

# How Many Slaps Is Equivalent to One Punch? New Approaches to Assessing the Relative Severity of Violent Acts

Magda Osman  
Queen Mary University London

Dzenana Pupic  
Karl-Franzens University Graz

Nicholas Baigent  
London School of Economics

**Objective:** The study assesses the extent to which responses to different judgment tasks align consistently to reveal underlying rank orderings of violent acts by their level of severity. **Method:** There were 4 studies (total  $N = 540$ , adult sample). Participants were presented with a variety of tasks: **ranking task**: report your ranking of violent acts from least to most violent; trade-off task: report how many of less violent acts is equivalent to 1 more violent act?; compensation task: report the financial compensation needed that would exactly compensate you for being a victim of a violent act. For these tasks they were required to consider 3 violent acts (Study 1), and 8 violent acts (Studies 2, 3, and 4). **Results:** The similarity in the rank ordering of violent acts by severity in each study was relatively high in each study (77–92%). In Study 4 the average compensation (USD) request was for a spit (\$8,929), slap (\$9,876), kick (\$10,499), punch (\$10,354), head-butt (\$19,636), threaten with a knife (\$11,631), choke (\$118,119), and stab (\$125,596). A meta-analysis conducted across studies (2–4) also revealed that there was general agreement in the way in which violent acts were ordered by severity ( $p < .0005$ ); from least to most severe: spit, slap, kick, punch, head-butt, threaten with a knife, choke, and stab. **Conclusion:** People generally agree on their ranking of violent acts according to their severity, and from this it is possible to develop ranking systems that are sensitive to people's judgments of the level of violence of various acts.

**Keywords:** violent acts, severity of violence, aggregation problem, rates of violence, assessment of violence

In order to convincingly or plausibly say that the overall level of violence in a population has decreased, or increased, a measure of total violence needs to do the following: it should differentiate violent acts (e.g., assault, rape, homicide) according to their levels of violence and this requires a ranking of acts according to their levels of severity. In other words, not all violent acts are equally violent, so when the total number of violent acts are aggregated so that rates of violence can be tracked from one time period to another, the aggregation needs to take into account the differences in the severity of the acts. Currently, many measures of violence do not adopt a measurement system like the one outlined here, and the consequence of this is that distortions arise in the aggregation of violent acts (i.e., aggregation problem; we consider this in more detail in the General Discussion section). The aim of this study is to examine whether indeed it is psychologically plausible for

people to rank violent acts according to their severity that suggests internal consistency in their judgments and decisions. What we mean by internal consistency here is that judgments regarding the severity of violent acts are complementary across a range of different measurement tools that probe for severity of violent acts in different ways (e.g., rank ordering of violent acts, trade-offs, monetary compensation requests). Such evidence would provide important insights for any measurement tool that aims to accurately capture an aggregate rate of violence.

To date, there are over 77,140 published social science research articles on Web of Science (from 1950–2016) on the subject of violence. Limiting attention to work published only in the past year, rates of violence have been studied for, among others, different ages (Fry et al., 2016), educational levels (Sherr et al., 2016), genders (Elmqvist et al., 2016), controlled substances (Stappenbeck, Gulati, & Fromme, 2016), psychiatric disorders (Gray & Snowden, 2016), problem gamblers (Dowling et al., 2016), cultures (Sabina, Cuevas, & Cotignola-Pickens, 2016), gun control laws (Prinsloo, Matzopoulos, Laubscher, Myers, & Bradshaw, 2016), work settings (Vladutiu, Casteel, Nocera, Harrison, & Peek-Asa, 2016), and immigration (Bui & Farrington, 2016). This work is not only necessary for understanding what leads to violent behavior, but also for establishing policies that aim to prevent and reduce violence. However, these studies acknowledge that there is no consistent definition of violence, and no agreed set of tools for quantifying violence. As we have argued, this has a

This article was published Online First November 10, 2016.

Magda Osman, Psychology Division, Biology and Experimental Psychology Centre, Queen Mary University London; Dzenana Pupic, Faculty of Economics, Karl-Franzens University Graz; Nicholas Baigent, Centre for Philosophy of Natural and Social Sciences, London School of Economics.

Correspondence concerning this article should be addressed to Magda Osman, Psychology Division, Biology and Experimental Psychology Centre, Queen Mary University London, London, Mile End Road, E14NS London, United Kingdom. E-mail: [m.osman@qmul.ac.uk](mailto:m.osman@qmul.ac.uk)

bearing on the plausibility of claims regarding overall increases or decreases in rates of violence. If some acts are deemed as violent (e.g., the threat of violence) along with more commonly judged acts (e.g., punching) then aggregating records of violent acts in a given population can lead to higher rates of violence being reported, as compared with other studies (e.g., Walsh et al., 2002). Therefore one important practical starting point is to develop an agreed way to measuring violent acts.

Some of the issues raised here are put sharply into focus in a review of measurement tools that are implemented in order to quantify levels of violence: in this case, in the mental health domain. Harris, Oakley, and Picchioni (2013) comprehensively evaluate nine tools, of which four do not distinguishing levels of severity of violent acts. The remaining five tools distinguished acts of violence and aggression according to levels of severity: Modified Overt Aggression Scale (MOAS; Kay, Wolkenfeld, & Murrill, 1988), MacArthur Community Violence Instrument (MCVI; Steadman et al., 1998), Liverpool Violence Assessment (Nathan et al., 2003), Quantification of Violence Scale (Tyrer et al., 2007), and the Attempted and Actual Assault Scale (Bowers et al., 2007). Harris et al.'s (2013) comments are worrying, given the popularity of these tools. The general conclusions are that the severity of violent acts in many of the tools is not validated, and the reliability of the assumed levels of violence is not sufficiently high enough across studies; moreover, these issues were raised by the authors of the tools themselves. Harris et al. (2013) conclude that the quantification of violence is still a key research challenge, and more work is needed to improve the sensitivity of the tools and the methodologies used to assess those tools. In addition, if the tools that researchers and clinicians use to quantify violence and aggression fail to adequately establish levels of severity, then this, in turn, is likely to impact the empirical studies that use them.

### Motivation Behind the Present Studies

In the present set of four studies we consider the very basic issue of whether, in fact, people have reliable intuitive conceptions of how to rank violent acts according to their severity. In the main we are concerned with the severity of violence of the acts themselves (e.g., a punch) without specific reference to the consequences of the acts. If, in fact, people show that their judgments of severity on a variety of measurement tools are similar, as well as intraconsistency across samples irrespective of age, gender, and culture, then this would provide compelling evidence that could help pave the way for further empirical assessment of scales that are used to quantify violence that use rank orderings, and from which weightings can be developed so that aggregation of reports of violence are more plausible/convincing.

Thus, the aim of the four studies in this paper is to examine the extent to which judgments are similar regarding the ordering of violent acts according to levels of severity in a range of participant populations, and across three types of tasks: ranking task, trade-off task, and compensation task. The ranking task provided a direct assessment of the ordering of violent acts according to levels of severity, and the other two were indirect measures of the rank ordering of violent acts. The first is the trade-off task in which people are required to trade-off one violent act against another (e.g., indicate how many slaps would need to be performed to be equivalent to 1 punch). In the other participants indicated the

financial compensation they would expect to receive after experiencing a violent act. This is commonly referred to as willingness-to-pay measures, which have been used in previous studies for the same purposes as ours (Bishop & Murphy, 2011). Trade-off type tasks are also used to assess the relative severity of violent acts in which one violent act is judged against another violent act (Conrad et al., 2010; Hudson & McIntosh, 1981; Peek-Asa, Garcia, McArthur, & Castro, 2002; Suglia et al., 2008). In addition, for Studies 2–4, we had the opportunity to examine the judgments and decisions from different samples, predominately United Kingdom, Bosnian, and Austrian, providing us with the possibility of investigating the extent to which judgments regarding the severity of violent acts were consistent across nationalities. Comparisons of this type have been employed to examine the extent to which different nationalities share similar patterns of judgments regarding the severity of violent acts. For instance, Peek-Asa et al. (2002) reported that their United States and Mexican samples showed significant overlap in the rank ordering of violent acts. By extension, we aim to explore similarity in judgments in different nationalities for different measurements of severity of violent acts (trade-off, compensation), in addition to a ranking task.

The selection of the violent acts included in the four studies was drawn from items that feature in several assessments of violence in mental health communities (e.g., MOAS; Kay et al., 1988; MCVI; Steadman et al., 1998), against women (e.g., Severity of Violence Against Women Scale [SVAWS]; Marshall, 1992), against families (Conflict Tactic Scale [CTS]; Straus, 1979), and in violent offenders (e.g., Violence Risk Appraisal Guide [VRAG]; Harris, Rice, & Quinsey, 1993). The reason for selecting the eight main violent acts that featured in Studies 2–4 was that while some assessments differentiate between the slap, kick, and punch, some do not, and they classify them into one broad category. However, in many cases violent acts such as choke and threaten with a knife are treated as separate items in many assessments. In addition, items referring to physical assaults often include stabbing and using blunt objects to create injury in a single category (as an exception, see Peek-Asa et al., 2002). Also, of the various assessment tools that exist, violent acts such as spitting are not often included, and so are not assessed directly alongside other acts of varying severity. Given that many assessment tools do not distinguish many of the aforementioned violent acts individually on an item-level basis, we included the following eight violent acts that commonly appear in assessment tools: spit, slap, kick, punch, threaten with a knife, head-butt, choke, and stab. In addition, we derive our predictions regarding the positioning of the 8 violent acts used in our investigations from the aforementioned assessment tools, and several key studies examining the severity of violent acts (Goodman et al., 1999; Hudson & McIntosh, 1981; Krahé & Busching, 2015; Peek-Asa et al., 2002).

