Situational factors shape moral judgments in the trolley dilemma in Eastern, Southern, and Western countries in a culturally diverse sample

Bence Bago[1], Marton Kovacs[2,3], John Protzko[4], Tamas Nagy[2], Zoltan Kekecs[2], Bence Palfi[5,6], Matus Adamkovic[7,8], Sylwia Adamus[9], Sumaya Albalooshi[10], Nihan Albayrak-Aydemir[11], Ilham N. Alfian[12], Sinan Alper[13], Sara Alvarez-Solas[14], Sara G. Alves[15], Santiago Amaya[16], Pia K. Andresen[17], Gulnaz Anjum[18], Daniel Ansari[19], Patrícia Arriaga[20], John Jamir Benzon R. Aruta[21], Alexios Arvanitis[22], Peter Babincak[7], Krystian Barzykowski[9], Bana Bashour[23], Ernest Baskin[24], Luisa Batalha[25], Carlota Batres[26], Jozef Bavolar[27], Fatih Bayrak[28], Maja Becker[29], Benjamin Becker[30], Anabel Belaus[31], Michał Białek[32], Ennio Bilancini[33,34], Daniel Boller[35], Leonardo Boncinelli[36], Jordane Boudesseul[37], Benjamin T. Brown[38], Erin M. Buchanan[39], Muhammad M. Butt[40], Dustin P. Calvillo[41], Nate C. Carnes[41], Jared B. Celniker[42], Christopher R. Chartier[43], William J. Chopik[44], Poom Chotikavan[45], Hu Chuan-Peng[46], Rockwell F. Clancy[47], Ogeday Çoker[48], Rita C. Correia[15], Vera Cubela Adoric[49], Carmelo P. Cubillas[50], Stefan Czoschke[51], Yalda Daryani[52], Job A. M. de Grefte[53,54], Wieteke C. de Vries[55], Elif G. Demirag Burak[56], Carina Dias[57], Barnaby J. W. Dixson[58], Xinkai Du[59], Francesca Dumančić[60], Andrei Dumbravă[61,62], Natalia B. Dutra[63], Janina Enachescu[64], Celia Esteban-Serna[65], Luis Eudave[66], Thomas R. Evans[67], Gilad Feldman[68], Fatima M. Felisberti[69], Susann Fiedler[70], Andrej Findor[71], Alexandra Fleischmann[72], Francesco Foroni[25], Radka Francová[73], Darius-Aurel Frank[74], Cynthia H. Y. Fu[75], Shan Gao[76], Omid Ghasemi[77], Ali-Reza Ghazi-Noori[75], Maliki E. Ghossainy[78], Isabella Giammusso[79], Tripat Gill[80], Biljana Gjoneska[81], Mario Gollwitzer[82], Aurélien Graton[83], Maurice Grinberg[84], Agata Groyecka-Bernard[32], Elizabeth A. Harris[85], Andree Hartanto[86], Widad A. N. M. Hassan[75], Javad Hatami[52], Katrina R. Heimark[87], Jasper J. J. Hidding[10], Evgeniya Hristova[84], Matej Hruška[71], Charlotte A. Hudson[88], Richard Huskey[89], Ayumi Ikeda[90], Yoel Inbar[91], Gordon P. D. Ingram[92], Ozan Isler[93,94], Chris Isloi[95], Aishwarya Iyer[96], Bastian Jaeger[97], Steve M. J. Janssen[98], William Jiménez-Leal[92], Biljana Jokić[99], Pavol Kačmár[27], Veselina Kadreva[84], Gwenaël Kaminski[29], Farzan Karimi-Malekabadi[100], Arno T. A. Kasper[10], Keith M. Kendrick[30], Bradley J. Kennedy[101], Halil E. Kocalar[102], Rabia I. Kodapanakkal[97], Marta Kowal[32], Elliott Kruse[103], Lenka Kučerová[73], Anton Kühberger[104], Anna O. Kuzminska[105], Fanny Lalot[106], Claus Lamm[107], Joris Lammers[72], Elke B. Lange[108], Anthony Lantian[109], Ivy Y.-M. Lau[86], Ljiljana B. Lazarevic[99], Marijke C. Leliveld[55], Jennifer N. Lenz[110], Carmel A. Levitan[111], Savannah C. Lewis[43], Manyu Li[112], Yansong Li[113,114], Haozheng Li[115], Tiago J. S. Lima[116], Samuel Lins[57], Marco Tullio Liuzza[117], Paula Lopes[57], Jackson G. Lu[118], Trent Lynds[119], Martin Máčel[73], Sean P. Mackinnon[120], Madhavilatha Maganti[121], Zoe Magraw-Mickelson[82], Leon F. Magson[88], Harry Manley[45], Gabriela M. Marcu[122], Darja Maslić Seršić[60], Celine-Justine Matibag[123], Alan D. A. Mattiassi[34], Mahdi Mazidi[124], Joseph P. McFall[125], Neil McLatchie[123], Michael C. Mensink[126], Lena Miketta[127], Taciano L. Milfont[128], Alberto Mirisola[79], Michal Misiak[32,129], Panagiotis Mitkidis[74,130], Mehrad Moeini-Jazani[10], Arash Monajem[52], David Moreau[131], Erica D. Musser[132], Erita Narhetali[133], Danielle

P. Ochoa[134], Jerome Olsen[127], Nicholas C. Owsley[135], Asil A. Özdoğru[136], Miriam Panning[137], Marietta Papadatou-Pastou[138], Neha Parashar[96], Philip Pärnamets[139,140], Mariola Paruzel-Czachura[141], Michal Parzuchowski[142], Julia V. Paterlini[43], Jeffrey M. Pavlacic[143], Mehmet Peker[144], Kim Peters[145], Liudmila Piatnitckaia[146], Isabel Pinto[147], Monica Renee Policarpio[148], Nada Pop-Jordanova[81], Annas J. Pratama[133], Maximilian A. Primbs[17], Ekaterina Pronizius[107], Danka Purić[99], Elisa Puvia[79], Vahid Qamari[52], Kun Qian[149], Alain Quiamzade[150,151], Beáta Ráczová[27], Diego A. Reinero[85], Ulf-Dietrich Reips[152], Cecilia Reyna[31], Kimberly Reynolds[153], Matheus F. F. Ribeiro[154], Jan P. Röer[155], Robert M. Ross[156], Petros Roussos[157], Fernando Ruiz-Dodobara[87], Susana Ruiz-Fernandez[158,159], Bastiaan T. Rutjens[160], Katarzyna Rybus[142], Adil Samekin[161], Anabela C. Santos[162,163], Nicolas Say[164], Christoph Schild[165], Kathleen Schmidt[166], Karolina A. Ścigała[167], MohammadHasan Sharifian[52], Jiaxin Shi[68], Yaoxi Shi[168], Erin Sievers[43], Miroslav Sirota[169], Michael Slipenkyj[170], Çağlar Solak[171], Agnieszka Sorokowska[32], Piotr Sorokowski[32], Sinem Söylemez[171], Niklas K. Steffens[172], Ian D. Stephen[173], Anni Sternisko[85], Laura Stevens-Wilson[88], Suzanne L. K. Stewart[101], Stefan Stieger[110], Daniel Storage[174], Justine Strube[175], Kyle J. Susa[176], Raluca D. Szekely-Copîndean[177,178], Natalia M. Szostak[142], Bagus Takwin[133], Srinivasan Tatachari[179], Andrew G. Thomas[180], Kevin E. Tiede[181], Lucas E. Tiong[86], Mirjana Tonković[60], Bastien Trémolière[182], Lauren V. Tunstead[88], Belgüzar N. Türkan[48], Mathias Twardawski[82], Miguel A. Vadillo[50], Zahir Vally[183,184], Leigh Ann Vaughn[185], Bruno Verschuere[160], Denis Vlašiček[60], Martin Voracek[107], Marek A. Vranka[186], Shuzhen Wang[115], Skye-Loren West[88], Stephen Whyte[93,94], Leigh S. Wilton[187], Anna Wlodarczyk[188], Xue Wu[90], Fei Xin[30], Su Yadanar[4], Hiroshi Yama[189], Yuki Yamada[190], Onurcan Yilmaz[191], Sangsuk Yoon[192], Danielle M. Young[193], Ilya Zakharov[194], Rizqy A. Zein[12], Ingo Zettler[195], Iris L. Žeželj[99], Don C. Zhang[196], Jin Zhang[115], Xiaoxiao Zheng[30,197], Rink Hoekstra[198], Balazs Aczel[2]

[1]IAST, Toulouse School of Economics, [2]Institute of Psychology, ELTE Eötvös Loránd University, Budapest, Hungary, [3]Doctoral School of Psychology, ELTE Eotvos Lorand University, Budapest, Hungary, [4]University of California, Santa Barbara, [5]School of Psychology, Universiy of Sussex, Brighton, UK, [6]Department of Surgery and Cancer, Imperial College London, London, UK, [7]Institute of Psychology, Faculty of Arts, University of Presov, [8]Institute for Research and Development of Education, Faculty of Education, Charles University, [9]Institute of Psychology, Faculty of Philosophy, Jagiellonian University, Krakow, Poland, [10]University of Groningen, [11]London School of Economics and Political Science & Open University, [12]Universitas Airlangga, [13]Yasar University, [14]Universidad Regional Amazónica Ikiam, Grupo de Biogeografía y Ecología Espacial (BioGeoE2), [15]Center for Psychology at University of Porto, Faculty of Psychology and Educational Sciences, University of Porto, Portugal, [16]Universidad de los Andes, Department of Philosophy, [17]Radboud University, [18]Department of Psychology, University of Oslo, [19]Department of Psychology and The Brain and Mind Institute, The University of Western Ontario, [20]Iscte-University Institute of Lisbon, CIS-IUL, Portugal, [21]De La Salle University, [22]Department of Psychology, University of Crete, [23]Department of Philosophy, American University of Beirut, [24]Saint Joseph's University, [25]School of Behavioural and Health Sciences, Australian Catholic University, [26]Department of Psychology, Franklin and Marshall College, [27]Department of psychology; Faculty of Arts; Pavol Jozef Šafárik University in

Košice, [28]Baskent University, [29]CLLE, Université de Toulouse, CNRS, UT2J, [30]The Clinical Hospital of Chengdu Brain Science Institute, MOE Key Laboratory for Neuroinformation, School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu, China, [31]Instituto de Investigaciones Psicológicas, Universidad Nacional de Córdoba (IIPsi, Conicet-UNC), [32]Instutute of Psychology, University of Wrocław, [33]IMT School for Advanced Studies Lucca, [34]GAME Science Research Center, [35]University of St. Gallen, [36]University of Florence, Department of Economics and Management, [37]Facultad de Psicología, Instituto de Investigación Científica, Universidad de Lima, [38]Georgia Gwinnett College, [39]Harrisburg University of Science and Technology, [40]Government College University, Lahore, [41]Department of Psychology, California State University San Marcos, [42]Department of Psychological Science; University of California, Irvine, [43]Ashland University, Department of Psychology, [44]Department of Psychology, Michigan State University, [45]Faculty of Psychology, Chulalongkorn University, Thailand., [46]School of Psychology, Nanjing Normal University, Nanjing, China, [47]Department of Values, Technology, and Innovation, Delft University of Technology, [48]Department of Psychology, Faculty of Science and Arts, Pamukkale University, [49]Department of Psychology, University of Zadar, [50]Department of Experimental Psychology, Autonomous University of Madrid, [51]Institute of Medical Psychology, Goethe University, Frankfurt am Main, [52]Department of Psychology, University of Tehran, [53]Faculty of Philosophy, University of Groningen, [54]Faculty of Economics, Econometrics, and Finance, University of Groningen, [55]Marketing department, University of Groningen, [56]Department of Psychology, Koç University, [57]University of Porto, Faculty of Psychology and Educational Sciences, [58]The School of Health and Behavioural Sciences, University of the Sunshine Coast, Australia, [59]Department of Psychology, University of Amsterdam, [60]Department of Psychology, Faculty of Humanities and Social Sciences, University of Zagreb, [61]Alexandru Ioan Cuza University, Iași, [62]George I.M. Georgescu Institute of Cardiovascular Diseases Iași, [63]Laboratório de Evolução do Comportamento Humano, Universidade Federal do Rio Grande do Norte, Brazil, [64]Department of Occupational, Economic and Social Psychology, Faculty of Psychology, University of Vienna, Austria, [65]Division of Psychology & Language Sciences, University College London, London, United Kingdom, [66]School of Education and Psychology, University of Navarra, [67]University of Greenwich, School of Human Sciences, UK, [68]Department of Psychology, University of Hong Kong, [69]Psychology Department, Kingston University London, [70]Vienna University of Economics and Business, [71]Faculty of Social and Economic Sciences, Comenius University in Bratislava, [72]Social Cognition Center Cologne, University of Cologne, [73]Charles University, Faculty of Arts, Department of Psychology, [74]Department of Management, Aarhus University, Aarhus, Denmark, [75]School of Psychology, University of East London, [76]School of Foreign Languages, University of Electronic Science and Technology of China, Chengdu, China, [77]Department of Cognitive Science, Macquarie University, [78]Wheelock College of Education and Human Development, Boston University, [79]University of Palermo, Department of Psychology, Educational Science and Human Movement, [80]Lazaridis School of Business & Economics, Wilfrid Laurier University, [81]Macedonian Academy of Sciences and Arts, [82]Department of Psychology, Ludwig-Maximilians-Universität München, [83]LIP / PC2S, Université Savoie Mont-Blanc, [84]Department of Cognitive Science and Psychology, New Bulgarian University, [85]Department

