

1 Inventory of Supporting Information

2

3 **Manuscript #:** NATHUMBEHAV-19046333G

4

5 **Corresponding author name(s):** Bence Bago

6 1. Extended Data

Figure #	Figure title	Filename	Figure Legend
Extended Data Fig. 1	Effect of physical force on moral dilemma judgments, no exclusion criteria applied	extended_fig1.jpg	<i>Results in Study 1 (effect of personal force without applying any exclusion criteria) on the Trolley (a) and Speedboat dilemma (b). Error bars are 95% Confidence Intervals around the mean. Scale ranged from 1 (completely unacceptable) to 9 (completely acceptable), n = 7,744.</i>
Extended Data Fig. 2	The effect of personal force on moral dilemma judgements (without applying any exclusion criteria)	extended_fig2.jpg	
Extended Data Fig. 3	Interaction between personal force and intention on moral judgments (without applying any exclusion criteria)	extended_fig3.jpg	
Extended Data Fig. 4	Effect of physical force	extended_fig4.jpg	<i>Results in Study 2 (personal force and intention</i>

	and intention on moral dilemma judgments, no exclusion criteria applied		<i>interaction without applying any exclusion criteria) on the Trolley (a) and Speedboat dilemma (b). Error bars represent 95% Confidence Intervals. Scale ranged from 1 (completely unacceptable) to 9 (completely acceptable), n = 19,340.</i>
Extended Data Fig. 5	Correlation between country level collectivism and personal force and intention interaction effect size, no exclusion criteria applied	extended_fig5.jpg	<i>Correlation between country-level collectivism and the Eta squared effect size of the interaction between personal force and intention with no exclusion criteria applied on the Trolley (a) and Speedboat dilemma (b). The size of the circles indicate the size of the sample in a given country. Blue line is the weighted regression line.</i>
Extended Data Fig. 6	Individualism/collectivism associations with the interaction between personal force and intention on moral judgments (without applying any exclusion criteria)	extended_fig6.jpg	

7

8

9 **2. Supplementary Information:**10 **A. Flat Files**

Item	Present?	Filename	A brief, numerical description of file contents.
Supplementary Information	Yes	Supplementary Information.pdf	Supplementary Methods 1-4, Supplementary Tables 1 - 10, Supplementary Figures 1-5,
Reporting Summary	Yes	nr-reporting-summary_moral.pdf	
Peer Review Information	Yes	PRFile_Bago.pdf	

11

12

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

15

16 Bence Bago¹, Marton Kovacs^{2,3}, John Protzko⁴, Tamas Nagy², Zoltan Kekecs², Bence Palfi^{5,6},
17 Matus Adamkovic^{7,8}, Sylwia Adamus⁹, Sumaya Albaloooshi¹⁰, Nihan Albayrak-Aydemir¹¹,
18 Ilham N. Alfian¹², Sinan Alper¹³, Sara Alvarez-Solas¹⁴, Sara G. Alves¹⁵, Santiago Amaya¹⁶,
19 Pia K. Andresen¹⁷, Gulnaz Anjum¹⁸, Daniel Ansari¹⁹, Patrícia Arriaga²⁰, John Jamir Benzon
20 R. Aruta²¹, Alexios Arvanitis²², Peter Babincak⁷, Krystian Barzykowski⁹, Bana Bashour²³,
21 Ernest Baskin²⁴, Luisa Batalha²⁵, Carlota Batres²⁶, Jozef Bavolar²⁷, Fatih Bayrak²⁸, Maja
22 Becker²⁹, Benjamin Becker³⁰, Anabel Belaus³¹, Michał Białek³², Ennio Bilancini^{33,34}, Daniel
23 Boller³⁵, Leonardo Boncinelli³⁶, Jordane Boudesseul³⁷, Benjamin T. Brown³⁸, Erin M.
24 Buchanan³⁹, Muhammad M. Butt⁴⁰, Dustin P. Calvillo⁴¹, Nate C. Carnes⁴¹, Jared B.
25 Celniker⁴², Christopher R. Chartier⁴³, William J. Chopik⁴⁴, Poom Chotikavan⁴⁵, Hu Chuan-
26 Peng⁴⁶, Rockwell F. Clancy⁴⁷, Ogeday Çoker⁴⁸, Rita C. Correia¹⁵, Vera Cubela Adoric⁴⁹,
27 Carmelo P. Cubillas⁵⁰, Stefan Czoschke⁵¹, Yalda Daryani⁵², Job A. M. de Grefte^{53,54}, Wieteke
28 C. de Vries⁵⁵, Elif G. Demirag Burak⁵⁶, Carina Dias⁵⁷, Barnaby J. W. Dixon⁵⁸, Xinkai Du⁵⁹,
29 Francesca Dumančić⁶⁰, Andrei Dumbravă^{61,62}, Natalia B. Dutra⁶³, Janina Enachescu⁶⁴, Celia
30 Esteban-Serna⁶⁵, Luis Eudave⁶⁶, Thomas R. Evans⁶⁷, Gilad Feldman⁶⁸, Fatima M.
31 Felisberti⁶⁹, Susann Fiedler⁷⁰, Andrej Findor⁷¹, Alexandra Fleischmann⁷², Francesco Foroni²⁵,
32 Radka Francová⁷³, Darius-Aurel Frank⁷⁴, Cynthia H. Y. Fu⁷⁵, Shan Gao⁷⁶, Omid Ghasemi⁷⁷,
33 Ali-Reza Ghazi-Noori⁷⁵, Maliki E. Ghossainy⁷⁸, Isabella Giammusso⁷⁹, Tripat Gill⁸⁰, Biljana
34 Gjoneska⁸¹, Mario Gollwitzer⁸², Aurélien Graton⁸³, Maurice Grinberg⁸⁴, Agata Groyecka-
35 Bernard³², Elizabeth A. Harris⁸⁵, Andree Hartanto⁸⁶, Widad A. N. M. Hassan⁷⁵, Javad
36 Hatami⁵², Katrina R. Heimark⁸⁷, Jasper J. J. Hidding¹⁰, Evgeniya Hristova⁸⁴, Matej Hruška⁷¹,
37 Charlotte A. Hudson⁸⁸, Richard Huskey⁸⁹, Ayumi Ikeda⁹⁰, Yoel Inbar⁹¹, Gordon P. D.
38 Ingram⁹², Ozan Isler^{93,94}, Chris Isloi⁹⁵, Aishwarya Iyer⁹⁶, Bastian Jaeger⁹⁷, Steve M. J.
39 Janssen⁹⁸, William Jiménez-Leal⁹², Biljana Jokic⁹⁹, Pavol Kačmár²⁷, Veselina Kadreva⁸⁴,
40 Gwenaël Kaminski²⁹, Farzan Karimi-Malekabadi¹⁰⁰, Arno T. A. Kasper¹⁰, Keith M.
41 Kendrick³⁰, Bradley J. Kennedy¹⁰¹, Halil E. Kocalar¹⁰², Rabia I. Kodapanakkal⁹⁷, Marta
42 Kowal³², Elliott Kruse¹⁰³, Lenka Kučerová⁷³, Anton Kühberger¹⁰⁴, Anna O. Kuzminska¹⁰⁵,
43 Fanny Lalot¹⁰⁶, Claus Lamm¹⁰⁷, Joris Lammers⁷², Elke B. Lange¹⁰⁸, Anthony Lantian¹⁰⁹, Ivy
44 Y.-M. Lau⁸⁶, Ljiljana B. Lazarevic⁹⁹, Marijke C. Leliveld⁵⁵, Jennifer N. Lenz¹¹⁰, Carmel A.
45 Levitan¹¹¹, Savannah C. Lewis⁴³, Manyu Li¹¹², Yansong Li^{113,114}, Haozheng Li¹¹⁵, Tiago J. S.
46 Lima¹¹⁶, Samuel Lins⁵⁷, Marco Tullio Liuzza¹¹⁷, Paula Lopes⁵⁷, Jackson G. Lu¹¹⁸, Trent
47 Lynds¹¹⁹, Martin Máčel⁷³, Sean P. Mackinnon¹²⁰, Madhavilatha Maganti¹²¹, Zoe Magraw-
48 Mickelson⁸², Leon F. Magson⁸⁸, Harry Manley⁴⁵, Gabriela M. Marcu¹²², Darja Maslić
49 Seršić⁶⁰, Celine-Justine Matibag¹²³, Alan D. A. Mattiassi³⁴, Mahdi Mazidi¹²⁴, Joseph P.
50 McFall¹²⁵, Neil McLatchie¹²³, Michael C. Mensink¹²⁶, Lena Miketta¹²⁷, Taciano L.
51 Milfont¹²⁸, Alberto Mirisola⁷⁹, Michal Misiak^{32,129}, Panagiotis Mitkidis^{74,130}, Mehrad Moeini-
52 Jazani¹⁰, Arash Monajem⁵², David Moreau¹³¹, Erica D. Musser¹³², Erita Narhetali¹³³, Danielle