### Research Purposes and Hypotheses

The overall research purpose is to assess the extent to which lay people's responses to different judgment tasks align to reveal reliable underlying rank orderings of violent acts by their level of severity.

*Hypothesis 1:* Based on previous work, the rank ordering of violent acts from low to high severity follow the order: spit, slap, kick, punch, threaten with a knife, head-butt, choke, stab.

*Hypothesis 2:* Trade-offs made between different violent acts will correspond with the same order of severity of violent acts as revealed in the explicit rank-ordering tasks on the same violent acts.

*Hypothesis 3:* The compensation amounts requested for experiencing the different violent acts will correspond with the same order of severity of violent acts as revealed in the explicit rank-ordering tasks on the same violent acts.

### Pilot Study 1a

The purpose of Pilot Study 1a was to explore the general possibility that two different judgment measures of violent acts (ranking task, trade-off task) could be used to assess the rank ordering of three violent acts in order of severity. The hypothesis tested in Pilot Study 1a was: if people have a stable representation of the severity of three violent acts, then an indirect assessment of their rank ordering of three violent acts involving a trading off between (assumed) lower level violent acts against the highest level violent act will be similar to an explicit ranking task of the three violent acts.

Pilot Study 1a recruited participants via a circular e-mail to a participant panel at Queen Mary University of London,  $N = 66$  (55 women, 10 men), with a mean age of  $M = 21.56$  years ( $SD = 7.52$ ), ranging from 18–67 years. All participants were a mixture of undergraduate and postgraduate students attending Queen Mary University of London. There was no financial compensation for taking part.

Pilot Study 1a was presented as an online study using the platform Qualtrics. There were a total of five questions presented

to participants. They involved two demographic details (age, gender), and a question regarding whether they had previously been a victim of crime in which responses were provided in a yes/no format. All participants were then presented with the ranking task, and then the trade-off task. For the ranking task participants were given a list of three violent acts (punched in the face, verbally assaulted, stabbed in the back), and were asked to indicate by allocating a number (1–3) to each of the violent acts (they could only allocate one number to each until all were used up); where 1 indicated the highest level of violence, 2 indicated moderate level of violence, and 3 indicated the lowest level of violence. After this they were then presented with the trade-off task in which they were told “please indicate in the box below how many Punch(s) in the face would someone have to receive to be equivalent to 1 stab in the back?” This was followed by, “please indicate in the box below how many verbal assault(s) would someone have to receive to be equivalent to 1 stab in the back?” The following alphas were reported as acceptable for scale reliability of responses to the trade-off task ( $\alpha = .89$ ), and for ordinal responses to the ranking task ( $\alpha = .85$ ).

The ranks for the three violent acts for each study were entered into a Kendall’s coefficient of concordance ( $W$ ) analysis, the mean rank scores for the three studies are presented in Figure 1. Study 1a revealed a concordance of  $W = .317$ , (Spearman’s  $\rho = .302$ , weak correlation), and a rejection of the null that there is no agreement  $\chi^2(2) = 41.84$ ,  $p < .001$ . The ordering of the violent acts from least to most severe were: punched in the face, verbally assaulted, and stabbed in the back (see Figure 1).

Both Krustall-Wallis  $H$  and Jonckheere-Terpstra tests were used to examine gender differences in the ranks, and for Study 1a,

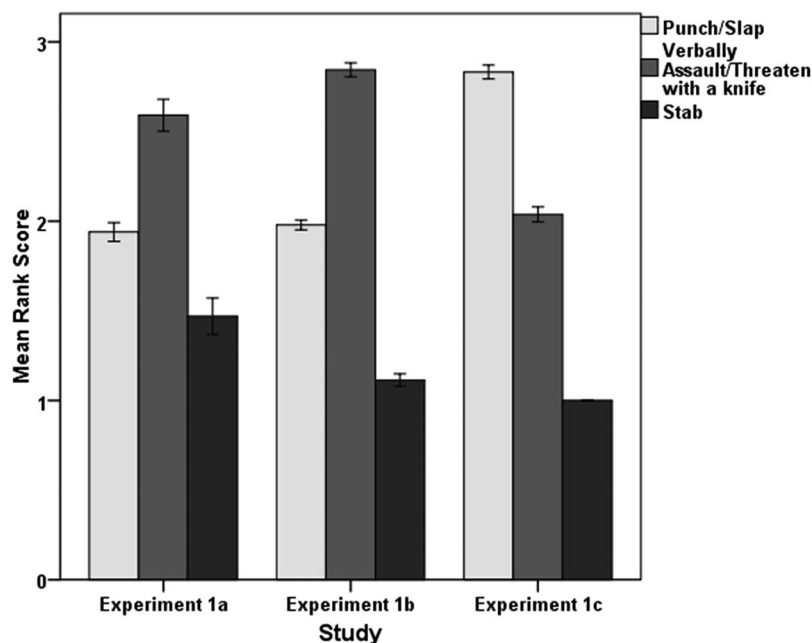


Figure 1. The mean rank scores ( $SE \pm 1$ ) for each of the three different violent acts presented in Studies 1a, 1b, and 1c. The rank positions were as follows: the lower the value the more severe the act. The bars also indicate the level of severity of the violent acts based on shading, from least severe (light gray) to most severe (very dark gray).

whether or not past experiences of being a victim of crime, but there was no significant difference in ranks by gender or past experience of crime. A Shapiro-Wilk test of normality revealed that for each trade-off that was given (Trade-off 1: punch against stab in the back, Trade-off 2: verbal assault against stab in the back) the values participants reported for each trade-off significantly deviated from normality. Log transformations reduced the deviation from normality but the responses still deviated significantly from normal. Analyses conducted before and after transformation revealed no differences in the pattern of results; the results are reported for the log transformed data. Mauchley's test indicated that for each study the assumption of sphericity had been violated,  $p = .0001$ , therefore Greenhouse-Geisser corrected tests are reported. A repeated measures analysis of variance (ANOVA) with age and gender as covariates revealed that there were differences in the trade-off scores between Trade-off 1 and Trade-off 2,  $F(1, 57) = 12.47$ ,  $p = .001$ ,  $\eta_p^2 = .18$ . Age and gender did not significantly impact the pattern of trade-offs. Looking at the mean logs, people needed to experience fewer verbal assaults ( $M = .95$ ,  $SD = .71$ ) than punches in the face ( $M = 1.42$ ,  $SD = 1.60$ ), when traded off against one stab. Thus, the trade-off task indirectly revealed the following orderings: punched in the face, verbally assaulted, and stabbed in the back.

The findings from Pilot Study 1a suggest that the most common ordering of the violent acts from least to most severe for both the direct rank ordering task and the indirect trade-off task was the same: punched in the face, verbally assaulted, and stabbed in the back. Age and gender did not significantly impact responses to the ranking and trade-off tasks.

### Pilot Study 1b

The purpose of Pilot Study 1b was to establish the reliability of the findings reported in Pilot Study 1a, and therefore it served as a replication. Three minor changes were made. The response format regarding the questions probing whether participants had been victims of crime were changed from yes/no to a scale response to enable a more nuanced response to the question. One of the violent acts was changed in order to avoid potential ambiguities, so "stabbed in the back" was replaced with "stabbed in the side of the body." Also, two additional questions were included that focused on the validity of asking people to perform ranking tasks on violent acts.

Pilot Study 1b recruited participants via Prolific Academic, an online crowd-sourcing method. All participants were United Kingdom residents and United Kingdom nationals, and the age restrictions were 18–70.  $N = 141$  (62 women, 79 men), with a mean age of  $M = 39.94$  years ( $SD = 11.28$ ) ranging from 18–67 years. The educational background of the participants ranged from high school level to postgraduate level. Participants were financially compensated for their time (70 cents).

Study 1b involved some minor changes to the materials and procedure used in Pilot Study 1a. Along with their age, gender, and educational background, participants were required to answer on a scale from 1 (*definitely yes*) to 5 (*definitely no*) regarding having been a victim of crime. The ranking and trade-off tasks were the same, but the violent act, stabbed in the back, was changed to stabbed in the side of the body; the rationale for this was that stabbed in the back may have been ambiguous and respondents

could have interpreted it as either a treacherous act or a literal act. In addition, in Study 1b participants were asked two further questions that probed their view of the validity of the questions with which they were confronted. They were asked: do you think that in reality it is possible to rank violent acts in order of their levels of violence? Do you think that for the purposes of crime statistics violent acts should be ranked in order of levels of violence? Answers were given on a scale from 1 (*definitely yes*) to 5 (*definitely no*). The following alphas were reported as acceptable for scale reliability of responses to the trade-off task ( $\alpha = .71$ ), validity questions ( $\alpha = .66$ ), and for ordinal responses to the ranking task ( $\alpha = .75$ ).