of Psychology, New York University, [86]School of Social Sciences, Singapore Management University, [87]Instituto de Investigación Científica, Universidad de Lima, [88]Department of Psychology, University of Portsmouth, [89]Department of Communication, Center for Mind and Brain, University of California Davis, [90]Graduate School of Human-Environment Studies, Kyushu University, [91]University of Toronto, [92]Department of Psychology, Universidad de los Andes, [93]School of Economics and Finance, Queensland University of Technology, Brisbane, Australia, [94]Centre for Behavioural Economics, Society and Technology, Brisbane, Australia, [95]Unaffiliated Researcher, [96]Montfort College, [97]Tilburg University, [98]School of Psychology, University of Nottingham Malaysia, [99]University of Belgrade, Faculty of Philosophy, Laboratory for Research of Individual Differences, [100]Department of Psychology, University of Southern California, [101]School of Psychology, University of Chester, [102]Department of Psychological Counseling and Guidance, Muğla Sıtkı Koçman University, [103]EGADE Business School, Instituto Tecnologico y de Estudios Superiores de Monterrey, [104]University of Salzburg, [105]Faculty of Management, University of Warsaw, [106]University of Kent, [107]Department of Cognition, Emotion, and Methods in Psychology, Faculty of Psychology, University of Vienna, Austria, [108]Department of Music, Max Planck Institute for Empirical Aesthetics, Frankfurt am Main, [109]Département de Psychologie, Laboratoire Parisien de Psychologie Sociale, UPL, Univ Paris Nanterre, [110]Department of Psychology and Psychodynamics, Karl Landsteiner University of Health Sciences, [111]Occidental College, [112]University of Louisiana at Lafayette, [113]Reward, Competition and Social Neuroscience Lab, Department of Psychology, School of Social and Behavioral Sciences, Nanjing University, Nanjing, China, [114]Institute for Brain Sciences, Nanjing University, Nanjing, China, [115]Institute of Applied Psychology,School of Public Administration,Northwest university(China), [116]Department of Social and Work Psychology, University of Brasília, [117]"Magna Graecia" University of Catanzaro, Department of Medical and Surgical Sciences, [118]Massachusetts Institute of Technology, [119]Lakehead University, Department of Psychology, [120]Department of Psychology and Neuroscience, Dalhousie University, [121]Ashoka University, [122]Department of Psychology, Lucian Blaga University of Sibiu, [123]Lancaster University, [124]Centre for the Advancement of Research on Emotion, School of Psychological Science, The University of Western Australia, [125]Department of Psychology, State University of New York at Fredonia, [126]Psychology Department, University of Wisconsin-Stout, [127]Max Planck Institute for Research on Collective Goods, [128]School of Psychology, University of Waikato, [129]School of Anthropology & Museum Ethnography, University of Oxford, [130]Social Science Research Institute, Duke University, Durham, United States, [131]School of Psychology and Centre for Brain Research, University of Auckland, [132]Florida International University, [133]Universitas Indonesia, [134]Department of Psychology, University of the Philippines Diliman, [135]Busara Center for Behavioral Economics, [136]Üsküdar University, [137]Maastricht University, [138]National and Kapodistrian University of Athens, Athens, Greece, [139]New York University, [140]Karolinska Institutet, [141]University of Silesia in Katowice, Institute of Psychology, [142]Center for Research on Cognition and Behavior, SWPS University of Social Sciences and Humanities in Sopot, [143]Department of Psychology, University of Mississippi, [144]Ege University, [145]University of Exeter, Department of Management, [146]Lomonosov Moscow State University, [147]University of Porto, Center of Psychology at University of Porto, [148]Department of Psychology, De La

Salle University, [149]Institute of Decision Science for a Sustainable Society, Kyushu University, [150]University of Geneva, [151]UniDistance Switzerland, [152]Research Methods, Assessment, and iScience, Department of Psychology, University of Konstanz, [153]Department of Psychology Nicholls State University, [154]University of Brasilia, [155]Witten/Herdecke University, [156]Department of Psychology, Macquarie University, [157]Department of Psychology, National and Kapodistrian University of Athens, [158]FOM University of Applied Sciences, Germany, [159]Leibniz-Institut für Wissensmedien, Tübingen, Germany, [160]University of Amsterdam, [161]M. Narikbayev KAZGUU University, Nur-Sultan, Kazakhstan, [162]Aventura Social and DESSH, Faculty of Human Kinetics, University of Lisbon; Institute of Environmental Health, Medicine Faculty, University of Lisbon, [163]Iscte-University Institute of Lisbon, CIS-IUL, [164]Prague University of Economics and Business, [165]University of Siegen, [166]School of Psychological and Behavioral Sciences, Southern Illinois University, [167]Aarhus University, [168]Imperial College Business School, Imperial College London, [169]Department of Psychology, University of Essex, [170]Department of Psychology and The Brain and Mind Institute, The University of Western Ontario, [171]Manisa Celal Bayar University, [172]The University of Queensland, [173]Department of Psychology, Macquarie University, Sydney, Australia, [174]Department of Psychology, University of Denver, [175]University of Bonn, [176]California State University, Bakersfield, [177]Romanian Academy, [178]Babeş-Bolyai University, [179]T A Pai Management Institute, [180]Psychology Department, Swansea University, [181]University of Konstanz, [182]Université de Nîmes, APSY-V, [183]Department of Clinical Psychology, United Arab Emirates University, [184]Wolfson College, University of Oxford, [185]Department of Psychology, Ithaca College, United States, [186]Charles University, [187]Department of Psychology, Skidmore College, [188]Universidad Católica del Norte, Escuela de Psicología, Antofagasta, Chile, [189]School of Literature and Human Sciences, Osaka City University, [190]Faculty of Arts and Science, Kyushu University, [191]Kadir Has University, [192]Department of Management and Marketing, University of Dayton, [193]Manhattan College, [194]Psychological Institute of Russian Academy of Education, [195]Department of Psychology, University of Copenhagen, [196]Louisiana State University, [197]Brain Cognition and Brain Disease Institute (BCBDI), Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China, 518055, [198]University of Groningen, Netherlands

Corresponding author: Bence Bago, bencebagok@gmail.com, Institute for Advanced Study in Toulouse - Université Toulouse 1 Capitole, 1, esplanade de l'Université, 31080 Toulouse, France

Abstract

The study of moral judgements often centers on moral dilemmas in which options consistent with deontological perspectives (i.e., emphasizing rules, individual rights, and duties) are in conflict with options consistent with utilitarian judgements (i.e., following the greater good based on consequences). Greene et al. (2009) showed that psychological and situational factors (e.g., the intent of the agent or the presence of physical contact between the agent and the victim) can play an important role in moral dilemma judgements (e.g., trolley problem). Our knowledge is limited concerning both the universality of these effects outside the United States and the impact of culture on the situational and psychological factors of moral judgements. Thus, we empirically tested the universality of the effects of intent and personal force on moral dilemma judgements by replicating the experiments of Greene et al. in 45 countries from all inhabited continents. We found that personal force and its interaction with intention, exert influence on moral judgements in the US and Western cultural clusters, replicating and expanding the original findings. Moreover, the personal force effect was present in all cultural clusters, suggesting it is culturally universal. The evidence for the cultural universality of the interaction effect was inconclusive in the Eastern and Southern cultural clusters (depending on exclusion criteria). We found no strong association between collectivism/individualism and moral dilemma judgements.

Introduction

Moral dilemmas can be portrayed as decisions between two main conflicting moral principles: utilitarian and deontological. Utilitarian (also referred to as consequentialist) philosophies[1] hold that an action is morally acceptable if it maximizes well-being for the greatest number of people (in terms of saved lives, for example). On the other hand, deontological philosophy[2] evaluates the morality of the action based on the intrinsic nature of the action (i.e., the deontological option often reflects greater concern for the individual rights and duties[3]). The dilemma between these two principles plays a prominent role in law and policy-making decisions, ranging from decisions of health budget allocations[4] to the dilemma of self-driving vehicles[5]. This inherent conflict is well illustrated by the so-called trolley problem, which has long interested both philosophers and psychologists. One version of the dilemma is presented as follows[6]:

You are a railway controller. There is a runaway trolley barrelling down the railway tracks. Ahead, on the tracks, there are 5 workmen. The trolley is headed straight for them and they will be killed if nothing is done. You are standing some distance off in the train yard, next to a lever. If you pull this lever, the trolley will switch to a side track and you can save the 5 workmen on the main track. You notice that there are 2 workmen on the side track. So there will be 2 workmen who will be killed if you pull the lever and change the tracks but the 5 workmen on the main track will be saved. Is it morally acceptable for you to pull the lever?

A deontological decision-maker would argue that pulling the lever is morally unacceptable, as it would be murder (Note that deontological principles are often more complicated than this. Some of the deontological rules would allow for killing in this situation. The terms "deontological" and "utilitarian/consequentialist" are labels we use to refer to certain responses). On the other hand, utilitarianism would suggest that it is morally acceptable to pull the lever, as it would maximize the number of saved lives.

In an alternative version of the dilemma, one has to push a man off a footbridge in front of the trolley ("footbridge" scenario). This man will die but will stop the trolley, and the five people in the way of the trolley will be saved. Interestingly, people are less likely to make a decision consistent with utilitarian perspectives in the footbridge scenario compared to the standard switch scenario (We call these "utilitarian" responses but the fact that these decisions are consistent with utilitarianism does not indicate that people gave them out of utilitarian principles; the same is true for "deontological" responses[7,8]). The difference between the utilitarian response rate in those scenarios became the basis of investigations of many influential cognitive theories in the field of moral judgement[3,7–13]. The fact that people respond differently to the two trolley dilemmas was proposed to be explained by people's adherence to the so-called doctrine of double effect[6,9]. A simple version of this doctrine is that harm is permissible as an unintentional side-effect of a good result. This doctrine is the basis of many policies in several countries all around the world concerning issues such as abortion[6], euthanasia[14], international armed conflict regulations[15,16], and even international business ethics[17]. According to this doctrine, it is morally impermissible to bomb civilians to win a war, even if ending the war would eventually save more lives. However, if civilians die

in a bombing of a nearby weapons factory as a side-effect, the bombing is morally acceptable. The way people perceive or act on these moral rules can influence the policies that are accepted or even followed - as we can already see in the case of driverless cars, which sometimes have to decide between sacrificing their own passengers and saving one or more pedestrians[5].

Greene et al.[18] and Cushman et al.[9] , however, argued that the difference in utilitarian response rates cannot simply be explained by the doctrine of double effect. Greene et al. presented evidence for the interaction of the intention of harm (i.e., harm as means or side effect; referring to the doctrine of double effect) and personal force (i.e., whether or not the agent had to use personal effort to kill the victim and save more people) on moral acceptability ratings. More concretely, people were less likely to judge sacrificing one person to save more people as morally acceptable when they had to use their personal force to kill the person *and* the death of this person was required to save more people (this is what is meant by *intending* the harm). Hence, they concluded that people are more sensitive to the doctrine of double effect when they have to use their own physical force. Despite some exceptions[26,27], most of the evidence for this conclusion comes from samples of WEIRD (Western, Educated, Industrialized, Rich, Democratic[23,24]) societies, leaving the question open of whether these effects are psychologically universal[25] or culture-specific.

This study tests three cross-cultural hypotheses:

(1) The effects of personal force on moral judgements are culturally universal.
(2) The interactional effect of personal force and intention on moral judgements is culturally universal.
(3) Collectivism-individualism has a moderating effect on the degree to which personal force and intention affect moral judgements in a way that their effect is stronger in more collectivistic cultures.

The first and second hypotheses, that the effects of personal force and intention on moral judgements are culturally universal, come from their relatedness to interpersonal violence. People seem to exhibit a general tendency to avoid causing violent harms (e.g., murder)[19,20], and they are more likely to perceive actions as violent or harmful when they are supposed to use personal force or intention[3]. As a result, people are more likely to behave in a deontological way when personal force or intention is present in the dilemma. As all cultures regulate interpersonal violence,[21] we expected to find that both intention and personal force, as well as their interaction, have an effect on moral judgements across cultures. The literature seems to be in accordance with these hypotheses. For example, Chinese[25–27] and Russian[28] participants responded similarly to moral dilemmas as Americans and Western Europeans, and even small-scale societies tended to be susceptible to the effect of intention[22,23].

Even though we anticipated that the effect of personal force and intention would emerge universally across cultures, we nonetheless expected cultural differences to moderate these effects. The effect of personal force on moral judgement has been attributed to emotional processes[9,24–26], specifically social emotions (such as guilt, shame or regret)[25,27]. The potential

use of personal force makes people feel guilt or shame before making a decision and, therefore, rating actions that use personal force as morally less acceptable. There is a convincing argument that these social emotions are universal[28–30], despite some cultural variation in their intensity and the social contexts in which they are experienced[28–30]. It has been argued that shame and guilt are more important in interdependent, collectivistic cultures (as their function is argued to be linked to social control). People living in East Asian countries have reported experiencing these emotions more frequently and more intensely[28–30]. Other findings suggest that it is anxiety that mediates the effect of intention and personal force[26], but anxiety (social anxiety in particular) has also been positively associated with collectivism[31], pointing to the same direction. Hence, we hypothesized that people living in collectivistic cultures would judge actions that involve personal force and intention as morally less acceptable than people in individualistic cultures. Utilitarian responding in moral dilemma judgements has also been associated with low levels of empathic concern[32] and people living in collectivistic cultures have been suggested to exhibit higher levels of empathic concern[33,34]. Hence, we predicted that individualism-collectivism would also have an effect on utilitarian responding: collectivists would be less utilitarian in general, due to their higher levels of empathic concern.

In addition to testing our confirmatory hypotheses, we also collected a number of additional country-level as well as individual measures for exploratory purposes. These measures have been previously shown to be related to moral judgement such as economic status[35], individual level individualism-collectivism[35], and religiosity[36]. We also administered an alternative measure of utilitarian responding[37–40].

The present investigation is crucial for advancing the field for the following reasons:

1) The original article has been very influential (515 citations so far), but replicability has not established yet.
2) Our knowledge is scarce on the cultural universality of the effect of personal force and intention in moral judgements.
3) The resulting database (with many types of trolley problems and additional measures) could assist and guide future research and applications on moral thinking.

Overview

In the first part of our study, we tested the universality of the role of personal force in moral judgements with a direct replication of Study 1a conducted by Greene et al.. In their study, the authors found evidence that the application of personal force decreases moral acceptability of the utilitarian action (Hypothesis 1a, 1b). In the second part, we tested the universality of the interactional effect of personal force and intention on moral dilemma judgements, by replicating Study 2 of Greene et al. (Hypothesis 2a, 2b) with partially different moral dilemmas. Furthermore, we tested our hypothesis that collectivism moderates the effect of intention and personal force (Hypothesis 3). In addition, we collected various additional measures for exploratory purposes.

Results

We collected data from 27,502 participants out of 45 countries. Due to our exclusion criteria, we had to exclude 80.6% of the sample for the main analysis (see Table 1 for the various exclusion criteria). Note that, as we registered, we conducted the analysis without excluding the data of the participants who were familiar with the trolley problem (36.2% exclusions), and we also conducted a post-hoc explorative analysis in which we applied no exclusion criteria. All participants were presented with two moral dilemmas that were equivalent in structure but were different in wording: trolley dilemmas and speedboat dilemmas (the former described a situation involving trolley and people on the tracks, the latter described a situation with people on a speedboat and others drowning in the sea). In Study 1, we tested the effect of personal force on moral dilemma judgments (Hypothesis 1a, 1b), while in Study 2, we tested the interaction effect between personal force and intention (Hypothesis 2a, 2b, 3).