53 P. Ochoa¹³⁴, Jerome Olsen¹²⁷, Nicholas C. Owsley¹³⁵, Asil A. Özdoğru¹³⁶, Miriam Panning¹³⁷,
 54 Marietta Papadatou-Pastou¹³⁸, Neha Parashar⁹⁶, Philip Pärnamets^{139,140}, Mariola Paruzel-
 55 Czachura¹⁴¹, Michal Parzuchowski¹⁴², Julia V. Paterlini⁴³, Jeffrey M. Pavlacic¹⁴³, Mehmet
 56 Peker¹⁴⁴, Kim Peters¹⁴⁵, Liudmila Piatnitskaia¹⁴⁶, Isabel Pinto¹⁴⁷, Monica Renee Policarpio¹⁴⁸,
 57 Nada Pop-Jordanova⁸¹, Annas J. Pratama¹³³, Maximilian A. Primbs¹⁷, Ekaterina Pronizius¹⁰⁷,
 58 Danka Purić⁹⁹, Elisa Puvia⁷⁹, Vahid Qamari⁵², Kun Qian¹⁴⁹, Alain Quiamzade^{150,151}, Beáta
 59 Ráczová²⁷, Diego A. Reinero⁸⁵, Ulf-Dietrich Reips¹⁵², Cecilia Reyna³¹, Kimberly
 60 Reynolds¹⁵³, Matheus F. F. Ribeiro¹⁵⁴, Jan P. Röer¹⁵⁵, Robert M. Ross¹⁵⁶, Petros Roussos¹⁵⁷,
 61 Fernando Ruiz-Dodobara⁸⁷, Susana Ruiz-Fernandez^{158,159}, Bastiaan T. Rutjens¹⁶⁰, Katarzyna
 62 Rybus¹⁴², Adil Samekin¹⁶¹, Anabela C. Santos^{162,163}, Nicolas Say¹⁶⁴, Christoph Schild¹⁶⁵,
 63 Kathleen Schmidt¹⁶⁶, Karolina A. Ściagała¹⁶⁷, MohammadHasan Sharifian⁵², Jiaxin Shi⁶⁸,
 64 Yaoxi Shi¹⁶⁸, Erin Sievers⁴³, Miroslav Sirota¹⁶⁹, Michael Slipenkyj¹⁷⁰, Çağlar Solak¹⁷¹,
 65 Agnieszka Sorokowska³², Piotr Sorokowski³², Sinem Söylemez¹⁷¹, Niklas K. Steffens¹⁷², Ian
 66 D. Stephen¹⁷³, Anni Sternisko⁸⁵, Laura Stevens-Wilson⁸⁸, Suzanne L. K. Stewart¹⁰¹, Stefan
 67 Stieger¹¹⁰, Daniel Storage¹⁷⁴, Justine Strube¹⁷⁵, Kyle J. Susa¹⁷⁶, Raluca D. Szekely-
 68 Copîndean^{177,178}, Natalia M. Szostak¹⁴², Bagus Takwin¹³³, Srinivasan Tatachari¹⁷⁹, Andrew
 69 G. Thomas¹⁸⁰, Kevin E. Tiede¹⁸¹, Lucas E. Tiong⁸⁶, Mirjana Tonković⁶⁰, Bastien
 70 Trémolière¹⁸², Lauren V. Tunstead⁸⁸, Belgüzar N. Türkan⁴⁸, Mathias Twardawski⁸², Miguel
 71 A. Vadillo⁵⁰, Zahir Vally^{183,184}, Leigh Ann Vaughn¹⁸⁵, Bruno Verschuere¹⁶⁰, Denis
 72 Vlašiček⁶⁰, Martin Voracek¹⁰⁷, Marek A. Vranka¹⁸⁶, Shuzhen Wang¹¹⁵, Skye-Loren West⁸⁸,
 73 Stephen Whyte^{93,94}, Leigh S. Wilton¹⁸⁷, Anna Wlodarczyk¹⁸⁸, Xue Wu⁹⁰, Fei Xin³⁰, Su
 74 Yadanar⁴, Hiroshi Yama¹⁸⁹, Yuki Yamada¹⁹⁰, Onurcan Yilmaz¹⁹¹, Sangsuk Yoon¹⁹², Danielle
 75 M. Young¹⁹³, Ilya Zakharov¹⁹⁴, Rizqy A. Zein¹², Ingo Zettler¹⁹⁵, Iris L. Žeželj⁹⁹, Don C.
 76 Zhang¹⁹⁶, Jin Zhang¹¹⁵, Xiaoxiao Zheng^{30,197}, Rink Hoekstra¹⁹⁸, Balazs Aczel²