The ranks for the three violent acts for each study were entered into a Kendall's coefficient of concordance ( $W$ ) analysis; the mean rank scores for the three studies are presented in Figure 1. Study 1b revealed a concordance of  $W = .791$ , ( $\rho = .789$ , strong correlation), and a rejection of the null that there is no agreement  $\chi^2(2) = 222.98$ ,  $p < .000005$ . The rank ordering of the violent acts from least to most severe were punched in the face, verbally assaulted, and stabbed in the side of the body. Both Krustall-Wallis  $H$  and Jonckheere-Terpstra tests failed to reveal gender differences in the ranks. The same analysis was conducted of trade-off responses (Trade-off 1: punch against stab in the side, Trade-off 2: verbal assault against stab in the side) for Pilot Study 1b as used in Pilot Study 1a. For the trade-off task a repeated measures ANOVA with age, gender, and educational background were entered as covariates, and revealed that there were differences in the trade-off scores between Trade-off 1 and Trade-off 2,  $F(1, 137) = 20.10$ ,  $p = .00002$ ,  $\eta_p^2 = .12$ . Age and gender (based on median splits) did not significantly impact the pattern of trade-offs. Also, educational background did not significantly predict responses. Looking at the mean logs, people needed to experience fewer verbal assaults ( $M = .83$ ,  $SD = .46$ ) than punches in the face ( $M = 1.98$ ,  $SD = 1.46$ ), when traded off against one stab. Thus, the trade-off task indirectly revealed the following orderings: punched in the face, verbally assaulted, and stabbed in the side of the body.

Pilot Study 1b also examined attitudes toward using ranks of violent acts as a way to assess the order of severity of violent acts. Participants were asked their views on whether it is, in fact, possible in reality to rank order violent acts ( $M = 2.15$ ,  $SD = 1.02$ ), and whether this type of measurement should be carried out ( $M = 2.27$ ,  $SD = .98$ ); for both tasks the scale ranged from 1 (*definitely yes*) to 5 (*definitely no*). In addition, the judgments regarding attitudes toward using ranks of violent acts in assessment were correlated with the responses to whether participants had judged themselves to be victims of violent crimes. There was no evidence of a correlation between the extent to which a participant had been a victim of a crime and their attitudes toward using ranking tasks as methods of assessment of violent acts.

The findings from Pilot Study 1b suggest that the most common ordering of the violent acts from least to most severe for both the direct rank-ordering task and the indirect trade-off task was the same: punched in the face, verbally assaulted, and stabbed in the side of the body; this also corresponds with the rank ordering of violent acts reported in Pilot Study 1a. Consistent with Pilot Study 1a, this study also found that age and gender as well as educational background did not significant impact responses in the trade-off and ranking tasks. Finally, the mean scores for both questions assessing the validity of the



tasks presented to participants revealed positive attitudes toward using rank-ordering tasks as a method of assessment for severity of violent acts.

### Pilot Study 1c

The purpose of Pilot Study 1c was to further assess the reliability of the pattern of consistency in the response to the measures developed in Pilot Study 1a and 1b, but in this case extended to new violent acts. Pilot Study 1c was identical to Pilot Study 1b, with the exception that three new violent acts were presented to participants, but the actual trade-off and ranking tasks were the same as the previous two pilot studies. Study 1c was an adapted version of Study 1b, and was also launched via Prolific Academic. The same criteria were applied for recruitment of this sample, with a total of 107 (47 females, 60 men); the mean age was 32.69 years ( $SD = 9.75$ ) ranging from 18–63 years. Participants were financially compensated for their time (70 cents). With the exception that the acts of violence for the ranking and trade-off task were **slapped in the face, threatened with a knife, and stabbed in the side of the body**, in all other respects the procedure and materials used in Pilot Study 1c were identical to Pilot Study 1b. The following alphas were reported as acceptable for scale reliability of responses to the trade-off task ( $\alpha = .98$ ), validity questions ( $\alpha = .62$ ), and for ordinal responses to the ranking task ( $\alpha = .81$ ).

For the ranking task Pilot Study 1c revealed a concordance of  $W = .870$ , ( $p = .867$ , strong correlation), and a rejection of the null hypothesis that there is no agreement  $\chi^2(2) = 186.23$ ,  $p < .000005$ . The rank ordering of the violent acts from least to most severe were slapped in the face, threatened with a knife, and stabbed in the side of the body (see Figure 1). The same analyses used in Pilot Studies 1a and 1b were used to examine responses to the trade-off task in Pilot Study 1c (Trade-off 1: slap against stab in the side, Trade-off 2: threaten with a knife against stab in the side). For the trade-off task a repeated measures ANOVA with age, gender, and educational background as covariates revealed that there were differences in the trade-off scores between Trade-off 1 and Trade-off 2,  $F(1, 100) = 3.16$ ,  $p = .02$ ,  $\eta_p^2 = .01$ . Age, gender and educational background did not significantly impact the pattern of trade-offs. Looking at the mean logs, people needed to experience fewer threats with a knife ( $M = 1.89$ ,  $SD = 4.01$ ) than slaps in the face ( $M = 2.39$ ,  $SD = 5.25$ ), when traded off against one stab. Thus, the trade-off task indirectly revealed the following orderings: slapped in the face, threatened with a knife, and stabbed in the side of the body.

Looking at participants' responses to their attitudes to using ranking tasks to assess severity of violent acts, they showed similar positive attitudes as revealed in Pilot Study 1b. For the first question they were asked whether it is, in fact, possible in reality to rank order violent acts ( $M = 2.34$ ,  $SD = 1.01$ ), and for the second they were asked whether this type of measurement should be carried out ( $M = 2.37$ ,  $SD = .88$ ); in which for both tasks the scale ranged from 1 (*definitely yes*) to 5 (*definitely no*).

The findings from Pilot Study 1c suggest that the most common ordering of the violent acts from least to most severe for both the direct rank-ordering task and the indirect trade-off task was the same: slapped in the face, threatened with a knife, and stabbed in the side of the body. Consistent with Pilot Studies 1a and 1b, there was no evidence to suggest that age, gender, and educational

background influence responses in the trade-off and ranking tasks. As with Pilot Study 1b, Pilot Study 1c also found that the responses to the validity questions revealed positive attitudes toward using rank-ordering tasks as a method of assessment for the severity of violent acts, and there was no correlation with whether participants had been victims of a crime. Overall, across all three pilot studies, the findings support the main purpose of this investigation, which is to demonstrate that people provide similar patterns of judgments regarding the level of severity of violent acts, indexed through direct and indirect judgment tools.

### Study 2a

The purpose of Study 2 was to provide a more focused examination of a more difficult version of the ranking task presented in the pilot studies. In order to fully focus participants' attention, participants were asked to simply rank order eight violent acts according to their level of severity. The eight violent acts were used in order to investigate support for  $H_1$ .

### Method

**Participants.** Study 2a comprised  $N = 35$  (19 women, 16 men). Participants were multinational (U.K. = 15, Spanish = 4, German = 5, U.S. = 6, Chinese = 2, Australian = 2, Dutch = 1) that volunteered to complete an online study presented on Qualtrics. The samples were recruited opportunistically via a circular e-mail distributed to various London judgment and decision-making university forums. The sample had a mean age of  $M = 27.28$  years ( $SD = 6.08$ ), ranging from 19–44 years. All participants were educated at graduate level or above, and were either in higher education in the United Kingdom or were seeking to continue in higher education in the United Kingdom. The samples were all proficient at English. Participants did not receive financial compensation for taking part.

**Materials.** Participants were presented with the questions online via the online platform Qualtrics. Participants were given three questions in total. They were asked to type their age in a text box, and provide details of their gender (male, female, rather not say), and to provide their responses regarding the eight violent acts by ranking them. This question appeared on a separate screen to the demographic questions. Participants were presented with a list of eight violent acts (choke, head-butt, kick, punch, slap, spit, stab, threaten with a knife) in random order on a small screen. Responses were provided by moving the items from the list into a separate box on screen, in which they could order the items by dragging and dropping them into the appropriate slot from least violent at the top, to most violent at the bottom of the list. The alpha for ordinal responses to the ranking task was acceptable at the level of ( $\alpha = .71$ ).

**Procedure.** Once participants had read the instructions and consented to take part, they were then shown the first screen in which they entered in their basic details and, once completed, they went on to complete the ranking task. Once this was complete, a final screen provided them with details about the basic rationale behind the study.

## Results and Discussion

Using Kendall's coefficient of concordance ( $W$ ) the ranks of all 35 participants were entered into the analysis, which revealed a concordance of  $W = .493$ , ( $p = .478$ , moderate correlation), and a rejection of the null hypothesis that there is no agreement  $\chi^2(7) = 120.81$ ,  $p = .0005$ . (Table 1). Both Krustall-Wallis  $H$  and Jonckheere-Terpstra tests failed to reveal the influence of gender differences in responses to the ranking task. The rank ordering of the violent acts from least to most severe based on Kendall's coefficient was: spit, slap, kick, punch, threaten with a knife, head-butt, choke, and stab. This supports  $H_1$ .