The effect of personal force

Findings are represented in Figure 1. To test the effect of personal force on moral judgement, we used one-sided *t*-tests. Consistent with our preregistration, we analysed only the continuous acceptability ratings (scale of 1-9), and not the binary choices. In each cultural cluster, we found at least strong evidence ($BF_{10} > 10$) of an effect of personal force on moral judgement, which implies the effect is culturally universal. The results indicate that, when personal force is seen to be necessary to save more lives, people are less likely to favourably judge a consequentialist outcome (i.e., save more people). The results remained robust across dilemma contexts (i.e., trolley or speedboat version) and when including participants who were very familiar with these trolley-problem type scenarios. Therefore, our results replicated the findings of Greene et al. in the original cultural setting (H1a) and in the Southern and Eastern cultural clusters (H1b). The statistical results are summarised in Table 2.

The interaction effect of personal force and intention

Figure 2 shows when we applied all exclusion criteria, we found strong evidence in the Western cluster (H2a) for the interaction between personal force and intention ($BF_{10} = 1.5*10^{11}$), but moderate inconclusive evidence in the Southern ($BF_{10} = 9.4$) and weak, inconclusive evidence in the Eastern clusters ($BF_{10} = 0.6$) (H2b). More concretely, in the Western cluster, participants judged the acceptability of consequentialist decisions much lower when both personal force and intention had to be applied (i.e., the personal force effect was numerically greater when intention also had to be applied). When we included participants who were familiar with the trolley dilemma, we still found strong evidence in the Western cluster ($BF_{10} = 1.28*10^{30}$) and, interestingly, we also found strong evidence in the Southern cluster ($BF_{10} =3.1*10^{6}$), but the evidence remained weak and inconclusive in the Eastern cluster ($BF_{10} = 2.9$). Although in the preregistration we expected the effect sizes to be smaller when participants familiar with the trolley problem were included, we observed the direct opposite: when including data of participants familiar with the trolley problem, we

found either equivalent or larger effect sizes in all cultural clusters. Notably, the size of the effect almost doubled in the Southern cluster when running the analysis on the sample with familiar and unfamiliar participants included (η_p^2 increased from .014 to .026). All statistical results are presented in Table 3.

On the speedboat dilemmas, we found strong evidence for the interaction in the Western cluster, regardless of the familiarity exclusion ($BF_{\text{all exclusions}} = 222$, $BF_{\text{with familiar}} = 4.8*10^7$). However, we found inconclusive evidence in the Eastern and Southern clusters, both before ($BF_{\text{Eastern}} = 0.4$; $BF_{\text{Southern}} = 0.4$) and after ($BF_{\text{Eastern}} = 0.4$; $BF_{\text{Southern}} = 1.1$) familiarity exclusions. Although our results were consistent in the Western and Eastern clusters for both the speedboat and trolley dilemmas, there was a divergence in the Southern cluster. Specifically, we found strong evidence only for the interaction in the Southern cluster when we included familiar participants in the analysis. In general, in all clusters, the observed effect sizes were smaller on the speedboat than on the trolley dilemma.

In summary, we conclude that we fully replicated the findings of Greene et al. with respect to the interaction of personal force and intention in the Western cluster (H2a) regardless of dilemma context or exclusion criteria. However, the evidence was inconclusive for all analyses of the Eastern cluster. In the Southern cluster, the conclusion is both context-dependent (i.e., the effect was only detectable in the trolley dilemma) and sensitive to exclusion criteria (i.e., the effect was only detectable when familiar participants were included).

To explore whether our results were sensitive to our choice of priors in the Bayesian analysis, we computed Robustness Regions ("RR") that indicate the region of priors within which our inference would remain unchanged. The width of this region shows how robust our inferences are to our selection of priors. The RRs were generally wide for all statistical tests (see Tables 2-3), indicating that our results were not sensitive to our choices of prior. Thus, we would arrive at the same conclusions with any possible prior within the realistic range. One exception to this finding where the final conclusion was prior-dependent can be found in the analysis of the Southern cluster in Study 2. Specifically, if the scale of the prior distribution had been $r = .21$ or higher (instead of $r = .19$), we could have concluded that there was strong evidence for the effect (instead of saying that the test is inconclusive). Here, we would like to stress that we did not reach our registered sample size in this cluster for Study 2 (we registered that for 95% power, we would need 1,800 participants in each cluster of which we only reached 690 - see the Methods for details on sample size estimation). This could explain why our results did not reach our evidence thresholds and remained inconclusive.

Cultural correlates

To test the "effects" of cultural variables, we used linear mixed models predicting moral acceptability ratings from different cultural variables with the random intercept of countries. We tested all five cultural variables one-by-one (i.e., country-level collectivism, and the four

individual-level measures of horizontal and vertical collectivism/individualism), in separate linear models on the data with and without familiarity exclusion.

H3 stated that we expected a three-way interaction between country-level collectivism, intention, and personal force. We first tested this hypothesis on the data with familiarity exclusion applied (see Table 4 for statistical results and Figure 3 for the graphical representation of findings). The results of the country-level collectivism scale were inconclusive (trolley: $BF_{10} = 1.2$; speedboat: $BF_{10} = 0.9$). When analysing the individual-level measures of horizontal and vertical collectivism/individualism, all results were inconclusive. We conducted the same analysis on the sample but this time including participants who were familiar with these types of moral dilemmas, but the results were still inconclusive (trolley: $BF_{10} = 2.2$; speedboat: $BF_{10} = 0.7$). Analysing the individual-level individualism/collectivism measures, we found inconclusive evidence in all the scales. In the Introduction (Stage 1), we also hypothesized that country level collectivism would be associated with decreased overall acceptability of the utilitarian option. This hypothesis was not included in the registered analysis plan. Nevertheless, we added this analysis to the Supplementary Analysis section 3. In short, we found no evidence for the association between country-level collectivism and moral acceptability rates. Interestingly, nevertheless, we found strong evidence for a positive correlation between vertical individualism and moral acceptability ratings.

We conducted the same analysis on the Speedboat dilemmas. Table 4 and Figure 4 presents the findings. Regardless of the familiarity exclusion criteria, we found inconclusive results in all cases.

Exploratory analysis

The effect of intention

We registered that we would test the main effect of intention by comparing the standard switch (no intention) and footbridge switch (intention) dilemmas. We found strong evidence in each cultural cluster and in each dilemma type for the effect of intention ($BF_{10} > 10$). Importantly, the effect of intention remained unchanged even when we included participants who were familiar with moral dilemmas in the sample ($BF_{10} > 10$). Tables 5-6 summarize the findings. As registered, we also tested the effect of physical force on moral judgement. In accordance with Greene et al., we found no evidence for this effect. See details in Supplementary Analysis section 2.1.

No exclusion analysis (post-hoc)

As the exclusion rate was very high in the above analyses (81%), we explored our results while applying no exclusion criteria (including all participants). In Study 1, we found strong evidence for the individual effects of personal force and intention, in each of the three cultural clusters, both in the speedboat and the trolley dilemmas—just as in our main analyses (see Extended Data Figures 1 and 2 for detailed results and data distribution).

For Study 2, Extended Data Figure 3 summarizes the statistical findings. Overall, we can conclude that almost all of our results regarding the effects of personal force and its interaction with intention are not sensitive to our exclusion. Only in the case of the Eastern cluster can we see a difference: without applying exclusions, strong evidence can be found for the effect of personal force and intention in the trolley dilemma, otherwise, we find inconclusive evidence. Here, we can only speculate whether the increased strength of evidence is due to the increased number of participants. The analysis on the speedboat dilemmas yielded the same results with and without exclusions: inconclusive evidence in the Eastern and Southern clusters, and strong evidence in the Western cluster (see Extended Data Figure 4 for the findings on Study 2). Thus, it appears that applying such strong exclusion criteria did not strengthen the replication effort nor substantially alter the inferences we draw about the replicability of the effect of force and intention.

We also conducted the cultural analysis without applying any exclusion criteria and we found that all of the results were inconclusive, with one exception. In the speedboat dilemma, we found moderate evidence that country level collectivism is positively associated with the interaction of personal force and intention (in line with our hypothesis; $BF_{10} = 5.1$; same test for the trolley dilemma: $BF_{10} = 2.8$). We also found moderate evidence ($BF_{10} = 9.8$) that in the trolley dilemma, the interaction between personal force and intention is positively associated with individual-level horizontal collectivism: being higher on horizontal collectivism means a heightened personal force and intention interaction effect size (see Extended Data Figures 5 and 6; same test in the speedboat dilemma was inconclusive: $BF_{10} = 0.54$). Thus, for the moderation of the effect by country-level collectivism, the strict exclusion criteria may have hurt our ability to detect these effects. Although these results appear in line with our prior hypothesis, this analysis was only exploratory, not registered a priori, and hence, should only be interpreted with caution.

As we registered, we added a figure showing the distribution of responses of both subscales of the Oxford Utilitarianism Scale for each country cluster, and also reported means and 95% confidence intervals, as registered. Moreover, we also added a post-hoc analysis correlating each subscales of the OUE with moral acceptability ratings of the moral dilemmas. We found that moral acceptability ratings correlate higher with the “instrumental harm” sub-scale ($r = 0.40 - 0.45$) than with the “impartial beneficence” sub-scale ($r = 0.05 - 0.20$) - with this latter correlation exhibiting somewhat larger cultural variations. Details can be found in the Supplementary Analysis section 2.4.

Discussion

For centuries, philosophers and psychologists have explored the determinants of moral judgments. Moral dilemmas that force life and death decisions help us explore what norms and psychological processes drive our moral preferences. Initially, researchers thought[41,42] that people are simply susceptible to the doctrine of double effects when making moral judgements; harm is permissible if it occurs as an unintentional side-effect of an overall good

outcome. Greene et al.[18], however, showed that the role of using physical force to kill one (and save more) influenced moral judgments even more than did the intentionality of an action.

In this research, we replicated the design of Greene et al.[18] using a culturally diverse sample across 45 countries to test the universality of their results. Overall, our results support the proposition that the effect of personal force on moral judgments is likely culturally universal. This finding makes it plausible that the personal force effect is influenced by basic cognitive or emotional processes that are universal for humans and independent of culture. Our findings regarding the interaction between personal force and intention were more mixed. We found strong evidence for the interaction of personal force and intention among participants coming from Western countries regardless of familiarity and dilemma contexts (trolley or speedboat), fully replicating the results of Greene et al.[18]. However, the evidence was inconclusive among participants from Eastern countries in all cases. Additionally, this interaction result was mixed for participants from countries in the Southern cluster; we only found strong enough evidence when people familiar with these dilemmas were included in the sample and only for the trolley (not speedboat) dilemma.

Our general observation is that the size of the interaction was smaller on the speedboat dilemmas in every cultural cluster. It is yet unclear whether this effect is caused by some deep-seated (and unknown) differences between the two dilemmas (e.g., participants experiencing smaller emotional engagement in the speedboat dilemmas that changes response patterns), or is caused by some unintended experimental confound (e.g., order effect of the presented dilemmas). Furthermore, in the Eastern and Southern clusters, more participants found the dilemmas confusing than in the Western cluster (see Table 2). The increased confusion rates might have played a role behind the fact that we found no evidence for the personal force and intention interaction in the speedboat dilemmas; participants from the Southern and Eastern clusters might have struggled to follow some versions of the speedboat dilemma, as it was originally written for U.S. participants.

Furthermore, we hypothesised that collectivism would enhance the effect of personal force and intention. This prediction was based on the notion that collectivism increases the sensitivity to certain emotions which mediate these effects. We found no evidence for this hypothesis when we executed our preregistered analysis plan. However, in the exploratory analysis (with no exclusion criteria were applied), we found some moderate evidence for the association of country level collectivism in the speedboat dilemma, and individual level horizontal collectivism in the trolley dilemma with the interactional effect of personal force and intention. Since this analysis was not preregistered, these results should be cautiously interpreted.

The interaction between intention and personal force was sensitive to whether we included participants familiar with moral dilemmas. In the Southern cluster, this led to inconclusive evidence regarding the trolley problem, but contrary to our expectations, the size of all of the interaction effects were larger when we included familiar participants in the analysis. This

increase could be due for at least two reasons: (1) familiarity is not the main reason behind the change in response patterns: familiarity correlates with an as yet unknown underlying variable, which induces a selection bias (e.g., educational background); and (2) familiarity is the main reason behind the change in response patterns: for example, being familiar with the trolley problem might have caused people to exhibit a lower emotional response to the problem or caused them to apply different reasoning that ended up affecting their responses. Our results cannot differentiate between the above described explanations (which are not necessarily mutually exclusive).

Although we found no strong evidence for the association between collectivism/individualism and the effects of personal force and intention, future research should test for other cultural variations. There are a number of interesting candidates that we did not examine, including cultural tightness[43] and social mobility[44]. Our database provides opportunities to the field to examine different aspects and cultural moderators of moral judgment.

This research has a number of limitations that future work will need to address. Although we call the personal force effect "universal", it is only universal to the cultures we tested. This puts a limit to the "universality" of the effects: we did not (nor intended to) reach small scale hunter gatherer societies for example. Moreover, while our sample was more diverse and less WEIRD than that of Greene et al.'s research, it consisted of mostly educated individuals from younger age groups with internet access, raising similar concerns (e.g., still Educated and Industrialized, and possibly Rich, though not strictly Western or Democratic). Secondly, the data collection was conducted before and during the COVID-19 pandemic which could have affected the participants' responding behaviour in some way (e.g., moral fatigue). Finally, 81% of the sample was not entered into the main confirmatory analyses because of our exclusion criteria, which might have resulted in unintended selection biases. For example, it is possible that more educated participants were more likely to get excluded due to being familiar with moral dilemmas from college. It is also possible that people with less working memory capacity or poor text comprehension abilities were more likely to be excluded due to the stringent attention checks. This is why we included an exploratory analysis in which we analysed data from all of our participants, without applying any exclusions. Our results on the full sample (no exclusion criteria applied) supported our previous conclusions (that were drawn based on the data with exclusions) except in the cultural analysis, in which we found strong evidence for cultural variations only when no data were excluded. Thus, future work, especially replication work, should take caution when applying stringent exclusion criteria as it may be entirely unnecessary and even hurt the discovery of new effects.

Another limitation of our study might come from the fact that we used a single continuous measure of deontological/utilitarian tendences. Although common in the field, such an approach has been criticized for being overly simplistic and not being able to pick up on more complex response patterns [45,46]. For example, maximizing outcome and rejecting harm are not necessarily symmetrical (as our continuous measure suggests). Hence, an interesting direction

for future research could be to identify whether personal force and intention increase reliance on deontological rules or decrease reliance on consequentialist thinking. Methodological approaches, such as process dissociation, are promising in this regard[40].