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

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

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

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

215

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

219

220

221 **Abstract**

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

239

240

241

242

243

244 Introduction

245 Moral dilemmas can be portrayed as decisions between two main conflicting moral
246 principles: utilitarian and deontological. Utilitarian (also referred to as consequentialist)
247 philosophies¹ hold that an action is morally acceptable if it maximizes well-being for the
248 greatest number of people (in terms of saved lives, for example). On the other hand,
249 deontological philosophy² evaluates the morality of the action based on the intrinsic nature of
250 the action (i.e., the deontological option often reflects greater concern for the individual rights
251 and duties³). The dilemma between these two principles plays a prominent role in law and
252 policy-making decisions, ranging from decisions of health budget allocations⁴ to the dilemma
253 of self-driving vehicles⁵. This inherent conflict is well illustrated by the so-called trolley
254 problem, which has long interested both philosophers and psychologists. One version of the
255 dilemma is presented as follows⁶:

256
257 You are a railway controller. There is a runaway trolley barrelling down the railway tracks.
258 Ahead, on the tracks, there are 5 workmen. The trolley is headed straight for them and they
259 will be killed if nothing is done. You are standing some distance off in the train yard, next to
260 a lever. If you pull this lever, the trolley will switch to a side track and you can save the 5
261 workmen on the main track. You notice that there are 2 workmen on the side track. So there
262 will be 2 workmen who will be killed if you pull the lever and change the tracks but the 5
263 workmen on the main track will be saved. Is it morally acceptable for you to pull the lever?
264

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

271
272 In an alternative version of the dilemma, one has to push a man off a footbridge in front of
273 the trolley (“footbridge” scenario). This man will die but will stop the trolley, and the five
274 people in the way of the trolley will be saved. Interestingly, people are less likely to make a
275 decision consistent with utilitarian perspectives in the footbridge scenario compared to the
276 standard switch scenario (We call these “utilitarian” responses but the fact that these
277 decisions are consistent with utilitarianism does not indicate that people gave them out of
278 utilitarian principles; the same is true for “deontological” responses^{7,8}). The difference
279 between the utilitarian response rate in those scenarios became the basis of investigations of
280 many influential cognitive theories in the field of moral judgement^{3,7-13}. The fact that people
281 respond differently to the two trolley dilemmas was proposed to be explained by people’s
282 adherence to the so-called doctrine of double effect^{6,9}. A simple version of this doctrine is
283 that harm is permissible as an unintentional side-effect of a good result. This doctrine is the
284 basis of many policies in several countries all around the world concerning issues such as
285 abortion⁶, euthanasia¹⁴, international armed conflict regulations^{15,16}, and even international
286 business ethics¹⁷. According to this doctrine, it is morally impermissible to bomb civilians to
287 win a war, even if ending the war would eventually save more lives. However, if civilians die

288 in a bombing of a nearby weapons factory as a side-effect, the bombing is morally
289 acceptable. The way people perceive or act on these moral rules can influence the policies
290 that are accepted or even followed - as we can already see in the case of driverless cars,
291 which sometimes have to decide between sacrificing their own passengers and saving one or
292 more pedestrians⁵.

293

294 Greene et al.¹⁸ and Cushman et al.⁹, however, argued that the difference in utilitarian
295 response rates cannot simply be explained by the doctrine of double effect. Greene et al.
296 presented evidence for the interaction of the intention of harm (i.e., harm as means or side
297 effect; referring to the doctrine of double effect) and personal force (i.e., whether or not the
298 agent had to use personal effort to kill the victim and save more people) on moral
299 acceptability ratings. More concretely, people were less likely to judge sacrificing one person
300 to save more people as morally acceptable when they had to use their personal force to kill
301 the person *and* the death of this person was required to save more people (this is what is
302 meant by *intending* the harm). Hence, they concluded that people are more sensitive to the
303 doctrine of double effect when they have to use their own physical force. Despite some
304 exceptions^{26,27}, most of the evidence for this conclusion comes from samples of WEIRD
305 (Western, Educated, Industrialized, Rich, Democratic^{23,24}) societies, leaving the question open
306 of whether these effects are psychologically universal²⁵ or culture-specific.