### Study 2b

The purpose of Study 2b was to explore the extent to which the rank ordering of the different violent acts would be influenced by the sample, and to provide further support for  $H_1$ . The sample used in Study 2a was a multinational group residing in the United Kingdom, therefore making it difficult to directly explore the possibility that different nationalities would rank order the eight violent acts differently. Therefore in Study 2b nationality was manipulated. Half the sample was United Kingdom nationals, and the remaining half were Bosnian nationals.

## Method

**Participants.** The sample comprised  $N = 65$  in total,  $n = 30$  were United Kingdom-born participants (12 females, 18 male) and  $n = 35$  were Bosnian-born participants (22 females, 17 male). They volunteered to complete an online study presented on Qualtrics. The United Kingdom sample had a mean age of 37.23 years ( $SD = 10.9$ ) ranging from 18–68 year, and the Bosnian sample had a mean age of 33.62 years ( $SD = 9.11$ ) ranging from 23–65. They were recruited via a circular e-mail distributed to participant panels at Queen Mary University of London, and through online university social forums (e.g., Facebook) in Bosnia. Participants did not receive financial compensation for taking part

**Materials.** The materials were identical to that of Study 2a, except that the materials were translated into Bosnian and responses were recorded in Bosnian, but then translated back into English for analysis. The alpha for ordinal responses to the ranking task was acceptable at the level of ( $\alpha = .85$ ).

**Procedure.** The procedure was identical to that used in study 2a.

## Results and Discussion

Using Kendall's coefficient of concordance ( $W$ ), the analysis revealed that, overall, combined across U.K. and Bosnian samples for the ranks of all 65 participants, there was a concordance of  $W = .589$ , ( $p = .582$ , strong correlation), and a rejection of the null that there is no agreement  $\chi^2(7) = 268.12$ ,  $p = .0003$  (see Table 1). Both Krustall-Wallis (KW)  $H$  and Jonckheere-Terpstra (J-T) tests revealed that there were differences in the ranks based on gender (J-T = 2.06,  $p = .039$ , KW = 4.26,  $p = .039$ ). In particular, women ranked spitting higher than men. In addition, we found that there were differences in the ranking of items by sample. For the items, spit (J-T = 3.09,  $p = .002$ , KW = 9.58,  $p = .002$ ), slap (J-T = 2.96,  $p = .003$ , KW = 8.79,  $p = .003$ ), and punch (J-T = 12.46,  $p = .0004$ , KW = 12.46,  $p = .0004$ ), the United Kingdom sample gave lower mean ranks than the Bosnian sample, but for kick (J-T = -2.99,  $p = .003$ , KW = 8.98,  $p = .003$ ), the United Kingdom sample gave a higher mean rank than the Bosnian sample. For the United Kingdom sample the rank ordering of the violent acts from least to most severe based on Kendall's coefficient was: spit, slap, kick, punch, head-butt, threaten with a knife, choke, and stab. For the Bosnian sample the rank ordering of the violent acts from least to most severe based on Kendall's coefficient was: spit, slap, kick, punch, threaten with a knife, head-butt, choke, and stab. The findings from Study 2b suggest that nationality does seem to lead to minor differences in the rank ordering of violent acts. Nevertheless, the ordering of the eight violent acts

Table 1

*Mean (and SD) Rank Orderings of the Eight Violent Acts by Study (Study 2a, 2b, Study 3, Study 4), and by Sample (Austrian, Bosnian, Multinational, UK)*

Study	Spit	Slap	Kick	Punch	Threaten with a knife	Head-butt	Choke	Stab
Study 2a	<b>2.06</b>	<b>2.66</b>	<b>3.46</b>	<b>4.43</b>	<b>4.97</b>	<b>5.26</b>	<b>5.80</b>	<b>7.29</b>
Multinational	(1.81)	(1.49)	(1.84)	(1.42)	(1.32)	(1.96)	(1.64)	(1.73)
Study 2b	<b>1.40</b>	<b>1.87</b>	<b>4.00</b>	<b>4.07</b>	<b>5.73</b>	<b>4.67</b>	<b>6.60</b>	<b>7.67</b>
UK	(.83)	(.64)	(.65)	(1.28)	(1.16)	(1.76)	(1.30)	(.61)
Study 2b	<b>2.31</b>	<b>2.80</b>	<b>3.11</b>	<b>4.94</b>	<b>4.74</b>	<b>5.06</b>	<b>5.74</b>	<b>7.20</b>
Bosnian	(1.80)	(1.51)	(1.94)	(1.39)	(1.74)	(2.00)	(1.70)	(1.71)
Study 3	<b>1.00</b>	<b>1.65</b>	<b>3.97</b>	<b>3.37</b>	<b>5.03</b>	<b>4.23</b>	<b>5.29</b>	<b>6.95</b>
UK		(1.06)	(1.43)	(1.48)	(1.18)	(1.82)	(1.44)	(.94)
Study 3	<b>1.00</b>	<b>2.39</b>	<b>2.50</b>	<b>3.91</b>	<b>4.31</b>	<b>5.02</b>	<b>5.46</b>	<b>6.67</b>
Austrian		(1.55)	(1.55)	(1.30)	(1.39)	(1.26)	(1.03)	(.72)
Study 4	<b>1.73</b>	<b>2.09</b>	<b>3.43</b>	<b>3.53</b>	<b>5.49</b>	<b>5.04</b>	<b>6.60</b>	<b>7.86</b>
UK	(.80)	(1.00)	(.89)	(1.33)	(1.18)	(1.07)	(1.10)	(.39)
Study 4	<b>1.91</b>	<b>2.59</b>	<b>3.98</b>	<b>3.94</b>	<b>4.61</b>	<b>4.44</b>	<b>7.02</b>	<b>7.30</b>
Bosnian	(1.52)	(1.31)	(1.27)	(1.31)	(1.60)	(1.68)	(1.05)	(.77)
Mean rank	<b>1.68</b>	<b>2.34</b>	<b>3.47</b>	<b>4.04</b>	<b>4.98</b>	<b>4.84</b>	<b>6.05</b>	<b>7.27</b>
position overall	(1.34)	(1.34)	(1.55)	(1.48)	(1.44)	(1.70)	(1.47)	(1.19)

Note. Means indicated by bold type; SD indicated by lightface type.

from both samples was generally consistent with the ordering, as predicted in  $H_1$ .

### Study 3

The purpose of Study 3 was to extend the work in Study 2, but instead of an explicit ranking task, participants were required to perform a simple trade-off task using the same eight violent acts used in Study 2. As a result Study 3 examines  $H_2$ . A simple manipulation was introduced which involved presenting the violent acts in such a way that participants were required to perform trade-offs. That is, they were asked to estimate the number of acts required to make a slap equivalent to the other seven violent acts; that is, “how many slaps would be required in order to be equivalent to (X)?”, where X may take the values: 1 spit, 1 kick, 1 punch, 1 head-butt, 1 threat with a knife, 1 choke, 1 stab.

### Method

**Participants.** Study 3 comprised a total of  $N = 58$  participants, of which  $n = 31$  participants were United Kingdom nationals residing in the United Kingdom (13 women, 18 men), and  $n = 27$  Austrians residing in Austria (19 women, 18 men). They volunteered to take part in the online task presented via Qualtrics. The United Kingdom sample had a mean age of 31.71 years ( $SD = 9.04$ ) with a range from 20–61 years. The Austrian sample had a mean age of 34.26 years ( $SD = 12.90$ ) with a range from 21–67 years. The United Kingdom sample was recruited via a circular e-mail distributed to participant panels at Queen Mary University of London. The Austrian sample was also recruited via email through Facebook. No participant received financial compensation for taking part.

**Materials.** Participants were presented with the questions online via the online platform Qualtrics. Participants were given three questions in total. They were asked to type their age in a text box,

provide details of their gender (male, female, rather not say), and to provide their responses regarding the seven trade-offs they were required to perform. For the trade-off task, just as those presented in the pilot study, participants were asked to estimate the number of acts required to make a slap equivalent to the other seven violent acts; that is, “how many slaps would be required in order to be equivalent to (X)?”, where X is 1 spit, 1 kick, 1 punch, 1 head-butt, 1 threat with a knife, 1 choke, or 1 stab. They were required to give a response for each trade-off, with the lowest value being 1. Participants had to perform seven trade-offs in which the violent acts that were compared against all other violent acts were slap (i.e., slap (?) = 1 spit; slap (?) = 1 kick; slap (?) = 1 punch; slap (?) = 1 head-butt; slap (?) = 1 threat with a knife; slap (?) = 1 choke; slap (?) = 1 stab). Each trade-off was performed separately on screen for each pairing, and the order of presentation was randomized for each participant. For the Austrian sample the materials were identical to those presented to the United Kingdom sample. They were translated into German and responses were recorded in German, but then back-translated into English in order to carry out analyses. The alpha reported as acceptable for scale reliability of responses to the trade-off task was ( $\alpha = .82$ ).

**Procedure.** Once participants had read the instructions and consented to take part, they were then shown the first screen in which they entered their basic details and, once completed, they then proceeded to the trade-off task. An example of how to perform the trade-off task was provided along with an explanation of how to respond on screen. Once this was completed, a final screen provided them with details about the basic rationale behind the study.