Conclusion

With this replication study, we present empirical results about how people around the world make judgments in moral dilemmas that have long interested moral philosophers and psychologists. Empirical studies in this field have been conducted mostly on WEIRD samples, with little attention paid to cultural universality and variations. Our research allows us to avoid some important selection biases by having participants take the survey in their native language from 45 countries. The shared dataset should allow the assessment of different effects on moral dilemma judgments, such as religion or second language effects.

Overall, we found (1) the negative main effects of personal force and intention on moral dilemma judgments is universal; (2) the interaction between intention and personal force was replicated in the Southern and Western clusters, finding people are less likely to support sacrificing one person's life for the sake of saving the lives of several others, if they have both to intentionally engage in an action to do this and to use personal force; and (3) this interaction is not associated strongly with individual nor country-level collectivism/individualism measures.

Method

Participants

A large culturally and demographically diverse sample of participants was recruited from collaborating laboratories through the Psychological Science Accelerator[47]. The data collection team was originally proposed to include 146 labs from 52 countries. All of these participating laboratories obtained IRB approval (verified before the last round of Stage 1 submission). Combined, these labs committed to collect a minimum number of 18,637 participants. More labs were expected to be recruited before data collection commences. Each lab will recruit participants for the study by sending out the survey link along with the consent form to their participant pool, online platforms (such as Mturk), or testing them in the research lab. Due to some dropouts, the data collection team included 140 labs from 45 countries. Eligibility for participation was based on age (≥ 18 years) and being a native speaker of the language of the test (more details on this criterion in the *Controlling for possible confounds* section). Data were collected either from local university participant pools or via data collection platforms (e.g., MTurk). Altogether, 41,090 participants started our survey, and 27,502 finished it whose data were analysed (17961 females, 7956 males, Mean age = 26.0 years, SD = 10.3 years; Study 1: 7744 participants, 4329 females, 2487 males,

Mean age = 26.8 years, SD = 11.1 years; Study 2: 19340 participants, 13632 females, 5469 males, Mean age = 25.8 years, SD = 9.98 years).

We did not collect any identifiable private data during the project that can be linked to individual survey responses. Each lab ascertained the agreement of the local institutional ethical review board with the proposed data collection. This study was conducted in accordance with the Declaration of Helsinki. The IRB approvals are available on our OSF project page: https://osf.io/j6kte/. Participants had to give an informed consent before starting the experiment. Only participants recruited through Mturk or Prolific received monetary compensation.

Materials

Moral dilemmas. We used a total of six trolley dilemmas, namely: *footbridge switch, standard footbridge, footbridge pole, loop, obstacle collide* (taken from Greene et al.), and *standard switch.* All the materials are provided in the Supplementary Methods sections 1-3. Each of these scenarios represents a different condition. For example, in the *standard footbridge* scenario both intention and personal force are required to push the man off the bridge. As in the original experiments, every participant was assigned to only one of these dilemmas. The problems were accompanied by a drawn sketch to aid understanding. Following the original procedure, after presenting each problem, participants were asked whether the described action (e.g., pushing the man to save five people) is morally acceptable or not (*Yes/No* response). After this judgement, participants were asked to indicate on a numbered Likert-type scale ranging from 1 (*completely unacceptable*) to 9 (*completely acceptable*), the extent to which they think that the given action is morally acceptable. Next, participants were asked to type the justification of their decision in an open question format. After participants were presented with the first trolley dilemma, they were presented with a second dilemma from the same condition, without drawn sketches. For the second dilemma, we used the so-called *speedboat dilemmas*. These dilemmas have been taken from Study 1b and 2b of Greene et al., and can be found in the Supplementary Methods section 1, with the exception of the dilemmas in the *obstacle collide* and *standard footbridge* conditions, which were provided by Joshua Greene during the review of the study. The order was fixed for dilemma presentation, so that the trolley version was always presented first. Study 1 was run before Study 2, but within study, participants were randomly assigned to one of the dilemmas within that study.

Additional measures. Although the exploration of individual-level factors associated with moral thinking is not the aim of the present research, to enrich our database for future studies and secondary analyses, we expanded our survey with additional individual-level measures: 1) total yearly household income; 2) place of living (urban or rural area); 3) position on the four-dimensional Individualism-Collectivism scale[34] (16 items) for disentangling cultural differences in participants' responses[48]; 4) religion: Specific religion of the participant will be asked, plus one question to measure their level of religiosity: "On a scale from 1 to 10, how religious are you?". Furthermore, we included the Oxford Utilitarianism Scale[28] (9 items).

Following these questions, participants' level of education, age, and sex were also recorded. We also recorded participants' country of origin, and whether the participant came from an immigrant background.

Procedure

The experiment was administered by using a centralised online survey that participants could answer remotely or in the lab. We used the original instructions of Greene et al., as presented in the Supplementary Methods section 1. After responding to the dilemmas, participants were asked to answer three questions: (1) a measure of careless responding (question about the specifics of the trolley scenario); (2) whether they found the material confusing; and (3) whether they found the description of the problem realistic. After these questions, participants were directed to our series of questionnaires: the Oxford Utilitarianism Scale, followed by the Individualism-Collectivism Scale, and the measures of religion. Next, we administered the demographic questions (income, place of living, country of origin, immigrant background, level of education, age, and sex). Afterwards, we asked three further questions to measure careless responses, participants' familiarity with research questions, and finally, we asked for further comments or any experienced technical problems.

Controlling for possible confounds. To avoid second language effects on moral judgement[49], only native speakers of the language of the experiment could participate. To ensure this, we asked participants to indicate their native language(s). Bilinguals could choose their preferred language. The data of anyone with a native language different from the language of the survey were removed from data analyses.

Following Greene et al.'s procedure, data from participants who reported that they found the material confusing were excluded from the analyses. Data from participants who reported having experienced technical problems during the experiment were also excluded from all analyses. To avoid careless responses, we added three bogus items at the end of the survey. We asked participants very basic questions (e.g., "I was born on February 30th.") to which incorrect answering indicates careless responding[50]. We excluded data from participants who gave an incorrect response to any of these questions. Moreover, we introduced two additional questions (presented right after the moral dilemmas), asking participants about the specifics of the trolley and speedboat scenarios that they had been presented with, to test whether they had paid attention when reading the scenarios (referred to as attention check in the later test). Specifically, participants were asked to select the option which most accurately described the situation that they had been presented with. Each option described the nature of the physical action that was the key manipulation in the experiment. As attention to the trolley and speedboat dilemmas was measured by different questions, when analysing the responses, we excluded the data for the correspondingly failed attention check question. For example, people who gave a correct response on the trolley, but not on the speedboat attention check question, were included when analysing the trolley dilemma, and excluded when analysing the speedboat version.

As moral dilemmas are becoming more and more common in psychological research and in summaries of this research in popular media and culture and teaching, it is possible that some participants may have previous knowledge of these dilemmas, which may affect their responses. To address this potential problem, at the end of the experiment participants were asked the following question: "Before this experiment, were you familiar with moral dilemmas of this kind, in which you can save more people by causing the death of one person?" Answers were given on a rating scale from 1 (*absolutely not familiar*) to 5 (*absolutely familiar*). Familiarity with the trolley problem or such moral dilemmas (participants who responded with 4 or 5 on this scale) was used as a further exclusion criterion. Additionally, participating labs were asked to avoid recruiting philosophers or philosophy students because they are likely to have heard about trolley problems, and we wanted to minimise the number of participants to be excluded following data collection.

Notable deviations between this study and the design of Greene et al.

Besides the multinational data collection that forms the crux of our project, the first important methodological difference between this study and the original study is that the original study was conducted by paper and pencil, whereas we administered the experiment online. Of note, recent research found no evidence for a difference between the behaviour of participants who took part in the experiment online versus those who took part in the experiment in the lab. We also added one change in the introduction of the experiment (see Supplementary Methods section 1); participants were not given the opportunity to ask the researcher any questions before the experiment (as the experiment can be administered online, they did not have the opportunity to do so).

The second important change in this experiment is that participants were presented with two moral dilemmas in one condition, instead of one. These additional dilemmas will be analysed separately, as they were in the original experiment. The third difference is that for Study 2, we used different moral dilemmas than those that were used by Greene et al.; the standard switch and footbridge dilemmas were used instead of the loop weight and obstacle push dilemmas, respectively. These dilemmas are not different from the ones used by Greene et al. in their structural characteristics, only on surface characteristics. That is, in the standard switch the harm is unintended and no personal force is required, while in the standard footbridge dilemma, the harm is intended and requires personal force. By including the standard switch and standard footbridge scenarios instead of the original ones, we gain further insight into the data. Imagine for example, that the personal force effect does not replicate in one of the cultural clusters. One explanation for this is that people are simply not sensitive to the effect of personal force in that cluster. However, it might also be the case that utilitarian response rates to similar dilemmas increase over time[51]. If so, we should see that the replicated difference between the standard footbridge and switch dilemmas is shrinking or disappeared. Furthermore, by comparing the standard footbridge to the footbridge pole dilemmas, we can test the effect of physical contact, and by comparing the standard switch case to the footbridge switch case to confirm the effect of intention.

Finally, in the original experiment, Greene et al. excluded participants who did not manage to suspend disbelief. Nevertheless, as they noted, this had no effect on their results. Thus, we decided that we would not use this exclusion criterion.

Cultural classification of countries. To test the cultural universality hypothesis, a comprehensive cultural classification is needed that encompasses multiple sources of cultural variability. Hence, to assess our first hypothesis on the universality of the effect of personal force and intention on moral judgements, we used the cultural classification of Awad et al.[35]. Based on surveyed moral preferences, they identified three distinct clusters of countries: Eastern, Southern, and Western. They argued that this cluster structure is broadly consistent with the alternative, but more complex Inglehart-Welzel cultural map[34]. Therefore, we assigned the countries of our participating labs to these cultural clusters, as listed in Supplementary Analysis Section 1, Table S1.

Language adaptation. The participating labs translated the survey items into the language of the participant pool, following the translation process of the PSA (https://psysciacc.org/translation-process/) detailed below.

1. Translation: Original document is translated from source to target language by A translators resulting in document Version A
2. Back-translation: Version A is translated back from target to source language by B Translators independently resulting in Version B
3. Discussion: Version A and B are discussed among translators and the language coordinator, discrepancies in Version A and B are detected and solutions are discussed. Version C is created.
4. External Readings: Version C is tested on two non-academics fluent in the target language. Members of the fluent group are asked how they perceive and understand the translation. Possible misunderstandings are noted and again discussed as in Step 3.
5. Cultural Adjustments: Data collection labs read materials and identify any needed adjustments for their local participant sample. Adjustments are discussed with the Language Coordinator, who makes any necessary changes, resulting in the final version for each site.

Planned analyses

Preregistered analysis

Confirmatory Replication Analyses

As explained in the introduction, we focused our analyses on the question of universality of Greene et al.'s two most important claims. We conducted independent analyses in each cultural cluster and reported them separately. We preregistered the following hypotheses:

Hypothesis 1a: There is an effect of personal force on moral judgement in the Western cluster (replication of the original effect).

Hypothesis 1b: If the effect of personal force is culturally universal, there is an effect of personal force on the moral acceptability ratings (Greene et al., Study 1) in the Southern and Eastern cultural clusters as well.

Hypothesis 2a: There is an interaction between personal force and intention (Greene et al., Study 2) in the Western cluster (replication of original effects). More specifically, the intention factor is larger when personal force is present compared to when personal force is absent.

Hypothesis 2b: If this effect is culturally universal, there is an effect in the Southern and Eastern cultural clusters as well.

Unlike in the original study, we employed Bayesian analyses to gain information from our data concerning the strength of evidence for the null and alternative hypotheses. The Bayes factor indicates the relative evidence provided by the data comparing two hypotheses[52]. Regarding the threshold of strong Bayesian evidence, we followed the recommendations of [53] and set the decision threshold of BF_{10} to > 10 for H_1 and < *1/10* for H_0. We used informed priors for the alternative model: a one-tailed Cauchy distribution with a mode of zero and a scale $r = 0.26$ (Hypothesis 1a and 1b) and $r = 0.19$ (Hypothesis 2a and 2b) on the standardized effect size using the BayesFactor package[54] in R for the analysis. These priors are based on the effect sizes that we expect to find as explained below in the sample size estimation section. We will implement all of our analyses with the R statistical software[55].

To test Hypothesis 1a and 1b, we compared the moral acceptability ratings given on the footbridge switch problem and footbridge pole dilemma, with the moral acceptability rating of the footbridge switch dilemma expected to be higher. More concretely, we performed three one-sided Bayesian *t*-tests with the same comparison in each cultural group. For each cultural cluster, we would conclude that we replicated the original effect if Bayes factor (BF_{10}) > 10, we would conclude that we found a null effect if $BF_{10} < 1/10$, and we would conclude that the results are inconclusive if we find a BF_{10} in between these numbers (see below for justification of these thresholds).

To test Hypothesis 2a and 2b, we tested the interaction of personal force and intention in each cultural cluster, separately. We conducted Bayesian linear regression analysis in each cultural cluster. The Bayes factor of interest is defined as the quotient of the model including the interaction and two main effects (numerator) and the model including only the two main effects (denominator). For each cultural group, we would conclude that we replicated the original effect if the Bayes factor of the interaction (BF_{10}) > 10, we would conclude that we found a null effect if $BF_{10} < 1/10$, and we would conclude that the results are inconclusive if we find a BF_{10} in between these values (see below for justification of these thresholds). To

further understand the direction of the interaction, we will plot out the results in each cultural cluster. To conclude the replication of the original effect, we should find that the intention effect is higher in the personal force condition than in the no personal force condition.

Note that we conducted and reported the frequentist version of the proposed analysis (e.g., *t* tests for each hypothesis, for each cultural class) for the sake of comparability of the original and our results. Nevertheless, we regarded the results of our Bayesian analyses the basis of our statistical inference. Although we registered that the frequentist statistics would only be added as the supplementary material, we added it to the main text for easier comparability. No inference was drawn from the frequentist statistics.

Test assumptions for the statistical tests (t-tests and linear regressions) were assumed to hold true, but they were not formally tested.

Robustness analyses

To probe the robustness of our conclusions to the scaling factor of the Cauchy distribution used as the prior of H1, we reported Robustness Regions for each Bayes factor. Robustness Regions were notated as RR[min, max], where min indicates the smallest and max indicates the largest scaling factor that would lead us to the same conclusion as the originally chosen scaling factor[56].

