
- Morgan RE, Oudekerk BA (2019). Criminal victimization, 2018. NCJ- Bureau of Justice Statistics (253043), 37
- Morgan RE, Truman JL (2018) Criminal victimization, 2017. NCJ- Bureau of Justice Statistics 252472, 30
- Mulla MM, Witte TH, Richardson K, Hart W, Kassing FL, Coffey CA, Hackman CL, Sherwood IM (2019) The causal influence of perceived social norms on intimate partner violence perpetration: converging cross-sectional, longitudinal, and experimental support for a social disinhibition model. *Pers Soc Psychol Bull* 45(4):652–668
- Oster E (2019) Unobservable selection and coefficient stability: theory and evidence. *J Bus Econ Stat* 37(2):187–204
- Overstreet NM, Quinn DM (2013) The intimate partner violence stigmatization model and barriers to help seeking. *Basic Appl Soc Psych* 35:109–122
- Parliamentary-Committee (2020) Commissione parlamentare di inchiesta sul femminicidio, nonché su ogni forma di violenza di genere- relazione sui dati riguardanti la violenza di genere e domestica nel periodo di applicazione delle misure di contenimento per l'emergenze da COVID19. Senato della Repubblica Doc. XXII-bis n, p 2
- Podsakoff PM, Organ DW (1986) Self-reports in organizational research: problems and prospects. *J Manage* 12(4):531–544
- Rawlings S, Siddique Z (2020) Domestic violence and child mortality in the developing world. *Oxf Bull Econ Stat* 82(4):723–750
- Rogers TB, Kuiper NA, Kirker WS (1977) Self-reference and the encoding of personal information. *J Person Soc Psychol* 35(9):677
- Roodman D, Nielsen MØ, MacKinnon JG, Webb MD (2019) Fast and wild: bootstrap inference in Stata using boottest. *Stata J* 19(1):4–60
- Sanga S, McCrary J (2021) The impact of the coronavirus lockdown on domestic violence
- Seff I (2021) Social norms sustaining intimate partner violence: a systematic review of methodologies for proxy measures. *Trauma Violence Abuse* 1–20
- Simes RJ (1986) An improved Bonferroni procedure for multiple tests of significance. *Biometrika* 73(3):751–754
- Tauchen HV, Witte AD, Long SK (1991) Domestic violence: a nonrandom affair. *Int Econ Rev* 32(2):491–511
- Tur-Prats A (2019a) Family types and intimate partner violence: a historical perspective. *Rev Econ Stat* 101(5):878–891
- Tur-Prats A (2019b) Unemployment and intimate-partner violence: a cultural approach. mimeo UC Merced 52
- UN Women (2020) Policy brief: the impact of COVID-19 on women
- Van Raaij WF (1988) Information processing and decision making cognitive aspects of economic behaviour. In *Handbook of economic psychology*, pp 74–106. Springer, Dordrecht
- Zhang Y, Breunig R (2023) Female breadwinning and domestic abuse: evidence from Australia. *J Popul Econ* 36(4):2925–2965

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