### Results

The mean scores for the estimated trade-offs are presented in Figure 2. A Shapiro-Wilk test of normality revealed that for each violent act the responses significantly deviated from a normal

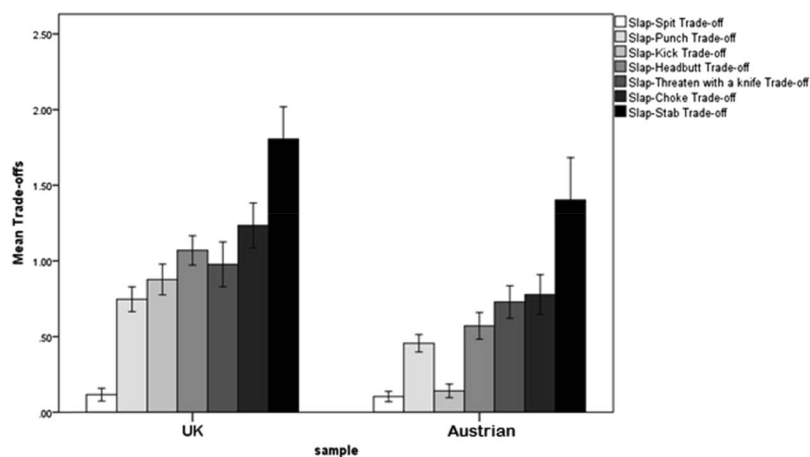


Figure 2. The mean ( $SE \pm 1$ ) trade-off estimates for each of the seven trade-offs presented to participants in Study 3. Participants were required to trade-off each violent act against a slap. The estimates were based on the number of slaps that needed to be performed to make them equivalent to each of the seven violent acts (spit, punch, kick, headbutt, threaten with a knife, choke, stab). Each estimate from each participant was log transformed (base 10), and the mean estimates for each trade-off are presented by sample. The darker the bar, the more severe the act being traded-off the slap. As the bars increase in height, this indicates the increasing number of slaps traded-off against the violent act.

distribution. Log transformations reduced the deviancy from normality but the responses still deviated significantly from normal. Though analyses conducted before and after transformation revealed no differences in the pattern of results, the results are reported for the log-transformed data. Mauchley's test indicated that the assumption of sphericity had been violated,  $\chi^2(20) = 195.98$ ,  $p = .0001$ , therefore Greenhouse-Geisser corrected tests are reported ( $\epsilon = .41$ ). An ANOVA with violent acts (7 trade-offs) as the within-subject variable, and sample (U.K., Austrian) as the between-subject variable was conducted, in which age and gender were entered as covariates. There was a main effect of violent acts,  $F(2.42, 130.97) = 3.02$ ,  $p = .004$ ,  $\eta_p^2 = .07$ . When compared against the trade-off slap-spit, all other trade-offs were significantly different, suggesting that more slaps were estimated as needing to be experienced to be equivalent to a single punch, kick, head-butt, threaten with a knife, choke, or stab. There was a main effect of sample, suggesting that the magnitude of the trade-offs conducted by the Austrian sample were overall lower than the United Kingdom sample,  $F(1, 54) = 9.38$ ,  $p = .004$ ,  $\eta_p^2 = .15$ . There was also a Sample  $\times$  Violent Act interaction,  $F(2.42, 130.97) = 2.43$ ,  $p = .03$ ,  $\eta_p^2 = .04$ . As shown in Figure 3, looking closely at the trade-offs that were carried out, the United Kingdom sample revealed that the increases in magnitude of the trade-offs were in the following order: the lowest trade-off was slap-spit, then, in increasing order, slap-punch, slap-kick, slap-threaten with a knife, slap-head-butt, slap-choke and slap-stab ( $t > 1$ ,  $p < .05$ ). For the Austrian sample, trade-offs performed between slap-spit, slap-kick were no different, but then the trade-offs increased in magnitude in the following order: slap-punch, slap-head-butt, slap-threaten with a knife, slap-choke, and slap-stab ( $t > 1$ ,  $p < .05$ ).

The responses to the trade-off task were used to infer the rank ordering of violent acts from least to most severe. The scoring simply involved using the trade-offs and ordering them in magnitude from least to most, which was taken as an indirect check on

the order of severity of the violent acts obtained directly. With respect to the two least violent acts, spit and slap, if participants responded with a score of 1, then spit and slap were scored as joint 1st in the ranking (as least violent). Thus, for each participant their trade-offs for the pairs of violent acts were used to rank order the acts, and the ranks were subjected to the same Kendall's coefficient of concordance ( $W$ ).

For each participant their trade-off responses were rank ordered, and the ranks were subjected to a Kendall's coefficient of concordance ( $W$ ) test. Separate analyses were conducted on the United Kingdom and the Austrian sample, given the different exposure in the main compensation task. For the United Kingdom sample there was a concordance of  $W = .74$ , ( $\rho = .75$ , strong correlation), and a rejection of the null hypothesis that there is no agreement  $\chi^2(7) = 162.00$ ,  $p = .0001$ . For the Austrian sample there was a concordance of  $W = .78$ , ( $\rho = .80$ , strong correlation), and a rejection of the null hypothesis that there is no agreement  $\chi^2(7) = 147.97$ ,  $p = .0001$ . Krustall-Wallis  $H$  and Jonckheere-Terpstra tests were used to examine gender differences; neither test revealing a significant impact of gender on ranks. For both samples, the ordering of violent acts based on Kendall's  $W$ , from least to most severe, as inferred from the compensation requests, differed but only with regards to two violent acts (see Table 1). For the United Kingdom sample the ordering of violent acts was as follows: slap, spit, kick, punch, threaten with a knife, head-butt, choke, and stab, and for the Austrian sample was: slap, spit, kick, punch, head-butt, threaten with a knife, choke, and stab. Thus the only difference between the two samples was the ordering of the acts, head-butt and threaten with a knife; for all other violent acts the ordering was the same as the United Kingdom sample.

To explore  $H_2$ , the rank ordering of the violent acts, as inferred indirectly from the trade-offs in Study 3, were compared to the direct rank orderings of the violent acts reported in Study 2b for each sample. For the United Kingdom sample, Jonckheere-

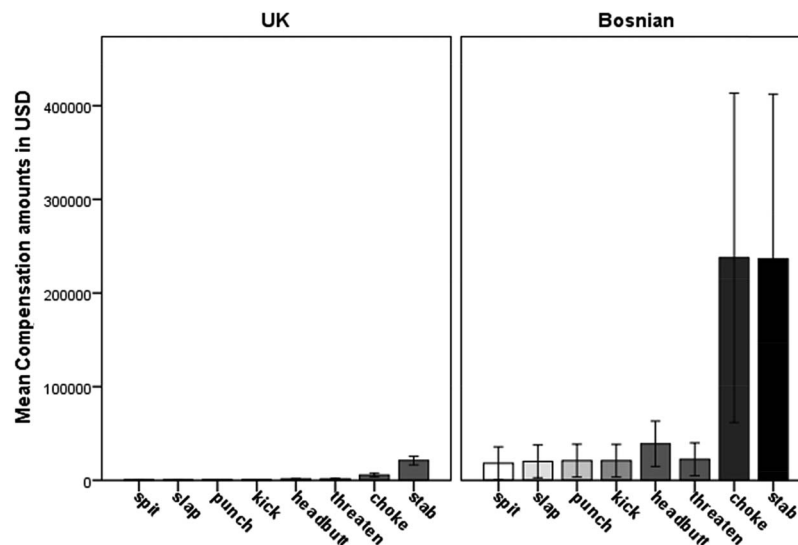


Figure 3. The figure presents the mean ( $SE \pm 1$ ) compensation requests (converted into USD) in Study 4, in which participants (U.K., Bosnia) were asked to report the amount of compensation they would request upon experiencing each of the eight violent acts. As the bars become darker, this indicates the violent acts increasing in severity. As the bars increase in height, this indicates the increasing amount of compensation in USD.



Terpstra tests revealed that the only differences between Study 3 and Study 2b were based on the violent acts slap ( $J-T = 4.31, p = .04$ ) and kick ( $J-T = 15.20, p = .0003$ ). For the Austrian sample, the same analysis revealed that the only difference between Study 3 and Study 2b was based on the rank positioning of the violent act slap ( $J-T = 5.12, p = .02$ ). With the exception of these differences, the inferred ordering of the violent acts from the trade-off task from Study 3 was broadly similar with the directly measured orderings of the violent acts in Study 2b.

### Study 4

The purpose of Study 4 was to use a different judgment task to indirectly examine the rank ordering of the same eight violent acts presented in Studies 2a, 2b, and 3. As a result, Study 4 was able to investigate  $H_3$ . Participants in this study were simply required to provide estimates of the financial compensation they would require to be exactly compensated for being subjected to each of the eight violent acts. Keeping to the same design as Study 2b and Study 3, the present study also used two different nationalities (U.K., Bosnian) for comparison purposes with the other studies included in this investigation.

### Method

**Participants.** Study 4 comprised a total of  $N = 68$  participants, of which  $n = 35$  participants were United Kingdom nationals residing in the United Kingdom (16 women, 19 men), and  $n = 33$  Bosnian nationals residing in Bosnia or abroad (19 females). All volunteered to take part in the online task. The United Kingdom sample had a mean age of  $M = 30.51$  years ( $SD = 9.67$ ) with a range from 18–51 years. The Bosnian sample had a mean age of  $M = 34.61$  years ( $SD = 5.05$ ) with a range from 21–45 years. The United Kingdom sample was recruited via a circular e-mail distributed to participant panels at Queen Mary university of London. The Bosnian sample was also recruited via email through Facebook. No participant received financial compensation for taking part.