Sampling plan and stopping rule

As the data were planned to be collected globally, our knowledge was insufficient concerning the noise of the measurement and the rate of exclusion in the various samples, which were needed for an accurate sample size estimation. For this reason, we proposed a sequential data acquisition. That is, first, we launched Study 1 (Hypotheses 1a and 1b), and collected data in sequences from 500 participants per cluster per condition; from 3,000 participants altogether (after all exclusions). We stop data collection after each sequence. At these stops, we conducted our planned Bayesian analyses. Should the BF reach the preset thresholds in a given cluster, we will stop data collection for that cluster. If, in a cluster, the BF thresholds were not reached, we would continue data collection with 200 additional participants per cluster per condition, and then re-analyse the data, repeating this procedure until one of the BF thresholds is reached, or the participant pool is exhausted. Note, however, that we deviated from this sampling plan. See “Deviations from registration” for details.
Should we not have reached this limit with our planned capacity of ~19,000 participants, we would have extended the data collection to a new semester. In the case that we would have not reached our evidence threshold within 12 months, we would have reported our final results, acknowledging the limited strength of the findings.
We launched Study 2 data collection in a given cluster only when the analysis of Study 1 was conclusive. In Study 2, we conducted the analysis only when we had exhausted our resources.

Sample Size estimation

To calculate our needs for data collection, we conducted a rough sample size estimation. Assuming that the original effect size is found in Study 1 ($d = 0.4$), our sample size estimation indicated that we would require 500 participants per condition per cluster (3,000 altogether), while if the original effect size is to be found in Study 2 ($d = 0.28$), our estimation indicated that we would need 1,800 participants per condition per cluster (21,600 altogether for Study 2) to obtain 95% of power in detecting the effect. A detailed description of the Sample Size estimation can be found in Supplementary Methods section 4.

Testing the association between country-level collectivism and the effects of personal force and intention

Our third hypothesis proposed that collectivism increases the effects of personal force and intention. As a measure of country-level individualism and collectivism, we added the Collectivism measure from the Cultural Distance WEIRD scale (countries' differences in terms of individualism from the United States)[57] as a continuous variable in our model. We tested whether collectivism interacted with personal force and intention (Hypothesis 3), as explained in the introduction. Hypothesis 3 expected to find a three-way interaction between collectivism, intention, and personal force, for which we used the dilemmas we used to test Hypotheses 2a and 2b. In this analysis, we used a Cauchy distribution with a scale of r = 0.37 (same we used to test Hypothesis 2a and 2b, i.e., the test of the interaction) as prior. Should we find evidence for null effect (BF < 1/10) of the interaction of individualism/collectivism, personal force, and intention, we would conclude that individualism/collectivism does not moderate the effect of personal force and intention.

Analysis of the additional moral dilemmas

Study 1.

As we explained above, each participant had to give a response on two moral dilemmas. For Study 1 (effect of personal force), we conducted the same analysis on the rest of the moral dilemmas, without the trolley versions, as in the original study (Study 1b; Greene et al.).

Study 2.

We conducted the same analysis (interaction of personal force and intention) on the rest of Speedboat dilemmas, without the trolley versions.

Further tests

Effect of physical contact and intention. With this set of items, we were able to assess the effect of physical contact, by comparing the standard footbridge and footbridge pole dilemmas. We also assessed the effect of intention by comparing the standard switch case with the footbridge switch case. These analyses were done in every cluster, and we used Bayesian t-tests for these comparisons. We used the same prior we use for the assessment of the effect of physical force ($r = 0.26$). This analysis was done separately on the trolley and speedboat dilemmas.

Comparing the standard switch and standard footbridge dilemmas. For the reasons explained earlier, we compared the standard footbridge and standard switch dilemmas, in each cultural cluster. For this, we conducted a Bayesian t-test, with the same prior previously used for the assessment of the effect of physical force ($d = 0.26$). This analysis was done separately for the trolley and speedboat dilemmas.

Oxford Utilitarianism Scale. We computed a figure showing the response distribution of each subscales of the Oxford Utilitarianism Scale[39] for each cultural cluster to explore potential cultural differences (along with means and 95% CI). The results of this can be found in the Supplementary Analysis section 2.4.

Individual-level horizontal and vertical individualism-collectivism. Triandis and Gelfand[45] defined individualistic and collectivistic cultural tendencies among 4 dimensions: *vertical individualism, vertical collectivism, horizontal individualism,* and *horizontal collectivism*. We added these continuous measures to our Bayesian linear regression analysis. The predictive power of all four measures were assessed separately.

Including familiar participants. A potentially large number of participants were excluded due to familiarity with the trolley dilemma, and there was a possibility that this exclusion criterion will affect the data from some countries or cultural clusters more than others. To avoid this potential sampling bias, we computed all above-listed analyses on moral dilemmas (confirmatory and exploratory) on the full sample in which we did not exclude the participants who were familiar with moral dilemmas. Second, we computed all analyses specifically on data coming from people who were familiar with moral dilemmas in order to compare the results of "familiar" and "unfamiliar" participants. This latter analysis can be found in the Supplementary Analysis section 2.3 and was limited to the confirmatory hypothesis tests.

Pilot testing. To ascertain that the survey software operates without any technical problems, we planned to conduct a pilot test in which each participating lab would have been expected to collect data from 10 participants. We would have only assessed the expected functioning of the survey software without analysing the collected data.

Timeline. We planned to finish data collection within six months from Stage 1 in principle acceptance and we planned to submit our report within one month from then.

Deviations from registration

We preregistered that we would collect data from 3,000 participants for Study 1 (test of personal force; H1a, H1b), after exclusions. Unexpectedly, the exclusion criteria led to 80.6% exclusion of our collected data. At the point when this was realized, it seemed likely that Study 1 would exhaust the available sample pool, not leaving capacity for Study 2. Therefore, with the agreement of the journal editor, we decided to collect participants for Study 1 only

until our Bayes Factor evidence thresholds were reached after all exclusion criteria were applied. This modification allowed us to collect data for Study 2 as well.

At the time of this decision, the distribution of responses has been taken into account: we had collected data from 3,473 participants: 1319 from the "Western cluster", 1762 from the "Southern" cluster, and 392 from the "Eastern" cluster. After exclusions, 789 participants remained (78% excluded): 296 from the "Western" cluster (78% excluded), 429 from the "Southern" cluster (76% excluded), and 64 from the "Eastern" cluster (84% excluded).

Instead of conducting a pilot study as preregistered, in order to avoid wasting any (much needed) participants, participating researchers from all labs tested the experiment before it was sent out to assure that there are no grammatical mistakes or functionality problems.

Due to COVID-19 crisis, data collection took 6 months longer than expected (with the agreement of the editor).

Exploratory analysis

During the data pre-processing, we excluded 229 participants from three US-based labs as they received a wrong survey link. Furthermore, 13,359 participants started, but did not finish the experiment, therefore their data were also dropped from further analyses. These participants did not count towards our final sample and are not part of the data in any way. The final sample used for data analyses consisted of 27,502 participants. Further information on the demographics of our participants can be found in the Supplementary Analysis section 1.

Note that we limited the use of Robustness Regions for the confirmatory hypothesis tests.

Data availability statement

Collected anonymised raw and processed data are publicly shared on the Github page of the project: https://github.com/marton-balazs-kovacs/trolleyMultilabReplication/tree/master/data.

Code availability statement

Code for data management and statistical analyses have been written in R and are available at: https://github.com/marton-balazs-kovacs/trolleyMultilabReplication.

Protocol Registration Information

The Stage 1 protocol for this Registered Report was accepted in principle on 30th January 2020. The protocol, as accepted by the journal, can be found at https://doi.org/10.6084/m9.figshare.11871324.v1

Acknowledgements

M.A. Vadillo was supported by 2016-T1/SOC-1395 and 2020-5A/SOC-19723 from Comunidad de Madrid; PSI2017-85159-P from AEI and UE/FEDER; M.P.-C. was supported by 2017/01/X/HS6/01332 from the National Science Centre, Poland; P.M. was supported by Aarhus University Research Foundation (AUFF), Starting Grant: AUFF-E-2019-9-4; B. Bago was supported by ANR grant ANR-17-EURE-0010 (Investissementts d'Avenir program), and ANR Labex IAST; R.M.R. was supported by Australian Research Council (DP180102384); N.B.D. was supported by CAPES grant number 88887.364180/2019-00; C.S., K.A.Ś., and I. Zettler were supported by Carlsberg Foundation (CF16-0444), Independent Research Fund Denmark (7024-00057B); J.L. was supported by EXC 2126/1– 390838866 under Germany´s Excellence Strategy; K.B. was supported by following grants from the National Science Centre, Poland: (1) while working on the data collection: No.: 2015/19/D/HS6/00641, (2) while working on the final version of the paper: No.: 2019/35/B/HS6/00528; A.W. was supported by FONDECYT 11190980, CONICYT; A. Fleischmann was supported by German Research Foundation (research unit grant FOR-2150, LA 3566/1-2); H.Y. was supported by JSPS Grant (18K03010); Y.Y. was supported by JSPS KAKENHI (16H03079, 17H00875, 18K12015, and 20H04581); K.Q. was supported by JSPS KAKENHI (17H06342, 20K03479, and 20KK0054); A. Ikeda was supported by JSPS KAKENHI (20J21976); K.M.K. was supported by National Natural Science Foundation of China (NSFC) (grant number 31530032) and Key Technological Projects of Guangdong Province (grant number 2018B030335001); J.B.C. was supported by National Science Foundation Graduate Research Fellowship (Grant No. DGE-1839285); M. Parzuchowski, K. Rybus, and N.M.S. were supported by Polish National Science Center & DFG Beethoven grant (2016/23/G/HS6/01775); A.C.S. was supported by Portuguese National Foundation for Science and Technology [Grant number SFRH/BD/126304/2016]; L. Boncinelli was supported by PRIN 2017, grant n. 20178293XT, Italian Ministry of Education and Research; M.F.F.R. was supported by PSA 006 BRA 008 Data Collection in Support of PSADM 001 Measurement Invariance Project; M. Misiak was supported by scholarship of the Foundation for Polish Science (START) and by the scholarship of the National Science Centre (2020/36/T/HS6/00256); P.B. was supported by Slovak Research and Development Agenc [project no. APVV-18-0140]; M.A. was supported by Slovak Research and Development Agency [project no. APVV-17-0418] & project PRIMUS/20/HUM/009; A. Findor and M.H. were supported by Slovak Research and Development Agency under contract no. APVV-17-0596; T.G. was supported by Social Sciences and Humanities Research Council of Canada (# 950-224884); P.P. was supported by Swedish Research Council (2016-06793); Y.L. was supported by The Project of Philosophy and Social Science Research in Colleges and Universities in Jiangsu Province (grant number: 2020SJA0017); M. Kowal was supported by the scholarship of the National Science Centre (2019/33/N/HS6/00054); P.A. was supported by UID/PSI/03125/2019 from the Portuguese National Foundation for Science and Technology.

The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author Contribution

Conceptualization: B. Bago and B.A.

Data curation: B. Bago, M. Kovacs, and T.N.

Formal analysis: B. Bago, M. Kovacs, T.N., Z.K., and B.P.

Funding acquisition: P.A., P.M., K.Q., I. Zettler, and R. Hoekstra

Investigation: B. Bago, M. Kovacs, M.A., S. Adamus, S. Albalooshi, N.A.-A., S. Alper, S.A.-S., S.G.A., S. Amaya, P.K.A., G.A., D.A., P.A., J.J.B.R.A., A.A., P.B., K.B., B. Bashour, E. Baskin, L. Batalha, C.B., J. Bavolar, F.B., M. Becker, B. Becker, A.B., M. Białek, E. Bilancini, D.B., L. Boncinelli, J. Boudesseul, B.T.B., E.M.B., M.M.B., D.P.C., N.C.C., J.B.C., C.R.C., W.J.C., P.C., H.C.-P., R.F.C., O.Ç., R.C.C., V.C.A., C.P.C., S.C., Y.D., J.A.M.d.G., W.C.d.V., E.G.D.B., C.D., B.J.W.D., X.D., F.D., A.D., N.B.D., J.E., C.E.-S., L.E., T.R.E., G.F., F.M.F., S.F., A. Findor, A. Fleischmann, F.F., R.F., D.-A.F., C.H.Y.F., S.G., O.G., A.-R.G.-N., M.E.G., I.G., T.G., B.G., M. Gollwitzer, A.G., M. Grinberg, A.G.-B., E.A.H., A.H., W.A.N.M.H., J.H., K.R.H., J.J.J.H., E.H., M.H., C.A.H., R. Huskey, A. Ikeda, Y.I., G.P.D.I., O.I., C.I., A. Iyer, B. Jaeger, S.M.J.J., W.J.-L., B. Jokić, P.K., V.K., G.K., F.K.-M., A.T.A.K., K.M.K., B.J.K., H.E.K., R.I.K., M. Kowal, E.K., L.K., A.K., A.O.K., F.L., C.L., J.L., E.B.L., A.L., I.Y.-M.L., L.B.L., M.C.L., J.N.L., C.A.L., S.C.L., M.L., Y.L., H.L., T.J.S.L., S.L., M.T.L., P.L., J.G.L., T.L., M. Máčel, S.P.M., M. Maganti, Z.M.-M., L.F.M., H.M., G.M.M., D.M.S., C.-J.M., A.D.A.M., M. Mazidi, J.P.M., N.M., M.C.M., L.M., T.L.M., A. Mirisola, M. Misiak, P.M., M.M.-J., A. Monajem, D.M., E.D.M., E.N., I.N., D.P.O., J.O., N.C.O., A.A.Ö., M. Panning, M.P.-P., N.P., P.P., M.P.-C., M. Parzuchowski, J.V.P., J.M.P., M. Peker, K.P., L.P., I.P., M.R.P., N.P.-J., A.J.P., M.A.P., E. Pronizius, D.P., E. Puvia, V.Q., K.Q., A.Q., B.R., D.A.R., U.-D.R., C.R., K. Reynolds, M.F.F.R., J.P.R., R.M.R., P.R., F.R.-D., S.R.-F., B.T.R., K. Rybus, A. Samekin, A.C.S., N.S., C.S., K.S., K.A.Ś., M. Sharifian, J. Shi, Y.S., E.S., M. Sirota, M. Slipenkyj, Ç.S., A. Sorokowska, P.S., S. Söylemez, N.K.S., I.D.S., A. Sternisko, L.S.-W., S.L.K.S., S. Stieger, D.S., J. Strube, K.J.S., R.D.S.-C., N.M.S., B. Takwin, S.T., A.G.T., K.E.T., L.E.T., M. Tonković, B. Trémolière, L.V.T., B.N.T., M. Twardawski, M.A. Vadillo, Z.V., L.A.V., B.V., D.V., M.V., M.A. Vranka, S. Wang, S.-L.W., S. Whyte, L.S.W., A.W., X.W., F.X., S. Yadanar, H.Y., Y.Y., O.Y., S. Yoon, D.M.Y., I. Zakharov, R.A.Z., I. Zettler, I.L.Ž., D.C.Z., J.Z., X.Z., and B.A.