307

308 This study tests three cross-cultural hypotheses:

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

315

316 The first and second hypotheses, that the effects of personal force and intention on moral
317 judgements are culturally universal, come from their relatedness to interpersonal violence.
318 People seem to exhibit a general tendency to avoid causing violent harms (e.g., murder)^{19,20},
319 and they are more likely to perceive actions as violent or harmful when they are supposed to
320 use personal force or intention³. As a result, people are more likely to behave in a
321 deontological way when personal force or intention is present in the dilemma. As all cultures
322 regulate interpersonal violence,²¹ we expected to find that both intention and personal force,
323 as well as their interaction, have an effect on moral judgements across cultures. The literature
324 seems to be in accordance with these hypotheses. For example, Chinese²⁵⁻²⁷ and Russian²⁸
325 participants responded similarly to moral dilemmas as Americans and Western Europeans,
326 and even small-scale societies tended to be susceptible to the effect of intention^{22,23}.

327

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

332 use of personal force makes people feel guilt or shame before making a decision and,
333 therefore, rating actions that use personal force as morally less acceptable. There is a
334 convincing argument that these social emotions are universal²⁸⁻³⁰, despite some cultural
335 variation in their intensity and the social contexts in which they are experienced²⁸⁻³⁰. It has
336 been argued that shame and guilt are more important in interdependent, collectivistic cultures
337 (as their function is argued to be linked to social control). People living in East Asian
338 countries have reported experiencing these emotions more frequently and more intensely²⁸⁻³⁰.
339 Other findings suggest that it is anxiety that mediates the effect of intention and personal
340 force²⁶, but anxiety (social anxiety in particular) has also been positively associated with
341 collectivism³¹, pointing to the same direction. Hence, we hypothesized that people living in
342 collectivistic cultures would judge actions that involve personal force and intention as
343 morally less acceptable than people in individualistic cultures. Utilitarian responding in moral
344 dilemma judgements has also been associated with low levels of empathic concern³² and
345 people living in collectivistic cultures have been suggested to exhibit higher levels of
346 empathic concern^{33,34}. Hence, we predicted that individualism-collectivism would also have
347 an effect on utilitarian responding: collectivists would be less utilitarian in general, due to
348 their higher levels of empathic concern.

349
350 In addition to testing our confirmatory hypotheses, we also collected a number of additional
351 country-level as well as individual measures for exploratory purposes. These measures have
352 been previously shown to be related to moral judgement such as economic status³⁵, individual
353 level individualism-collectivism³⁵, and religiosity³⁶. We also administered an alternative
354 measure of utilitarian responding³⁷⁻⁴⁰.

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

- 356
- 357 1) The original article has been very influential (515 citations so far), but replicability
358 has not established yet.
 - 359 2) Our knowledge is scarce on the cultural universality of the effect of personal force
360 and intention in moral judgements.
 - 361 3) The resulting database (with many types of trolley problems and additional measures)
362 could assist and guide future research and applications on moral thinking.

363 **Overview**

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

373

374 **Results**

375

376 We collected data from 27,502 participants out of 45 countries. Due to our exclusion criteria,
 377 we had to exclude 80.6% of the sample for the main analysis (see Table 1 for the various
 378 exclusion criteria). Note that, as we registered, we conducted the analysis without excluding
 379 the data of the participants who were familiar with the trolley problem (36.2% exclusions),
 380 and we also conducted a post-hoc explorative analysis in which we applied no exclusion
 381 criteria. All participants were presented with two moral dilemmas that were equivalent in
 382 structure but were different in wording: trolley dilemmas and speedboat dilemmas (the
 383 former described a situation involving trolley and people on the tracks, the latter described a
 384 situation with people on a speedboat and others drowning in the sea). In Study 1, we tested
 385 the effect of personal force on moral dilemma judgments (Hypothesis 1a, 1b), while in Study
 386 2, we tested the interaction effect between personal force and intention (Hypothesis 2a, 2b,
 387 3).

388

389 **The effect of personal force**

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

401

402

403 **The interaction effect of personal force and intention**

404 Figure 2 shows when we applied all exclusion criteria, we found strong evidence in the
 405 Western cluster (H2a) for the interaction between personal force and intention ($BF_{10} =$
 406 $1.5 * 10^{11}$), but moderate inconclusive evidence in the Southern ($BF_{10} = 9.4$) and weak,
 407 inconclusive evidence in the Eastern clusters ($BF_{10} = 0.6$) (H2b). More concretely, in the
 408 Western cluster, participants judged the acceptability of consequentialist decisions much
 409 lower when both personal force and intention had to be applied (i.e., the personal force effect
 410 was numerically greater when intention also had to be applied). When we included
 411 participants who were familiar with the trolley dilemma, we still found strong evidence in the
 412 Western cluster ($BF_{10} = 1.28 * 10^{30}$) and, interestingly, we also found strong evidence in the
 413 Southern cluster ($BF_{10} = 3.1 * 10^6$), but the evidence remained weak and inconclusive in the
 414 Eastern cluster ($BF_{10} = 2.9$). Although in the preregistration we expected the effect sizes to be
 415 smaller when participants familiar with the trolley problem were included, we observed the
 416 direct opposite: when including data of participants familiar with the trolley problem, we

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

421
 422 On the speedboat dilemmas, we found strong evidence for the interaction in the Western
 423 cluster, regardless of the familiarity exclusion ($BF_{\text{all exclusions}} = 222$, $BF_{\text{with familiar}} = 4.8 \times 10^7$).
 424 However, we found inconclusive evidence in the Eastern and Southern clusters, both before
 425 ($BF_{\text{Eastern}} = 0.4$; $BF_{\text{Southern}} = 0.4$) and after ($BF_{\text{Eastern}} = 0.4$; $BF_{\text{Southern}} = 1.1$) familiarity
 426 exclusions. Although our results were consistent in the Western and Eastern clusters for both
 427 the speedboat and trolley dilemmas, there was a divergence in the Southern cluster.
 428 Specifically, we found strong evidence only for the interaction in the Southern cluster when
 429 we included familiar participants in the analysis. In general, in all clusters, the observed
 430 effect sizes were smaller on the speedboat than on the trolley dilemma.