**Materials.** Participants were presented with the questions online via the online platform Qualtrics. Participants were given three questions in total. They were asked to type their age in a text box, and provide details of their gender (male, female, rather not say), and to provide their responses regarding compensation required for experiencing each of the eight violent acts. They were simply told that for each of the eight violent acts presented to them they would be required to report the compensation in British pounds (for the Bosnian sample the same instructions were given, but the amount of compensation was in Euros) that they would require if they had experienced the violent act. Each estimate of compensation was performed separately on screen for each violent act, and the order of presentation of the violent act was randomized for each participant. For the Bosnian sample the materials were identical to those presented to the United Kingdom sample. They were translated into Bosnian and responses were recorded in Bosnian, but then translated back into English for analyses. The alpha reported as acceptable for scale reliability of responses to the compensation task was ( $\alpha = .94$ ).

**Procedure.** Once participants had read the instructions and consented to take part, they were then shown the first screen in

which they entered in their basic details and, once completed, they then proceeded to the compensation task. Once this was completed, a final screen provided them with details about the basic rationale behind the study.

### Results and Discussion

Both currencies (GBP and Euros) were converted into USD (conversion was carried out as appropriate at the time of study). The mean and standard deviation for requested compensation for each violent act is presented in Figure 3. The scores converted into USD were then used as a basis to conduct the following analyses. A Shapiro-Wilk test of normality revealed that for each violent act the responses significantly deviated from a normal distribution. Log transformations reduced the deviancy from normality but the responses still deviated significantly from normal. Though analyses conducted before and after transformation revealed no differences in the pattern of results, the results are reported for the log-transformed data. Mauchley's test indicated that the assumption of sphericity had been violated,  $\chi^2(27) = 239.27, p = .0005$ , therefore Greenhouse-Geisser corrected tests are reported ( $\epsilon = .44$ ).

A repeated measures ANOVA with age and gender as covariates show a main effect of violent acts on the amount of compensation requested. This suggests that people asked for different amounts depending on the violent act that had been performed,  $F(3.11, 190.26) = 6.71, p = .0008, \eta_p^2 = .01$ . All pairwise comparisons were significantly different with the exception of spit and slap,  $t(66) = -.31, p = .75$ , kick and punch,  $t(66) = .93, p = .35$ . There was a main effect of sample,  $F(1, 61) = 5.98, p = .02, \eta_p^2 = .09$ , suggesting that the United Kingdom sample required less compensation overall ( $M = 3,828.85$  USD,  $SD = 12,445.71$  USD) as compared with the Bosnian sample ( $M = 76,981.80$  USD,  $SD = 515,465.25$  USD). There was also a significant Sample  $\times$  Violent Act interaction,  $F(3.11, 190.26) = 2.75, p = .0009, \eta_p^2 = .04$ . Independent sample  $t$  tests were conducted, and revealed that for slap,  $t(65) = 2.85, p < .01$ , punch,  $t(65) = 2.76, p < .01$ , kick,  $t(65) = 3.07, p < .01$ , and choke,  $t(65) = 2.65, p < .01$ , the United Kingdom sample required less in compensation than the Bosnian sample. No other analyses or interactions were significant.

The responses to the compensation task were used to infer the rank ordering of violent acts from least to most severe. For each participant their requests for compensation by violent act were rank ordered, and the ranks were subjected to a Kendall's coefficient of concordance ( $W$ ) test. Separate analyses were conducted on the United Kingdom and the Bosnian samples, given the differences revealed in the main compensation task. For the United Kingdom sample there was a concordance of  $W = .814, (p = .81, \text{strong correlation})$ , and a rejection of the null hypothesis that there is no agreement  $\chi^2(7) = 150.56, p = .0001$ . For the Bosnian sample there was a concordance of  $W = .694, (p = .70, \text{strong correlation})$ , and a rejection of the null hypothesis that there is no agreement  $\chi^2(7) = 199.37, p = .0001$ . For both samples, the ordering of violent acts based on Kendall's  $W$ , from least to most severe, as inferred from the compensations, were identical (spit, slap, kick, punch, threaten with a knife, head-butt, choke, and stab; see Table 1). In addition, both Krustall-Wallis  $H$  and Jonckheere-Terpstra tests were used to examine gender differences, neither revealing a significant impact of gender on the ranks.

To explore  $H_3$ , the rank orderings of the violent acts, as inferred indirectly from the compensation task, were compared to the direct rank orderings of the violent acts reported in Study 2b for each sample. For the United Kingdom sample, Jonckheere-Terpstra tests revealed that the only differences between Study 4 and Study 2b were based on the violent acts punch ( $J-T = 6.09, p = .01$ ) and choke ( $J-T = 3.86, p = .049$ ). For the Bosnian and Austrian samples, the same analysis revealed that the only difference between Study 4 and Study 2b was based on the relative ranks of the acts slap ( $J-T = 11.35, p = .001$ ), kick ( $J-T = 6.95, p = .01$ ), and choke ( $J-T = 17.40, p = .0001$ ). With the exception of these differences, the inferred ordering of the violent acts from the compensation tasks in Study 4 were broadly consistent with the directly measured orderings of the violent acts in Study 2b.

**Meta-analysis of Studies 2, 3 and 4.** To further explore the general purpose of the study, and to investigate  $H_1$ , the explicit ranking orderings (Studies 2a and 2b) and inferred rank-ordering scores from the trade-off task (Study 3), and the compensation task (Study 4) were analyzed. The mean ranks scores for these studies are presented in Table 1. The rank scores ( $N = 211$ , women = 110, men 101) were subjected to Kendall's coefficient of concordance ( $W$ ) and revealed a concordance of  $W = .631$ , ( $p = .62$ , strong correlation), and a rejection of the null that there is no agreement  $\chi^2(7) = 931.42, p = .00005$ . This suggests that, despite the differences in the measurement of judgments regarding the eight violent acts, the rank ordering scores support the general purpose of the study, which is to show that there is consistency in the way in which people rank violent acts according to severity. A Jonckheere-Terpstra test for differences in rank scores between the different studies (Study 2a, Study 2b U.K., Study 2b Bosnian, Study 3 U.K., Study 3 Austrian, Study 4 U.K., Study 4 Bosnian) was examined, and revealed differences in the rank scores for the following violent acts: punch ( $J-T = 20.89, p = .0001$ ), head-butt ( $J-T = 9.85, p = .04$ ), and choke ( $J-T = 50.33, p = .0006$ ). The overall rank ordering of acts from least to most severe were as follows: spit, slap, kick, punch, head-butt, threaten with a knife, choke, and stab, which supports  $H_1$ . Gender was also examined separately but was not found to impact significantly the rank scores of the eight violent acts. While the rank ordering of the eight violent acts was generally consistent across Studies 2–4, there were overall differences between the United Kingdom ( $n = 96$ ) and Bosnian samples ( $n = 95$ ). A Jonckheere-Terpstra test of difference in rank scores between the United Kingdom and Bosnian samples revealed differences in the rank scores for the following violent acts: slap ( $J-T = 10.82, p = .001$ ), punch ( $J-T = 5.96, p = .01$ ), kick ( $J-T = 5.14, p = .02$ ) and threaten with a knife ( $J-T = 11.63, p = .001$ ). Nevertheless, the differences in these analyses reflect the difference between the rank-ordering scores and, based on the Kendall's coefficient of concordance ( $W$ ), looking at the actual rank ordering of the eight violent acts for both samples overall, was generally the same.

### General Discussion

The main purpose of the study was met. The general pattern of findings in Studies 1–4 suggest that when people are required to provide judgments regarding the severity of violent acts, either directly or indirectly, the ordering of the acts according to severity is broadly similar. In addition, Study 1 also revealed that peoples'

attitudes toward assessing the severity of violent acts through rank ordering tasks is positive, and they were generally in favor of crime statistics reporting the rank ordering of violent acts according to their severity. Both  $H_2$  and  $H_3$  were generally supported based on comparisons between judgment tasks across studies, suggesting that both the trade-off task and the compensation task were able to indirectly reveal the general pattern of ordering of violent acts as in the direct-ranking task used in Study 2.

This investigation showed that generally gender does not influence judgment behavior, with one exception. In Study 2b, across both United Kingdom and Bosnian samples, the mean rank position for spitting was higher for women compared to men. These differences were not consistently found in the other studies, and so this raises the question of how stable these gender differences are. We speculate though, that one reason why it might be that men regard the relative severity of spitting lower than women, is that they may well be exposed (not necessarily directly—i.e., in TV and films, as well as in sporting contexts) to more examples of other men spitting on other men, whereas there may well be fewer examples in which women are seen either spitting on other men or other women. In fact, this reveals an important issue with regard to base rate information/level of exposure to particular violent events (Fazel & Bjørkly, 2016), and how it can bias judgments about the same events on different criteria (Kahneman & Tversky, 1973). This particular issue may also help to explain the minor but prevailing differences that were identified between the United Kingdom and Bosnian/Austrian samples in Study 2b, Study 3, and Study 4, in which the mean relative weighting of the ranking scores of the violent acts head-butting and threatening with a knife differed between samples, the magnitude of trade-offs was lower in the Austrian sample compared to the United Kingdom, and the magnitude of financial compensation requests in the United Kingdom was lower than the Bosnian sample. While we speculate that the reason for these minor differences may be the result of base rate effects, this requires further empirical testing.