Methodology: B. Bago, Z.K., B.P., R. Hoekstra, and B.A.

Project administration: B. Bago, M. Kovacs, J.P., and B.A.

Resources: B. Bago, M. Kovacs, S.G.A., G.A., P.A., H.C.-P., R.C.C., Y.D., X.D., W.J.-L., F.K.-M., C.L., H.M., A.A.Ö., V.Q., A.C.S., Y.S., J. Strube, N.M.S., M.V., I. Zakharov, and B.A.

Supervision: B. Bago, J.P., P.A., A.A.Ö., P.S., M.V., and B.A.

Validation: B. Bago, M. Kovacs, T.N., Z.K., B.P., and B.A.

Visualization: B. Bago, M. Kovacs, T.N., and B.P.

Writing - original draft: B. Bago and B.A.

Writing - review & editing: B. Bago, M. Kovacs, J.P., B.P., M.A., S. Adamus, S. Albalooshi, N.A.-A., S. Alper, S.A.-S., S.G.A., S. Amaya, P.K.A., G.A., D.A., P.A., J.J.B.R.A., A.A., P.B., K.B., B. Bashour, E. Baskin, L. Batalha, C.B., J. Bavolar, F.B., M. Becker, B. Becker, A.B., M. Białek, E. Bilancini, D.B., L. Boncinelli, J. Boudesseul, B.T.B.,

E.M.B., M.M.B., D.P.C., N.C.C., J.B.C., C.R.C., W.J.C., P.C., H.C.-P., R.F.C., O.Ç., R.C.C., V.C.A., C.P.C., S.C., Y.D., J.A.M.d.G., W.C.d.V., E.G.D.B., C.D., B.J.W.D., X.D., F.D., A.D., N.B.D., J.E., C.E.-S., L.E., T.R.E., G.F., F.M.F., S.F., A. Findor, A. Fleischmann, F.F., R.F., D.-A.F., C.H.Y.F., S.G., O.G., A.-R.G.-N., M.E.G., I.G., T.G., B.G., M. Gollwitzer, A.G., M. Grinberg, A.G.-B., E.A.H., A.H., W.A.N.M.H., J.H., K.R.H., J.J.J.H., E.H., M.H., C.A.H., R. Huskey, A. Ikeda, Y.I., G.P.D.I., O.I., C.I., A. Iyer, B. Jaeger, S.M.J.J., W.J.-L., B. Jokić, P.K., V.K., G.K., F.K.-M., A.T.A.K., K.M.K., B.J.K., H.E.K., R.I.K., M. Kowal, E.K., L.K., A.K., A.O.K., F.L., C.L., J.L., E.B.L., A.L., I.Y.-M.L., L.B.L., M.C.L., J.N.L., C.A.L., S.C.L., M.L., Y.L., H.L., T.J.S.L., S.L., M.T.L., P.L., J.G.L., T.L., M. Máčel, S.P.M., M. Maganti, Z.M.-M., L.F.M., H.M., G.M.M., D.M.S., C.-J.M., A.D.A.M., M. Mazidi, J.P.M., N.M., M.C.M., L.M., T.L.M., A. Mirisola, M. Misiak, P.M., M.M.-J., A. Monajem, D.M., E.D.M., E.N., I.N., D.P.O., J.O., N.C.O., A.A.Ö., M. Panning, M.P.-P., N.P., P.P., M.P.-C., M. Parzuchowski, J.V.P., J.M.P., M. Peker, K.P., L.P., I.P., M.R.P., N.P.-J., A.J.P., M.A.P., E. Pronizius, D.P., E. Puvia, V.Q., K.Q., A.Q., B.R., D.A.R., U.-D.R., C.R., K. Reynolds, M.F.F.R., J.P.R., R.M.R., P.R., F.R.-D., S.R.-F., B.T.R., K. Rybus, A. Samekin, A.C.S., N.S., C.S., K.S., K.A.Ś., M. Sharifian, J. Shi, Y.S., E.S., M. Sirota, M. Slipenkyj, Ç.S., A. Sorokowska, P.S., S. Söylemez, N.K.S., I.D.S., A. Sternisko, L.S.-W., S.L.K.S., S. Stieger, D.S., J. Strube, K.J.S., R.D.S.-C., N.M.S., B. Takwin, S.T., A.G.T., K.E.T., L.E.T., M. Tonković, B. Trémolière, L.V.T., B.N.T., M. Twardawski, M.A. Vadillo, Z.V., L.A.V., B.V., D.V., M.V., M.A. Vranka, S. Wang, S.-L.W., S. Whyte, L.S.W., A.W., X.W., F.X., S. Yadanar, H.Y., Y.Y., O.Y., S. Yoon, D.M.Y., I. Zakharov, R.A.Z., I. Zettler, I.L.Ž., D.C.Z., J.Z., X.Z., and B.A.

Competing interests statement

The authors declare no competing interests.

Table 1

Summary of sample sizes and exclusions in all cultural clusters

	Eastern	Southern	Western	All
Reason to exclude				
N without exclusion	3,877	5,333	18,292	27,502
Careless responding	156 (4.0%)	82 (1.5%)	256 (1.4%)	494 (1.8%)
Confusion	752 (19.4%)	658 (12.3%)	1,718 (9.4%)	3,128 (11.4%)
Familiarity with moral dilemmas	1,669 (43.0%)	2,501 (46.9%)	10,332 (56.5%)	14,502 (52.7%)
Technical problem	531 (13.7%)	413 (7.7%)	1,225 (6.7%)	2,169 (7.9%)
Non-native speaker	347 (9.0%)	177 (3.3%)	1,305 (7.1%)	1,829 (6.7%)
Failed attention check (Study1a)	720 (18.6%)	943 (17.7%)	1,311 (7.2%)	2,974 (10.8%)
Failed attention check (Study 1b)	849 (21.9%)	1,042 (19.5%)	1,336 (7.3%)	3,227 (11.7%)
Failed attention check (Study 2a)	1,102 (28.4%)	1,071 (20.1%)	4,900 (26.8%)	7,073 (25.7%)
Failed attention check (Study 2b)	1,195 (30.8%)	1,367 (25.6%)	5,528 (30.2%)	8,090 (29.4%)
Final sample				
Study1a	381	622	566	1,569
Study1b	327	553	546	1,426
Study2a	323	690	2,971	3,984
Study2b	277	576	2,660	3,513

Note. Study 1b and Study 2b refers to the Speedboat dilemmas (recall, all of our subjects answered to one trolley and a speedboat dilemmas)

Table 2

The effect of personal force on moral dilemma judgements

Dilemma	Exclusion	Cluster	BF	RR	t	df	p	Cohen's d	Raw effect	89% CI
Trolley	Exclude	Eastern	$1.9*10^{2}$	$7.00*10^{-3}$, 14.00	-3.69	366.23	<.001	0.38	0.85	[0.39, 1.12]
		Southern	$2.44*10^{7}$	$1.00*10^{-5}$, $2.80*10^{6}$	-6.32	619.93	<.001	0.51	1.10	[0.76, 1.33]
		Western	80.1	$1.20*10^{-2}$, 4.30	-3.41	553.15	0.001	0.29	0.59	[0.24, 0.79]
	Including familiar	Eastern	$9.21*10^{4}$	$<1.50*10^{-5}$, $6.50*10^{3}$	-5.19	806.76	<.001	0.36	0.79	[0.51, 1]
		Southern	$5.91*10^{12}$	$<1.00*10^{-5}$, $5.50*10^{11}$	-8.09	1345.85	<.001	0.44	0.94	[0.73, 1.1]
		Western	$4.95*10^{5}$	$<1.00*10^{-5}$, $2.90*10^{4}$	-5.51	1338.48	<.001	0.30	0.65	[0.43, 0.8]
Speedboat	Exclude	Eastern	$1.16*10^{5}$	$1.80*10^{-5}$, $1.70*10^{4}$	-5.26	283.92	<.001	0.59	1.18	[0.77, 1.47]
		Southern	$1.01*10^{3}$	$1.30*10^{-3}$, 74.00	-4.19	436.86	<.001	0.37	0.72	[0.37, 0.93]
		Western	25.2	$3.30*10^{-2}$, 1.20	-3.01	437.36	0.003	0.27	0.51	[0.18, 0.72]
	Including familiar	Eastern	$2.4*10^{4}$	$<6.00*10^{-5}$, $1.70*10^{3}$	-4.88	680.10	<.001	0.37	0.74	[0.46, 0.95]
		Southern	$7.8*10^{6}$	$<1.00*10^{-5}$, $5.50*10^{5}$	-5.94	908.97	<.001	0.36	0.69	[0.49, 0.85]
		Western	$5.53*10^{7}$	$<1.00*10^{-5}$, $4.0*10^{6}$	-6.34	1140.72	<.001	0.35	0.71	[0.51, 0.87]

Note. BF = Bayes Factor, RR = Robustness Region of the prior

Table 3

Interaction between personal force and intention on moral judgments

Dilemma	Exclusion	Cluster	BF	RR	b	89% CI	p	Partial η^2	Raw effect
Trolley	Exclusion	Eastern	0.59	$2.20*10^{-2}$, 0.64	0.027	[-0.16, 0.19]	0.84	0.000	0.11
		Southern	9.35	$2.75*10^{-2}$, 0.2	-0.250	[-0.35, -0.09]	0.002	0.014	-1.00
		Western	$1.54*10^{11}$	$5.80*10^{-5}$, $1.80*10^{3}$	-0.306	[-0.36, -0.24]	<.001	0.019	-1.23
	Include familiar	Eastern	2.85	$2.50*10^{-2}$, 1.35	-0.213	[-0.33, -0.03]	0.031	0.008	-0.85
		Southern	$3.08*10^{6}$	$2.23*10^{-3}$, 60	-0.348	[-0.43, -0.25]	<.001	0.026	-1.39
		Western	$1.28*10^{30}$	$<1.00*10^{-5}$, $3.70*10^{9}$	-0.292	[-0.33, -0.25]	<.001	0.018	-1.17
Speedboat	Exclusion	Eastern	0.43	$4.60*10^{-2}$, 0.69	-0.007	[-0.17, 0.2]	0.959	0.000	-0.03
		Southern	0.36	$5.10*10^{-2}$, 0.65	0.028	[-0.12, 0.16]	0.794	0.000	0.11
		Western	222	$3.60*10^{-2}$, 1.15	-0.160	[-0.22, -0.08]	<.001	0.005	-0.64
	Include familiar	Eastern	0.42	$4.50*10^{-2}$, 0.6	0.010	[-0.14, 0.16]	0.926	0.000	0.04
		Southern	1.13	$3.20*10^{-2}$, 0.94	-0.132	[-0.23, 0.01]	0.097	0.002	-0.53
		Western	$4.75*10^{7}$	$6*10^{-4}$, 75	-0.152	[-0.19, -0.11]	<.001	0.005	-0.61

Note. BF = Bayes Factor, RR = Robustness Region of the prior

Table 4

Individualism/collectivism associations with the interaction between personal force and intention on moral judgments (Trolley dilemmas)

		With familiarity exclusion				No familiarity exclusion			
Dilemma	**Variable**	**BF**	**b**	**89% CI**	**p**	**BF**	**b**	**89% CI**	**p**
Trolley	Country-level collectivism	1.17	-1.13	[-3.17, 1.12]	0.405	2.17	-1.27	[-2.53, -0.11]	0.096
	H. Collectivism	1.66	-0.03	[-0.06, 0.01]	0.263	2.31	-0.03	[-0.05, 0]	0.096
	H. Individualism	0.70	0.00	[-0.04, 0.04]	0.921	0.94	0.02	[-0.01, 0.04]	0.325
	V. Collectivism	0.88	0.00	[-0.03, 0.04]	0.988	0.71	-0.01	[-0.03, 0.01]	0.538
	V. Individualism	0.72	-0.02	[-0.05, 0.02]	0.451	0.45	-0.01	[-0.03, 0.01]	0.607
Speedboat	Country-level collectivism	0.91	0.66	[-1.43, 2.9]	0.631	0.66	-0.32	[-1.61, 0.83]	0.684
	H. Collectivism	3.11	-0.04	[-0.08, 0]	0.114	0.91	-0.01	[-0.04, 0.01]	0.396
	H. Individualism	1.11	-0.01	[-0.05, 0.03]	0.611	0.70	0.00	[-0.02, 0.03]	0.852
	V. Collectivism	1.53	0.02	[-0.01, 0.06]	0.311	0.96	0.01	[-0.01, 0.04]	0.357
	V. Individualism	0.70	0.00	[-0.04, 0.03]	0.952	0.54	0.01	[-0.01, 0.03]	0.590

Table 5

The effect of intention on moral dilemma judgements (Trolley dilemmas)

Exclusion	Cluster	BF	t	df	p	Cohen's d	Raw effect	89% CI
Exclusion	Eastern	35.5	-3.13	159.97	0.002	0.41	0.99	[0.34, 1.36]
	Southern	$4.29*10^{6}$	-6.00	214.10	<.001	0.64	1.47	[0.99, 1.78]
	Western	$1.95*10^{15}$	-8.90	571.04	<.001	0.70	1.46	[1.17, 1.7]
Include familiar	Eastern	$6.05*10^{2}$	-3.93	234.76	<.001	0.40	0.91	[0.49, 1.2]
	Southern	$5.29*10^{13}$	-8.63	499.67	<.001	0.61	1.34	[1.04, 1.55]
	Western	$3.3*10^{34}$	-12.84	1278.97	<.001	0.64	1.33	[1.15, 1.47]
No exclusion	Eastern	30.6	-3.07	1060.61	0.002	0.17	0.39	[0.18, 0.57]
	Southern	$1.61*10^{14}$	-8.46	1421.86	<.001	0.40	0.89	[0.7, 1.04]
	Western	$2.89*10^{26}$	-11.01	2999.62	<.001	0.34	0.72	[0.62, 0.82]

Table 6

The effect of intention on moral dilemma judgements (Speedboat dilemmas)

Exclusion	Cluster	BF	t	df	p	Cohen's d	Raw effect	89% CI
Exclusion	Eastern	10.6	-2.67	192.91	0.008	0.35	0.78	[0.2, 1.12]
	Southern	$2.81*10^{5}$	-5.51	407.77	<.001	0.54	1.06	[0.68, 1.3]
	Western	$3.15*10^{9}$	-7.23	327.02	<.001	0.54	1.09	[0.81, 1.31]
Include familiar	Eastern	$3.83*10^{4}$	-4.99	319.39	<.001	0.48	1.03	[0.64, 1.3]
	Southern	$9.55*10^{6}$	-6.10	872.90	<.001	0.41	0.81	[0.57, 0.99]
	Western	$2.51*10^{16}$	-8.77	769.66	<.001	0.43	0.84	[0.68, 0.98]
No exclusion	Eastern	29.6	-3.06	1062.72	0.002	0.17	0.38	[0.18, 0.56]
	Southern	$1.83*10^{7}$	-6.12	1400.39	<.001	0.29	0.60	[0.43, 0.74]
	Western	$2.42*10^{12}$	-7.65	3006.15	<.001	0.23	0.47	[0.37, 0.56]

Figure 1

Results on Trolley and Speedboat dilemmas in Study 1 (effect of personal force) with all exclusion criteria applied (A,C) and including familiar participants (B,D). Error bars are 95% Confidence Intervals around the mean. Scale ranged from 1 (completely unacceptable) to 9 (completely acceptable). In the trolley problem n = 1,569 when all exclusion criteria is applied, and n = 3,524 when the familiarity exclusion not applied. In the speedboat dilemma, n = 1,426 when all exclusion criteria is applied, and n = 3,295 when the familiarity exclusion not applied.