431
 432 In summary, we conclude that we fully replicated the findings of Greene et al. with respect to
 433 the interaction of personal force and intention in the Western cluster (H2a) regardless of
 434 dilemma context or exclusion criteria. However, the evidence was inconclusive for all
 435 analyses of the Eastern cluster. In the Southern cluster, the conclusion is both context-
 436 dependent (i.e., the effect was only detectable in the trolley dilemma) and sensitive to
 437 exclusion criteria (i.e., the effect was only detectable when familiar participants were
 438 included).

439
 440 To explore whether our results were sensitive to our choice of priors in the Bayesian analysis,
 441 we computed Robustness Regions (“RR”) that indicate the region of priors within which our
 442 inference would remain unchanged. The width of this region shows how robust our
 443 inferences are to our selection of priors. The RRs were generally wide for all statistical tests
 444 (see Tables 2-3), indicating that our results were not sensitive to our choices of prior. Thus,
 445 we would arrive at the same conclusions with any possible prior within the realistic range.
 446 One exception to this finding where the final conclusion was prior-dependent can be found in
 447 the analysis of the Southern cluster in Study 2. Specifically, if the scale of the prior
 448 distribution had been $r = .21$ or higher (instead of $r = .19$), we could have concluded that
 449 there was strong evidence for the effect (instead of saying that the test is inconclusive). Here,
 450 we would like to stress that we did not reach our registered sample size in this cluster for
 451 Study 2 (we registered that for 95% power, we would need 1,800 participants in each cluster
 452 of which we only reached 690 - see the Methods for details on sample size estimation). This
 453 could explain why our results did not reach our evidence thresholds and remained
 454 inconclusive.

455 **Cultural correlates**

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

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

462

463 H3 stated that we expected a three-way interaction between country-level collectivism,
 464 intention, and personal force. We first tested this hypothesis on the data with familiarity
 465 exclusion applied (see Table 4 for statistical results and Figure 3 for the graphical
 466 representation of findings). The results of the country-level collectivism scale were
 467 inconclusive (trolley: $BF_{10} = 1.2$; speedboat: $BF_{10} = 0.9$). When analysing the individual-level
 468 measures of horizontal and vertical collectivism/individualism, all results were inconclusive.
 469 We conducted the same analysis on the sample but this time including participants who were
 470 familiar with these types of moral dilemmas, but the results were still inconclusive (trolley:
 471 $BF_{10} = 2.2$; speedboat: $BF_{10} = 0.7$). Analysing the individual-level individualism/collectivism
 472 measures, we found inconclusive evidence in all the scales. In the Introduction (Stage 1), we
 473 also hypothesized that country level collectivism would be associated with decreased overall
 474 acceptability of the utilitarian option. This hypothesis was not included in the registered
 475 analysis plan. Nevertheless, we added this analysis to the Supplementary Analysis section 3.
 476 In short, we found no evidence for the association between country-level collectivism and
 477 moral acceptability rates. Interestingly, nevertheless, we found strong evidence for a positive
 478 correlation between vertical individualism and moral acceptability ratings.

479

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

483

484 **Exploratory analysis**

485

486 **The effect of intention**

487

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

496

497 **No exclusion analysis (post-hoc)**

498

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

504

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

517

518 We also conducted the cultural analysis without applying any exclusion criteria and we found
519 that all of the results were inconclusive, with one exception. In the speedboat dilemma, we
520 found moderate evidence that country level collectivism is positively associated with the
521 interaction of personal force and intention (in line with our hypothesis; $BF_{10} = 5.1$; same test
522 for the trolley dilemma: $BF_{10} = 2.8$). We also found moderate evidence ($BF_{10} = 9.8$) that in
523 the trolley dilemma, the interaction between personal force and intention is positively
524 associated with individual-level horizontal collectivism: being higher on horizontal
525 collectivism means a heightened personal force and intention interaction effect size (see
526 Extended Data Figures 5 and 6; same test in the speedboat dilemma was inconclusive: $BF_{10} =$
527 0.54). Thus, for the moderation of the effect by country-level collectivism, the strict
528 exclusion criteria may have hurt our ability to detect these effects. Although these results
529 appear in line with our prior hypothesis, this analysis was only exploratory, not registered a
530 priori, and hence, should only be interpreted with caution.

531

532 As we registered, we added a figure showing the distribution of responses of both subscales
533 of the Oxford Utilitarianism Scale for each country cluster, and also reported means and 95%
534 confidence intervals, as registered. Moreover, we also added a post-hoc analysis correlating
535 each subscales of the OUE with moral acceptability ratings of the moral dilemmas. We found
536 that moral acceptability ratings correlate higher with the “instrumental harm” sub-scale ($r =$
537 $0.40 - 0.45$) than with the “impartial beneficence” sub-scale ($r = 0.05 - 0.20$) - with this latter
538 correlation exhibiting somewhat larger cultural variations. Details can be found in the
539 Supplementary Analysis section 2.4.