As well as finding strong support for the fact that different judgment tasks revealed similar patterns in the rank ordering of violent acts, what was also of note was the fact that there was general support for  $H_1$ , which predicted, based on previous findings from several studies, the rank orderings of the eight violent acts presented in Studies 2–4. There were of course minor differences in the orderings across studies, but in the main the overall ranks (via direct and indirect measurements) suggested that spit and/or slap were judged the lowest in severity, then were kick and punch, followed by head-butt and threaten with a knife, with the second most severe violent act being choke, and the most severe of the violent acts being a stab. The order also seemed to track the degree to which serious injuries would be sustained, with the exception of being threatened with a knife.

### Limitations

There were several limitations in the present investigation which require further detailed empirical exploration. It is likely that a bigger sample would enable finer-grained analyses with regard to potential individual differences, which we did not tend to find in our studies (e.g., age, gender, educational background). In addition, the violent acts were described as neutrally as possible in Studies 2–4, and this reflected the same typical labels used in

various assessments tools (e.g., CTS, MCVI, MOAS, SVAWS, VRAG). The assumption made here is that participants focused on the acts themselves rather than the consequences of the acts when making their judgments. However, this is an empirical question, and would require running the same studies again and contrasting judgments with and without explicit details regarding the consequences of the violent acts. Moreover, it would be important to assess the extent to which the rank ordering of violent acts is strongly or weakly affected when details of the consequences, and details of the perpetrator of the violent acts, are presented. Also, an important factor that was not assessed was the extent to which the rank orderings of the violent acts would remain constant, if the number of violent acts included in the various judgment tasks increased. A more thorough investigation would include as comprehensive a range of violent acts as possible to capture the full range of violent acts and categories of acts that appear most commonly in assessment tools. Given these limitations, future work is required to determine a more exhaustive response to the main purpose of this investigation, particularly in working toward developing a weighting as well as a ranking system of violent acts.

## Research Implications

The starting point for our study was the aggregation problem, and the implications it entails with regard to accurately reporting the rates of violence between populations, and at different times (Figure 4). This issue, along with a need to develop tools of assessment that are far more nuanced (Grych & Hamby, 2014; Harris et al., 2013) was a key motivation in the present study for examining how lay people judge the severity of different violent acts. The general appeal for using systems that rank and weight violent acts in order of severity has been mounting for some time (Conrad et al., 2010; Goodman et al., 1999; Harris et al., 2013; Hudson & McIntosh, 1981; Peek-Asa et al., 2002; Suglia et al.,

2008). In the present study we use a variety of measurement tools, many of which have been employed by others (e.g., Bishop & Murphy, 2011; Conrad et al., 2010; Hudson & McIntosh, 1981; Peek-Asa et al., 2002, use a trading off task in which people judge the least violent act against all other listed violent acts). The advantage of using simple measurement tools such as ranking, trade-off, and compensation tasks, in which individual items are judged according to their severity, enables researchers to establish whether there is convergence in the rank ordering of violent acts in general. For instance, there were five violent acts (i.e., slap, kick, punch, choke, threaten with a knife) included in Peek-Asa et al.'s (2002) study, and in the present study. The ordering of the violent acts in their study and the present study were the same, which is encouraging to those researchers seeking to show that, in principle, lay people judge violent acts according to their severity in similar ways across studies, samples, and measurement tools.

Tools of assessment that not only record the acts on a scale of severity, or weight them according to severity, may seem insensitive to those that have experienced violence. However, when asked to provide their views on the types of methods used to rank violent acts, in the present study participants tended to be positively disposed toward the use of ranking tasks. Recall that in Studies 1b and 1c participants were asked whether it is possible in real world experiences of violent acts that they can be rank ordered, and whether ranking tasks ought to be included in methods of assessment. Crucially, the findings suggested that their past experiences of crime did not influence their views, which were overwhelmingly positive regarding the implementation of ranking systems to record violent acts. This at least suggests that, in principle, in the view of lay people, there is merit to a process that assesses violent acts according to levels of severity through various measurement tools.

A			B		
Levels of Violence	Year 1	Year 2	Levels of Violence	Year 1	Year 2
Spit	40	55	Spit = Low (violent act *1)	40 (40)	55 (55)
Punch	20	15	Punch = Medium (violent act *5)	20 (100)	15 (75)
Stab	5	2	Stab = High (violent act *10)	5 (50)	2 (20)
Aggregated non-weighted sum	65	72	Aggregated weighted sum	190	150

**Figure 4.** Aggregation problem. This figure illustrates the problem of summing across reports of violent acts with and without a weighted sum. Imagine that the reports of violent acts below are from a Prison X that tracks the levels of violence from year to year. (Panel A) The last row presents the total number of violent acts recorded in Prison X in Year 1 and in Year 2. Without taking into account the different levels of severity of violent acts, what is clear is that by simply aggregating the number of violent acts in Year 1 (total = 65) and again in Year 2 (total = 72), there is an increase in the total level of violence from Year 1 to Year 2. (Panel B) The last row presents the total number of violent acts recorded in Prison X for the same years, based on the same records. However, in this case the violent acts are classified and weighted according to levels of severity (weights are allocated as follows: low severity \*1, medium severity \*5, high severity \*10). The weighted aggregation now shows that, when taking severity of different violent acts into account, the total level of violence in Prison X decreases from Year 1 (total = 190) to Year 2 (total = 150).



## Clinical and Policy Implications

The issues addressed in this paper may be related to the concerns of the World Health Organization (WHO) and other prominent agencies such as the Centers for Disease Control. The role of such organizations is to encourage and enable effective violence reduction measures by other agencies. To this end, in their seminal 2002 report (e.g., Etienne, Linda, James, Anthony, & Rafael, 2002), WHO offers a typology of violence (self-directed, interpersonal, and collective) with further distinctions involving types (physical, sexual, psychological, neglect) and modes (with or without a weapon). However, none of these distinctions are the same as distinctions based purely on the level of violence. For example, self-directed violence may be nonfatal self-harming or suicidal; physical or psychological violence admits many different levels and it is conceivable that a violent act with a weapon may be less violent than one committed with bare hands. This is not to say that the WHO rejects the use or relevance of distinguishing acts according to their levels of violence. For example, where appropriate, they sanction the use of the Conflict Tactic Scales. (e.g., Wessels et al., 2013). Moreover, the overall level of violence is a component of the prominent Global Peace Index (GPI). Different levels of violence are weighted where the weights are obtained from expert panels using a version of the Delphi method (Linstone & Turoff, 2002). This involves an international group of academic and nonacademic experts in the domain of peace, scoring several indices of violence according to qualitative and quantitative scales, which are then normalized (Global Peace Index, 2016). In the present study we made no assumptions about indexes or their form; unlike the linear form presumed for the GPI, this requires empirical testing. What we do show is that by using various quantitative measures the rank position of violent acts is determined by levels of severity, and this could form the starting point from which a more accurate weighting system could be developed.

To further expose the potential implication of our work for policy, we return to the main issue of general interest, which is accurately aggregating acts of violence so as to assess the overall level of violence in a population and whether it has decrease, or increased. Imagine a study had been conducted to examine rates of violence in two prisons in neighboring states. Suppose that two levels of violent acts can be distinguished, low (e.g., slap), and high (e.g., stab), and that in Prison A in one month there are five acts of low violence and one act of high violence, while in Prison B there is one act of low violence and three acts of high violence; the comparison may also be of violence in the same prison in different years rather than of two prisons in the same year. Which prison was the more violent? Any plausible answer surely cannot treat all violent acts as though they are equally violent. An approach which does not differentiate between different levels of violence can therefore be wildly misleading. For instance, the United States Department of Justice "violent crime" metric for the population of the United States 1995–2014, for each year, sums across data on offenses that include murder, rape (legacy definition), robbery, and aggravated assault. Any policy intervention designed to target apparent rising rates of violence is likely to be problematic if the central concern is: is violence increasing? If the different violent acts are not systematically distinguished by their different levels of violence then the question may not be answered convincingly. Our empirical work shows that by considering violent acts according their severity at an item level reveals that lay people do indeed treat violent acts differently according to how they rank them,

perform trade-offs, and the compensation requests they would make. Work of this kind could go some way to improving the accuracy of tools of assessment used to determine the overall level of violence in a population by being sensitive to the relative severity at the item level (individual violent acts) from which a more meaningful aggregation of records of violence could be conducted.

## Conclusion

The aim of the empirical work in this paper was to show that across a variety of direct (ranking task) and indirect measures (trade-off task, compensation task) the findings revealed a general level of agreement in the rank orderings of violent acts according to their severity. Overall, there were no effects of gender, and sample differences that involved comparison between different samples were minor, suggesting that the samples had more in common with regard to their judgments on the severity of violent acts than differences. These findings should be taken as an important step toward showing that through a variety of assessments, it is possible, in principle, to develop tools of assessment regarding violence that are directly informed by lay people's rankings of violent acts, from which a weighting system for those violent acts could be developed. This, in turn, would go some way to improving the systematic treatment of violent acts, which, in turn, will lead to ever more acceptable assessments of rates of violence.