Figure 2

Results on Trolley and Speedboat dilemmas in Study 2 (personal force and intention interaction) with all exclusion criteria applied (A,C) and including familiar participants (B,D). Error bars represent 95% Confidence Intervals. Scale ranged from 1 (completely unacceptable) to 9 (completely acceptable). In the trolley problem n = 3,984 when all exclusion criteria is applied, and n = 9,844 when the familiarity exclusion not applied. In the speedboat dilemma, n = 3,513 when all exclusion criteria is applied, and n = 9,006 when the familiarity exclusion not applied.

Figure 3

Correlation between country-level collectivism and the Eta squared effect size of the interaction between personal force and intention with all exclusion criteria applied (left panel) and including participants familiar with the trolley problem (right panel) . The size of the circles indicate the size of the sample in a given country. Blue line is the weighted regression line.

Figure 4

Correlation between country-level collectivism and the Eta squared effect size of the interaction between personal force and intention with all exclusion criteria applied (left) and

including familiar participants (right) on the speedboat problem. The size of the circles indicate the size of the sample in a given country. Blue line is the weighted regression line.

References

1. Mill, J. S. & Bentham, J. *Utilitarianism and other essays*. (Penguin, 1987).
2. Kant, I. *Groundwork for the metaphysics of morals.* (Yale University Press, 1785).
3. Greene, J. D. *Moral tribes: emotion, reason and the gap between us and them*. (Penguin Press, 2013).
4. London, J. A. How should we model rare disease allocation decisions? *Hastings Cent. Rep.* **42**, 3 (2014).
5. Bonnefon, J.-F., Shariff, A. & Rahwan, I. The social dilemma of autonomous vehicles. *Science* **352**, 1573–1576 (2016).
6. Foot, P. The problem of abortion and the doctrine of double effect. *Oxf. Rev.* **5**, 5–15 (1967).
7. Baron, J. Utilitarian vs. deontological reasoning: method, results, and theory. in *Moral inferences* (eds. Bonnefon, J.-F. & Trémolière, B.) 137–151 (Psychology Press, 2017).
8. Baron, J. & Gürçay, B. A meta-analysis of response-time tests of the sequential two-systems model of moral judgment. *Mem. Cognit.* **45**, 566–575 (2017).
9. Cushman, F., Young, L. & Hauser, M. The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. *Psychol. Sci.* **17**, 1082–1089 (2006).
10. Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M. & Cohen, J. D. The neural bases of cognitive conflict and control in moral judgment. *Neuron* **44**, 389–400 (2004).
11. Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M. & Cohen, J. D. An fMRI investigation of emotional engagement in moral judgment. *Science* **293**, 2105–2108 (2001).
12. Gürçay, B. & Baron, J. Challenges for the sequential two-system model of moral judgement. *Think. Reason.* **23**, 49–80 (2017).
13. Mikhail, J. Universal moral grammar: Theory, evidence and the future. *Trends Cogn. Sci.* **11**, 143–152 (2007).
14. Boyle, J. Medical ethics and double effect: the case of terminal sedation. *Theor. Med.*

Bioeth. **25**, 51–60 (2004).

15. Gross, M. L. Bioethics and armed conflict: Mapping the moral dimensions of medicine and war. *Hastings Cent. Rep.* **34**, 22–30 (2004).
16. Gross, M. L. Killing civilians intentionally: double effect, reprisal, and necessity in the Middle East. *Polit. Sci. Q.* **120**, 555–579 (2005).
17. Tully, P. A. The doctrine of double effect and the question of constraints on business decisions. *J. Bus. Ethics* **58**, 51–63 (2005).
18. Greene, J. D. *et al.* Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition* **111**, 364–371 (2009).
19. Cushman, F. Action, outcome, and value: A dual-system framework for morality. *Personal. Soc. Psychol. Rev.* **17**, 273–292 (2013).
20. Cushman, F., Gray, K., Gaffey, A. & Mendes, W. B. Simulating murder: The aversion to harmful action. *Emotion* **12**, 2 (2012).
21. Ellsworth, R. M. & Walker, R. S. Sociobiology of lethal violence in small-scale societies. in *The Routledge International Handbook of Biosocial Criminology* 85–102 (Routledge, 2014).
22. Abarbanell, L. & Hauser, M. D. Mayan morality: An exploration of permissible harms. *Cognition* **115**, 207–224 (2010).
23. Barrett, H. C. *et al.* Small-scale societies exhibit fundamental variation in the role of intentions in moral judgment. *Proc. Natl. Acad. Sci.* **113**, 4688–4693 (2016).
24. Cushman, F., Young, L. & Greene, J. D. Our multi-system moral psychology: Towards a consensus view. in *The Oxford handbook of moral psychology* 47–71 (2010).
25. Koenigs, M. *et al.* Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature* **446**, 908–911 (2007).
26. Perkins, A. M. *et al.* A dose of ruthlessness: Interpersonal moral judgment is hardened by the anti-anxiety drug lorazepam. *J. Exp. Psychol. Gen.* **142**, 612–620 (2013).
27. Szekely, R. D. & Miu, A. C. Incidental emotions in moral dilemmas: The influence of emotion regulation. *Cogn. Emot.* **29**, 64–75 (2015).

1315 28. Johnson, R. C. *et al.* Guilt, shame, and adjustment in three cultures. *Personal. Individ.*
1316 *Differ.* **8**, 357–364 (1987).
1317 29. Tracy, J. L. & Matsumoto, D. The spontaneous expression of pride and shame:
1318 Evidence for biologically innate nonverbal displays. *Proc. Natl. Acad. Sci.* **105**, 11655–
1319 11660 (2008).
1320 30. Scollon, C. N., Diener, E., Oishi, S. & Biswas-Diener, R. Emotions across cultures and
1321 methods. *J. Cross-Cult. Psychol.* **35**, 304–326 (2004).
1322 31. Heinrichs, N. *et al.* Cultural differences in perceived social norms and social anxiety.
1323 *Behav. Res. Ther.* **44**, 1187–1197 (2006).
1324 32. Gleichgerrcht, E. & Young, L. Low levels of empathic concern predict utilitarian moral
1325 judgment. *PloS One* **8**, e60418 (2013).
1326 33. Luo, S. *et al.* Interaction between oxytocin receptor polymorphism and interdependent
1327 culture values on human empathy. *Soc. Cogn. Affect. Neurosci.* **10**, 1273–1281 (2015).
1328 34. Cheon, B. K. *et al.* Cultural influences on neural basis of intergroup empathy.
1329 *NeuroImage* **57**, 642–650 (2011).
1330 35. Awad, E. *et al.* The Moral Machine experiment. *Nature* **563**, 59–64 (2018).
1331 36. Koenig, L. B., McGue, M., Krueger, R. F. & Bouchard Jr, T. J. Genetic and
1332 environmental influences on religiousness: Findings for retrospective and current
1333 religiousness ratings. *J. Pers.* **73**, 471–488 (2005).
1334 37. Kahane, G. On the wrong track: Process and content in moral psychology. *Mind Lang.*
1335 **27**, 519–545 (2012).
1336 38. Kahane, G., Everett, J. A., Earp, B. D., Farias, M. & Savulescu, J. 'Utilitarian'judgments
1337 in sacrificial moral dilemmas do not reflect impartial concern for the greater good.
1338 *Cognition* **134**, 193–209 (2015).
1339 39. Kahane, G. *et al.* Beyond sacrificial harm: A two-dimensional model of utilitarian
1340 psychology. *Psychol. Rev.* **125**, 131–164 (2017).
1341 40. Conway, P., Goldstein-Greenwood, J., Polacek, D. & Greene, J. D. Sacrificial utilitarian
1342 judgments do reflect concern for the greater good: Clarification via process dissociation

and the judgments of philosophers. *Cognition* **179**, 241–265 (2018).

41. Hauser, M. *Moral minds: How nature designed our universal sense of right and wrong.* (Ecco/HarperCollins Publishers, 2006).
42. Hauser, M. D., Young, L. & Cushman, F. Reviving Rawls' linguistic analogy. *Moral Psychol.* **2**, 107–143 (2008).
43. Gelfand, M. J., Nishii, L. H. & Raver, J. L. On the nature and importance of cultural tightness-looseness. *J. Appl. Psychol.* **91**, 1225 (2006).
44. Awad, E., Dsouza, S., Shariff, A., Rahwan, I. & Bonnefon, J.-F. Universals and variations in moral decisions made in 42 countries by 70,000 participants. *Proc. Natl. Acad. Sci.* **117**, 2332–2337 (2020).
45. Gawronski, B., Armstrong, J., Conway, P., Friesdorf, R. & Hütter, M. Consequences, norms, and generalized inaction in moral dilemmas: The CNI model of moral decision-making. *J. Pers. Soc. Psychol.* **113**, 343 (2017).
46. Conway, P. & Gawronski, B. Deontological and utilitarian inclinations in moral decision making: a process dissociation approach. *J. Pers. Soc. Psychol.* **104**, 216 (2013).
47. Moshontz, H. *et al.* The Psychological Science Accelerator: Advancing psychology through a distributed collaborative network. *Adv. Methods Pract. Psychol. Sci.* **1**, 501–515 (2018).
48. Bond, M. H. & van de Vijver, F. J. R. Making scientific sense of cultural differences in psychological outcomes: Unpackaging the Magnum Mysterium. in *Culture and psychology. Cross-cultural research methods in psychology* (eds. Matsumoto, D. & van de Vijver, F. J. R.) 75–100 (Cambridge University Press, 2011).
49. Costa, A. *et al.* Your morals depend on language. *PloS One* **9**, e94842 (2014).
50. Meade, A. W. & Craig, S. B. Identifying careless responses in survey data. *Psychol. Methods* **17**, 437–455 (2012).
51. Hannikainen, I. R., Machery, E. & Cushman, F. A. Is utilitarian sacrifice becoming more morally permissible? *Cognition* **170**, 95–101 (2018).
52. Dienes, Z. Using Bayes to get the most out of non-significant results. *Front. Psychol.* **5**,

781 (2014).

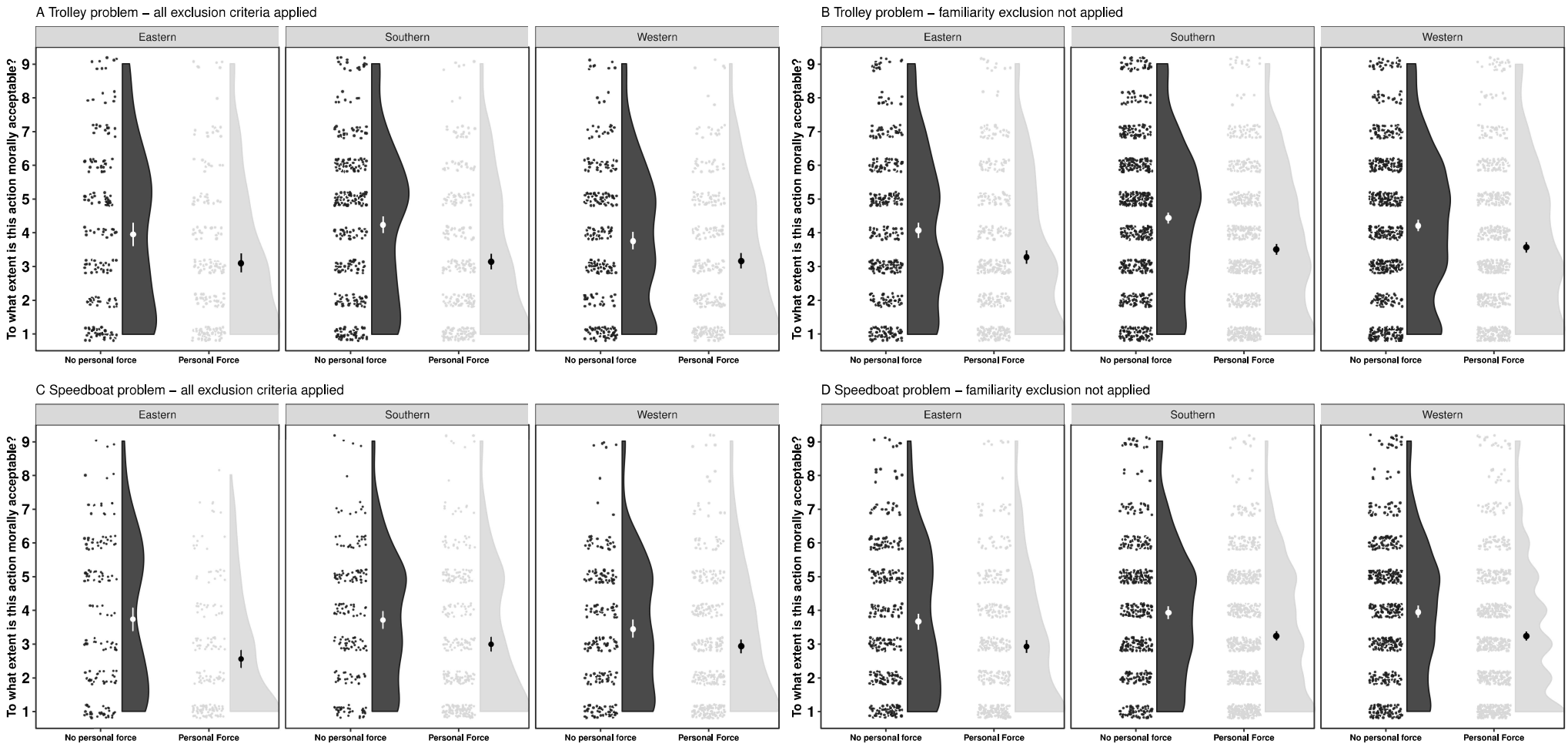
53. Schönbrodt, F. D. & Wagenmakers, E.-J. Bayes factor design analysis: Planning for compelling evidence. *Psychon. Bull. Rev.* **25**, 128–142 (2018).