540

541

542 Discussion

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

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

551

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

565

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

577

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

587

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

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

600

601 Although we found no strong evidence for the association between
602 collectivism/individualism and the effects of personal force and intention, future research
603 should test for other cultural variations. There are a number of interesting candidates that we
604 did not examine, including cultural tightness⁴³ and social mobility⁴⁴. Our database provides
605 opportunities to the field to examine different aspects and cultural moderators of moral
606 judgment.

607

608 This research has a number of limitations that future work will need to address. Although we
609 call the personal force effect “universal”, it is only universal to the cultures we tested. This
610 puts a limit to the “universality” of the effects: we did not (nor intended to) reach small scale
611 hunter gatherer societies for example. Moreover, while our sample was more diverse and less
612 WEIRD than that of Greene et al.’s research, it consisted of mostly educated individuals from
613 younger age groups with internet access, raising similar concerns (e.g., still Educated and
614 Industrialized, and possibly Rich, though not strictly Western or Democratic). Secondly, the
615 data collection was conducted before and during the COVID-19 pandemic which could have
616 affected the participants’ responding behaviour in some way (e.g., moral fatigue). Finally,
617 81% of the sample was not entered into the main confirmatory analyses because of our
618 exclusion criteria, which might have resulted in unintended selection biases. For example, it
619 is possible that more educated participants were more likely to get excluded due to being
620 familiar with moral dilemmas from college. It is also possible that people with less working
621 memory capacity or poor text comprehension abilities were more likely to be excluded due to
622 the stringent attention checks. This is why we included an exploratory analysis in which we
623 analysed data from all of our participants, without applying any exclusions. Our results on the
624 full sample (no exclusion criteria applied) supported our previous conclusions (that were
625 drawn based on the data with exclusions) except in the cultural analysis, in which we found
626 strong evidence for cultural variations only when no data were excluded. Thus, future work,
627 especially replication work, should take caution when applying stringent exclusion criteria as
628 it may be entirely unnecessary and even hurt the discovery of new effects.

629

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

635 for future research could be to identify whether personal force and intention increase reliance
636 on deontological rules or decrease reliance on consequentialist thinking. Methodological
637 approaches, such as process dissociation, are promising in this regard⁴⁰.

638

639

640 **Conclusion**

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

648

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

656

657

658

659 **Method**

660 **Participants**

661 A large culturally and demographically diverse sample of participants was recruited from
662 collaborating laboratories through the Psychological Science Accelerator⁴⁷. The data
663 collection team was originally proposed to include 146 labs from 52 countries. All of these
664 participating laboratories obtained IRB approval (verified before the last round of Stage 1
665 submission). Combined, these labs committed to collect a minimum number of 18,637
666 participants. More labs were expected to be recruited before data collection commences. Each
667 lab will recruit participants for the study by sending out the survey link along with the
668 consent form to their participant pool, online platforms (such as Mturk), or testing them in the
669 research lab. Due to some dropouts, the data collection team included 140 labs from 45
670 countries. Eligibility for participation was based on age (≥ 18 years) and being a native
671 speaker of the language of the test (more details on this criterion in the *Controlling for*
672 *possible confounds* section). Data were collected either from local university participant pools
673 or via data collection platforms (e.g., MTurk). Altogether, 41,090 participants started our
674 survey, and 27,502 finished it whose data were analysed (17961 females, 7956 males, Mean
675 age = 26.0 years, SD = 10.3 years; Study 1: 7744 participants, 4329 females, 2487 males,

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

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

685

686 **Materials**

687

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

710

711 **Additional measures.** Although the exploration of individual-level factors associated with
712 moral thinking is not the aim of the present research, to enrich our database for future studies
713 and secondary analyses, we expanded our survey with additional individual-level measures:
714 1) total yearly household income; 2) place of living (urban or rural area); 3) position on the
715 four-dimensional Individualism-Collectivism scale³⁴ (16 items) for disentangling cultural
716 differences in participants' responses⁴⁸; 4) religion: Specific religion of the participant will be
717 asked, plus one question to measure their level of religiosity: "On a scale from 1 to 10, how
718 religious are you?". Furthermore, we included the Oxford Utilitarianism Scale²⁸ (9 items).

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

722

723

724 **Procedure**

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

737

738

739 ***Controlling for possible confounds.*** To avoid second language effects on moral judgement⁴⁹,
740 only native speakers of the language of the experiment could participate. To ensure this, we
741 asked participants to indicate their native language(s). Bilinguals could choose their preferred
742 language. The data of anyone with a native language different from the language of the
743 survey were removed from data analyses.

744

745 Following Greene et al.'s procedure, data from participants who reported that they found the
746 material confusing were excluded from the analyses. Data from participants who reported
747 having experienced technical problems during the experiment were also excluded from all
748 analyses. To avoid careless responses, we added three bogus items at the end of the survey.
749 We asked participants very basic questions (e.g., "I was born on February 30th.") to which
750 incorrect answering indicates careless responding⁵⁰. We excluded data from participants who
751 gave an incorrect response to any of these questions. Moreover, we introduced two additional
752 questions (presented right after the moral dilemmas), asking participants about the specifics
753 of the trolley and speedboat scenarios that they had been presented with, to test whether they
754 had paid attention when reading the scenarios (referred to as attention check in the later test).
755 Specifically, participants were asked to select the option which most accurately described the
756 situation that they had been presented with. Each option described the nature of the physical
757 action that was the key manipulation in the experiment. As attention to the trolley and
758 speedboat dilemmas was measured by different questions, when analysing the responses, we
759 excluded the data for the correspondingly failed attention check question. For example,
760 people who gave a correct response on the trolley, but not on the speedboat attention check
761 question, were included when analysing the trolley dilemma, and excluded when analysing
762 the speedboat version.