## References

- Bishop, K. C., & Murphy, A. D. (2011). Estimating the willingness to pay to avoid violent crime: A dynamic approach. *The American Economic Review*, 101, 625–629. <http://dx.doi.org/10.1257/aer.101.3.625>
- Bowers, L., Nijman, H., & Palmstierna, T. (2007). The Attempted and Actual Assault Scale (Attacks). *International Journal of Methods in Psychiatric Research*, 16, 171–176. <http://dx.doi.org/10.1002/mpr.219>
- Bui, L., & Farrington, D. P. (2016). First-generation immigrants feel socially excluded and have greater pro-violence attitudes than the native population in England and Wales. *Journal of Aggression, Conflict and Peace Research*, 8, 46–60. <http://dx.doi.org/10.1108/JACPR-08-2014-0134>
- Conrad, K. J., Riley, B. B., Conrad, K. M., Chan, Y. F., & Dennis, M. L. (2010). Validation of the Crime and Violence Scale (CVS) against the Rasch measurement model including differences by gender, race, and age. *Evaluation Review*, 34, 83–115. <http://dx.doi.org/10.1177/0193841X10362162>
- Dowling, N., Suomi, A., Jackson, A., Lavis, T., Patford, J., Cockman, S., . . . Abbott, M. (2016). Problem Gambling and Intimate Partner Violence: A Systematic Review and Meta-Analysis. *Trauma, Violence, & Abuse*, 17, 43–61. <http://dx.doi.org/10.1177/1524838014561269>
- Elmqvist, J., Wolford-Clevenger, C., Zapor, H., Febres, J., Shorey, R. C., Hamel, J., & Stuart, G. L. (2016). A gender comparison of motivations for physical dating violence among college students. *Journal of Interpersonal Violence*, 31, 186–203. <http://dx.doi.org/10.1177/0886260514555130>
- Etienne, G., Linda, L. D., James, A. M., Anthony, B. Z., & Rafael, L. (2002). *World report on violence and health*. Geneva, Switzerland: WHO.
- Fazel, S., & Bjørkly, S. (2016). Methodological considerations in risk assessment research. In J. P. Singh, S. Bjørkly, and S. Fazel (Eds.), *International perspectives on violence risk assessment*. New York, NY: Oxford University Press.
- Fry, D., Anderson, J., Hidalgo, R. J., Elizalde, A., Casey, T., Rodriguez, R., . . . Fang, X. (2016). Prevalence of violence in childhood and adolescence and the impact on educational outcomes: Evidence from the 2013



- Peruvian national survey on social relations. *International Health*, 8, 44–52. <http://dx.doi.org/10.1093/inthealth/ihv075>
- Global Peace Index. (2016). *Quantifying peace and its benefits*. Sydney, Australia: Institute for Economics and Peace.
- Goodman, L. A., Thompson, K. M., Weinfurt, K., Corl, S., Acker, P., Mueser, K. T., & Rosenberg, S. D. (1999). Reliability of reports of violent victimization and posttraumatic stress disorder among men and women with serious mental illness. *Journal of Traumatic Stress*, 12, 587–599. <http://dx.doi.org/10.1023/A:1024708916143>
- Gray, N. S., & Snowden, R. J. (2016). Psychopathy in women: Prediction of criminality and violence in UK and USA psychiatric patients resident in the community. *Psychiatry Research*, 237, 339–343. <http://dx.doi.org/10.1016/j.psychres.2016.01.014>
- Grych, J., & Hamby, S. (2014). Advancing the measurement of violence: Challenges and opportunities. *Psychology of Violence*, 4, 363–368. <http://dx.doi.org/10.1037/a0037886>
- Harris, G. T., Rice, M. E., & Quinsey, V. L. (1993). Violent recidivism of mentally disordered offenders: The development of a statistical prediction instrument. *Criminal Justice and Behavior*, 20, 315–335. <http://dx.doi.org/10.1177/0093854893020004001>
- Harris, S. T., Oakley, C., & Picchioni, M. (2013). Quantifying violence in mental health research. *Aggression and Violent Behavior*, 18, 695–701. <http://dx.doi.org/10.1016/j.avb.2013.07.022>
- Hudson, W. W., & McIntosh, S. R. (1981). The assessment of spouse abuse: Two quantifiable dimensions. *Journal of Marriage and the Family*, 43, 873–888. <http://dx.doi.org/10.2307/351344>
- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80, 237–251. <http://dx.doi.org/10.1037/h0034747>
- Kay, S. R., Wolkenfeld, F., & Murrill, L. M. (1988). Profiles of aggression among psychiatric patients. I. Nature and prevalence. *Journal of Nervous and Mental Disease*, 176, 539–546. <http://dx.doi.org/10.1097/00005053-198809000-00007>
- Krahé, B., & Busching, R. (2015). Breaking the vicious cycle of media violence use and aggression: A test of intervention effects over 30 months. *Psychology of Violence*, 5, 217–226. <http://dx.doi.org/10.1037/a0036627>
- Linstone, H. A., & Turoff, M. (Eds.). (2002). *The Delphi method: Techniques and applications*. Retrieved from <http://www.is.njit.edu/pubs/delphibook>
- Marshall, L. L. (1992). Development of the severity of violence against women scales. *Journal of Family Violence*, 7, 103–121. <http://dx.doi.org/10.1007/BF00978700>
- Nathan, R., Rollinson, L., Harvey, K., & Hill, J. (2003). The Liverpool Violence Assessment: An investigator-based measure of serious violence. *Criminal Behaviour and Mental Health*, 13, 106–120. <http://dx.doi.org/10.1002/cbm.534>
- Peek-Asa, C., Garcia, L., McArthur, D., & Castro, R. (2002). Severity of intimate partner abuse indicators as perceived by women in Mexico and the United States. *Women & Health*, 35, 165–180. [http://dx.doi.org/10.1300/J013v35n02\\_11](http://dx.doi.org/10.1300/J013v35n02_11)
- Prinsloo, M., Matzopoulos, R., Laubscher, R., Myers, J., & Bradshaw, D. (2016). Validating homicide rates in the Western Cape Province, South Africa: Findings from the 2009 Injury Mortality Survey. *South African Medical Journal*, 106, 193–195. <http://dx.doi.org/10.7196/SAMJ.2016.v106i2.10211>
- Sabina, C., Cuevas, C. A., & Cotignola-Pickens, H. M. (2016). Longitudinal dating violence victimization among Latino teens: Rates, risk factors, and cultural influences. *Journal of Adolescence*, 47, 5–15. <http://dx.doi.org/10.1016/j.adolescence.2015.11.003>
- Sherr, L., Hensels, I. S., Skeen, S., Tomlinson, M., Roberts, K. J., & Macedo, A. (2016). Exposure to violence predicts poor educational outcomes in young children in South Africa and Malawi. *International Health*, 8, 36–43.
- Stappenbeck, C. A., Gulati, N. K., & Fromme, K. (2016). Daily associations between alcohol consumption and dating violence perpetration among men and women: Effects of self-regulation. *Journal of Studies on Alcohol and Drugs*, 77, 150–159. <http://dx.doi.org/10.15288/jsad.2016.77.150>
- Steadman, H. J., Mulvey, E. P., Monahan, J., Robbins, P. C., Appelbaum, P. S., Grisso, T., . . . Silver, E. (1998). Violence by people discharged from acute psychiatric inpatient facilities and by others in the same neighborhoods. *Archives of General Psychiatry*, 55, 393–401. <http://dx.doi.org/10.1001/archpsyc.55.5.393>
- Straus, M. A. (1979). Measuring intrafamily conflict and violence: The conflict tactics (CT) scales. *Journal of Marriage and the Family*, 41, 75–88. <http://dx.doi.org/10.2307/351733>
- Suglia, S. F., Ryan, L., & Wright, R. J. (2008). Creation of a community violence exposure scale: Accounting for what, who, where, and how often. *Journal of Traumatic Stress*, 21, 479–486. <http://dx.doi.org/10.1002/jts.20362>
- Tyrer, P., Cooper, S., Herbert, E., Duggan, C., Crawford, M., Joyce, E., . . . Maden, T. (2007). The Quantification of Violence Scale: A simple method of recording significant violence. *International Journal of Social Psychiatry*, 53, 485–497. <http://dx.doi.org/10.1177/0020764007083870>
- Vladutiu, C. J., Casteel, C., Nocera, M., Harrison, R., & Peek-Asa, C. (2016). Characteristics of workplace violence prevention training and violent events among home health and hospice care providers. *American Journal of Industrial Medicine*, 59, 23–30. <http://dx.doi.org/10.1002/ajim.22543>
- Walsh, E., Buchanan, A., & Fahy, T. (2002). Violence and schizophrenia: Examining the evidence. *The British Journal of Psychiatry*, 180, 490–495. <http://dx.doi.org/10.1192/bjp.180.6.490>
- Wessels, I., Mikton, C., Ward, C. L., Kilbane, T., Alves, R., Campello, G., . . . Madrid, B. (2013). *Preventing violence: Evaluating outcomes of parenting programmes*. Geneva, Switzerland: World Health Organization.

Received March 16, 2016

Revision received October 7, 2016

Accepted October 11, 2016 ■