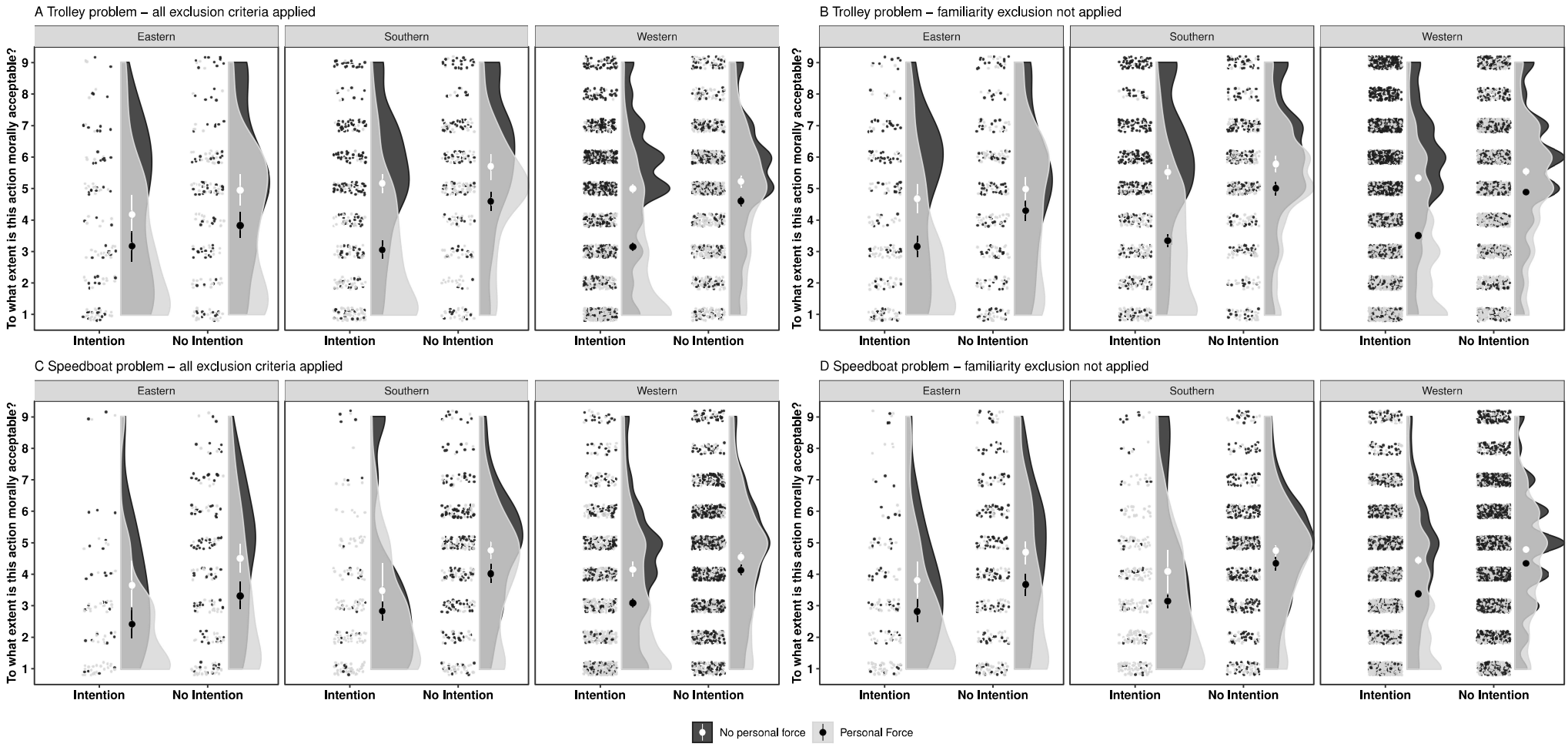
54. Morey, R. D., Rouder, J. N. & Jamil, T. *BayesFactor: Computation of Bayes factors for common designs*. (2015). [Computer Software]

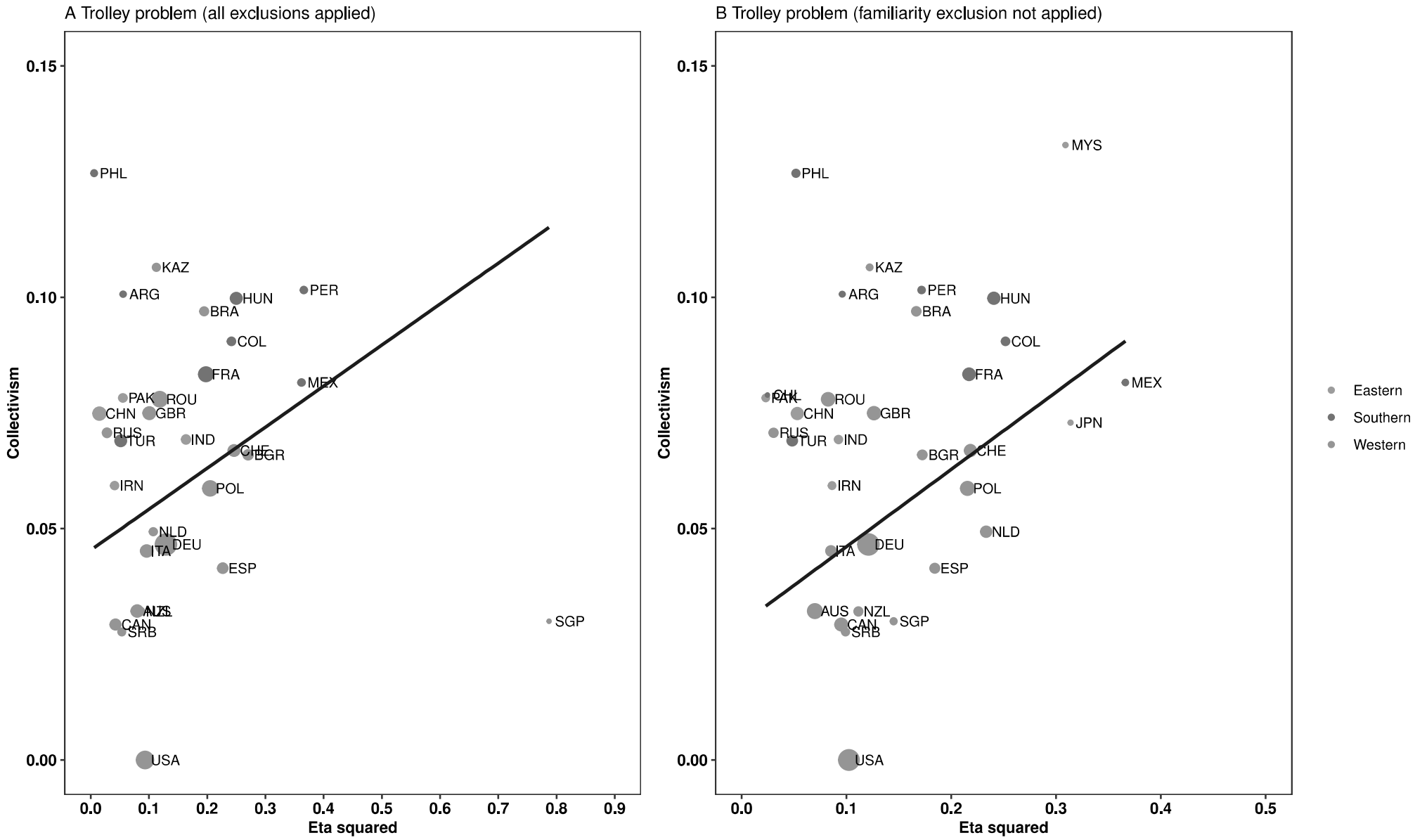
55. R Core Team. *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.* (2018). [Computer Software]

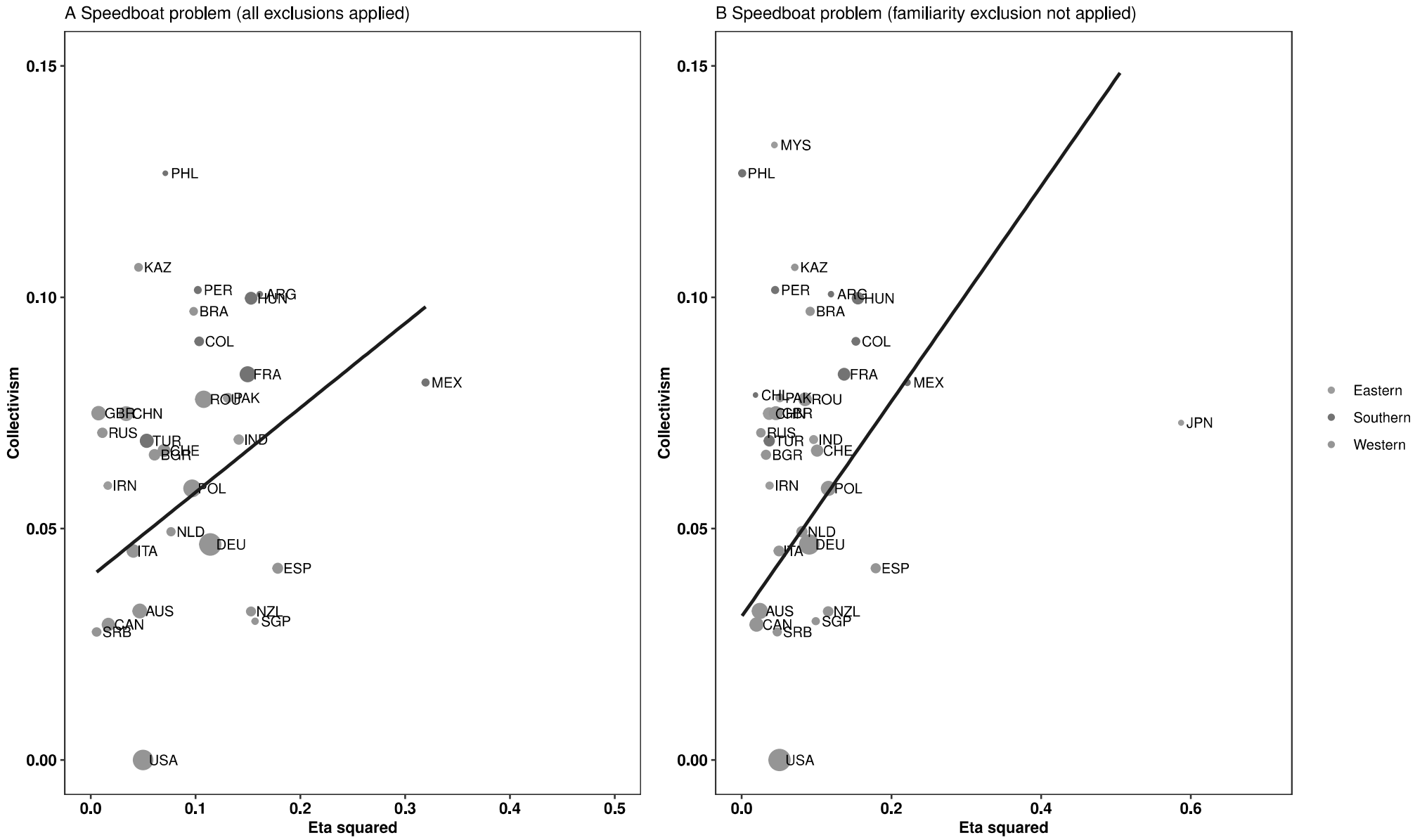
56. Dienes, Z. How do I know what my theory predicts? Accessed via: https://psyarxiv.com/yqaj4. *Adv. Methods Pract. Psychol. Sci.* **2**, 364–377 (2019).

57. Muthukrishna, M. *et al.* Beyond WEIRD psychology: measuring and mapping scales of cultural and psychological distance. *Psychol. Sci.* **31.** 678-701 (2018).


A Trolley problem – all exclusion criteria applied
B Trolley problem – familiarity exclusion not applied
C Speedboat problem – all exclusion criteria applied
D Speedboat problem – familiarity exclusion not applied
Eastern
Southern
Western
To what extent is this action morally acceptable?
1
2
3
4
5
6
7
8
9
No personal force
Personal Force


A Trolley problem – all exclusion criteria applied
B Trolley problem – familiarity exclusion not applied
C Speedboat problem – all exclusion criteria applied
D Speedboat problem – familiarity exclusion not applied
Eastern
Southern
Western
To what extent is this action morally acceptable?
1
2
3
4
5
6
7
8
9
Intention
No Intention
No personal force
Personal Force


A Trolley problem (all exclusions applied)
B Trolley problem (familiarity exclusion not applied)
Collectivism
Eta squared
0.00
0.05
0.10
0.15
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
PHL
KAZ
ARG
HUN
PER
BRA
COL
FRA
MEX
PAK
ROU
CHN
GBR
RUS
TUR
IND
CHE
BGR
IRN
POL
NLD
DEU
ITA
ESP
AUS
NZL
CAN
SRB
SGP
USA
MYS
JPN
Eastern
Southern
Western


A Speedboat problem (all exclusions applied)
B Speedboat problem (familiarity exclusion not applied)
Collectivism
Eta squared
0.00
0.05
0.10
0.15
0.0
0.1
0.2
0.3
0.4
0.5
0.6
PHL
KAZ
PER
ARG
HUN
BRA
COL
FRA
MEX
ROU
PAK
GBR
CHN
RUS
TUR
IND
CHE
BGR
IRN
POL
NLD
DEU
ITA
ESP
AUS
NZL
SGP
CAN
SRB
USA
MYS
CHL
JPN
Eastern
Southern
Western