763

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

776

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

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

787 The second important change in this experiment is that participants were presented with two
788 moral dilemmas in one condition, instead of one. These additional dilemmas will be analysed
789 separately, as they were in the original experiment. The third difference is that for Study 2,
790 we used different moral dilemmas than those that were used by Greene et al.; the standard
791 switch and footbridge dilemmas were used instead of the loop weight and obstacle push
792 dilemmas, respectively. These dilemmas are not different from the ones used by Greene et al.
793 in their structural characteristics, only on surface characteristics. That is, in the standard
794 switch the harm is unintended and no personal force is required, while in the standard
795 footbridge dilemma, the harm is intended and requires personal force. By including the
796 standard switch and standard footbridge scenarios instead of the original ones, we gain
797 further insight into the data. Imagine for example, that the personal force effect does not
798 replicate in one of the cultural clusters. One explanation for this is that people are simply not
799 sensitive to the effect of personal force in that cluster. However, it might also be the case that
800 utilitarian response rates to similar dilemmas increase over time⁵¹. If so, we should see that
801 the replicated difference between the standard footbridge and switch dilemmas is shrinking or
802 disappeared. Furthermore, by comparing the standard footbridge to the footbridge pole
803 dilemmas, we can test the effect of physical contact, and by comparing the standard switch
804 case to the footbridge switch case to confirm the effect of intention.

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

808

809 ***Cultural classification of countries.*** To test the cultural universality hypothesis, a
 810 comprehensive cultural classification is needed that encompasses multiple sources of cultural
 811 variability. Hence, to assess our first hypothesis on the universality of the effect of personal
 812 force and intention on moral judgements, we used the cultural classification of Awad et al.³⁵.
 813 Based on surveyed moral preferences, they identified three distinct clusters of countries:
 814 Eastern, Southern, and Western. They argued that this cluster structure is broadly consistent
 815 with the alternative, but more complex Inglehart-Welzel cultural map³⁴. Therefore, we
 816 assigned the countries of our participating labs to these cultural clusters, as listed in
 817 Supplementary Analysis Section 1, Table S1.

818

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

822

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

837

838 **Planned analyses**

839

840 **Preregistered analysis**

841

842 **Confirmatory Replication Analyses**

843

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

847

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

850

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

854

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

859

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

862

863

864 Unlike in the original study, we employed Bayesian analyses to gain information from our
 865 data concerning the strength of evidence for the null and alternative hypotheses. The Bayes
 866 factor indicates the relative evidence provided by the data comparing two hypotheses⁵².

867 Regarding the threshold of strong Bayesian evidence, we followed the recommendations of⁵³
 868 and set the decision threshold of BF_{10} to > 10 for H_1 and $< 1/10$ for H_0 . We used informed
 869 priors for the alternative model: a one-tailed Cauchy distribution with a mode of zero and a
 870 scale $r = 0.26$ (Hypothesis 1a and 1b) and $r = 0.19$ (Hypothesis 2a and 2b) on the
 871 standardized effect size using the BayesFactor package⁵⁴ in R for the analysis. These priors
 872 are based on the effect sizes that we expect to find as explained below in the sample size
 873 estimation section. We will implement all of our analyses with the R statistical software⁵⁵.

874

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

883

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

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

895

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

902

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

905

906 **Robustness analyses**

907 To probe the robustness of our conclusions to the scaling factor of the Cauchy distribution
908 used as the prior of H1, we reported Robustness Regions for each Bayes factor. Robustness
909 Regions were notated as RR[*min*, *max*], where *min* indicates the smallest and *max* indicates
910 the largest scaling factor that would lead us to the same conclusion as the originally chosen
911 scaling factor⁵⁶.

912

913 **Sampling plan and stopping rule**

914 As the data were planned to be collected globally, our knowledge was insufficient concerning
915 the noise of the measurement and the rate of exclusion in the various samples, which were
916 needed for an accurate sample size estimation. For this reason, we proposed a sequential data
917 acquisition. That is, first, we launched Study 1 (Hypotheses 1a and 1b), and collected data in
918 sequences from 500 participants per cluster per condition; from 3,000 participants altogether
919 (after all exclusions). We stop data collection after each sequence. At these stops, we
920 conducted our planned Bayesian analyses. Should the BF reach the preset thresholds in a
921 given cluster, we will stop data collection for that cluster. If, in a cluster, the BF thresholds
922 were not reached, we would continue data collection with 200 additional participants per
923 cluster per condition, and then re-analyse the data, repeating this procedure until one of the
924 BF thresholds is reached, or the participant pool is exhausted. Note, however, that we
925 deviated from this sampling plan. See “Deviations from registration” for details.

926 Should we not have reached this limit with our planned capacity of ~19,000 participants, we
927 would have extended the data collection to a new semester. In the case that we would have
928 not reached our evidence threshold within 12 months, we would have reported our final
929 results, acknowledging the limited strength of the findings.

930 We launched Study 2 data collection in a given cluster only when the analysis of Study 1 was
931 conclusive. In Study 2, we conducted the analysis only when we had exhausted our resources.

932

933

934 **Sample Size estimation**

935

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

943

944 **Testing the association between country-level collectivism and the effects of personal** 945 **force and intention**

946

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

959

960 **Analysis of the additional moral dilemmas**

961 **Study 1.**

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

965

966

967 **Study 2.**

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

970

971 **Further tests**

972

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

980

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

986

987 **Oxford Utilitarianism Scale.** We computed a figure showing the response distribution of
988 each subscales of the Oxford Utilitarianism Scale³⁹ for each cultural cluster to explore
989 potential cultural differences (along with means and 95% CI). The results of this can be found
990 in the Supplementary Analysis section 2.4.

991

992 **Individual-level horizontal and vertical individualism-collectivism.** Triandis and
993 Gelfand⁴⁵ defined individualistic and collectivistic cultural tendencies among 4 dimensions:
994 *vertical individualism, vertical collectivism, horizontal individualism, and horizontal*
995 *collectivism*. We added these continuous measures to our Bayesian linear regression analysis.
996 The predictive power of all four measures were assessed separately.

997

998 **Including familiar participants.** A potentially large number of participants were excluded
999 due to familiarity with the trolley dilemma, and there was a possibility that this exclusion
1000 criterion will affect the data from some countries or cultural clusters more than others. To
1001 avoid this potential sampling bias, we computed all above-listed analyses on moral dilemmas
1002 (confirmatory and exploratory) on the full sample in which we did not exclude the
1003 participants who were familiar with moral dilemmas. Second, we computed all analyses
1004 specifically on data coming from people who were familiar with moral dilemmas in order to
1005 compare the results of “familiar” and “unfamiliar” participants. This latter analysis can be
1006 found in the Supplementary Analysis section 2.3 and was limited to the confirmatory
1007 hypothesis tests.

1008

1009 **Pilot testing.** To ascertain that the survey software operates without any technical problems,
1010 we planned to conduct a pilot test in which each participating lab would have been expected
1011 to collect data from 10 participants. We would have only assessed the expected functioning
1012 of the survey software without analysing the collected data.

1013

1014 **Timeline.** We planned to finish data collection within six months from Stage 1 in principle
1015 acceptance and we planned to submit our report within one month from then.

1016

1017 **Deviations from registration**

1018 We preregistered that we would collect data from 3,000 participants for Study 1 (test of
1019 personal force; H1a, H1b), after exclusions. Unexpectedly, the exclusion criteria led to 80.6%
1020 exclusion of our collected data. At the point when this was realized, it seemed likely that
1021 Study 1 would exhaust the available sample pool, not leaving capacity for Study 2. Therefore,
1022 with the agreement of the journal editor, we decided to collect participants for Study 1 only

1023 until our Bayes Factor evidence thresholds were reached after all exclusion criteria were
1024 applied. This modification allowed us to collect data for Study 2 as well.

1025

1026 At the time of this decision, the distribution of responses has been taken into account: we had
1027 collected data from 3,473 participants: 1319 from the “Western cluster”, 1762 from the
1028 “Southern” cluster, and 392 from the “Eastern” cluster. After exclusions, 789 participants
1029 remained (78% excluded): 296 from the “Western” cluster (78% excluded), 429 from the
1030 “Southern” cluster (76% excluded), and 64 from the “Eastern” cluster (84% excluded).

1031

1032 Instead of conducting a pilot study as preregistered, in order to avoid wasting any (much
1033 needed) participants, participating researchers from all labs tested the experiment before it
1034 was sent out to assure that there are no grammatical mistakes or functionality problems.

1035

1036 Due to COVID-19 crisis, data collection took 6 months longer than expected (with the
1037 agreement of the editor).

1038

1039 **Exploratory analysis**

1040 During the data pre-processing, we excluded 229 participants from three US-based labs as
1041 they received a wrong survey link. Furthermore, 13,359 participants started, but did not finish
1042 the experiment, therefore their data were also dropped from further analyses. These
1043 participants did not count towards our final sample and are not part of the data in any way.
1044 The final sample used for data analyses consisted of 27,502 participants. Further information
1045 on the demographics of our participants can be found in the Supplementary Analysis section
1046 1.

1047

1048 Note that we limited the use of Robustness Regions for the confirmatory hypothesis tests.

1049

1050

1051 **Data availability statement**

1052 Collected anonymised raw and processed data are publicly shared on the Github page of the
1053 project: <https://github.com/marton-balazs-kovacs/trolleyMultilabReplication/tree/master/data>.

1054

1055 **Code availability statement**

1056 Code for data management and statistical analyses have been written in R and are available
1057 at: <https://github.com/marton-balazs-kovacs/trolleyMultilabReplication>.

1058

1059 **Protocol Registration Information**

1060 The Stage 1 protocol for this Registered Report was accepted in principle on 30th January
1061 2020. The protocol, as accepted by the journal, can be found at

1062 <https://doi.org/10.6084/m9.figshare.11871324.v1>

1063

1064

1065

1066

1067 **Acknowledgements**

1068 M.A. Vadillo was supported by 2016-T1/SOC-1395 and 2020-5A/SOC-19723 from
 1069 Comunidad de Madrid; PSI2017-85159-P from AEI and UE/FEDER; M.P.-C. was supported
 1070 by 2017/01/X/HS6/01332 from the National Science Centre, Poland; P.M. was supported by
 1071 Aarhus University Research Foundation (AUFF), Starting Grant: AUFF-E-2019-9-4; B. Bago
 1072 was supported by ANR grant ANR-17-EURE-0010 (Investissements d’Avenir program), and
 1073 ANR Labex IAST; R.M.R. was supported by Australian Research Council (DP180102384);
 1074 N.B.D. was supported by CAPES grant number 88887.364180/2019-00; C.S., K.A.Ś., and I.
 1075 Zettler were supported by Carlsberg Foundation (CF16-0444), Independent Research Fund
 1076 Denmark (7024-00057B); J.L. was supported by EXC 2126/1– 390838866 under Germany’s
 1077 Excellence Strategy; K.B. was supported by following grants from the National Science
 1078 Centre, Poland: (1) while working on the data collection: No.: 2015/19/D/HS6/00641, (2)
 1079 while working on the final version of the paper: No.: 2019/35/B/HS6/00528; A.W. was
 1080 supported by FONDECYT 11190980, CONICYT; A. Fleischmann was supported by German
 1081 Research Foundation (research unit grant FOR-2150, LA 3566/1-2); H.Y. was supported by
 1082 JSPS Grant (18K03010); Y.Y. was supported by JSPS KAKENHI (16H03079, 17H00875,
 1083 18K12015, and 20H04581); K.Q. was supported by JSPS KAKENHI (17H06342, 20K03479,
 1084 and 20KK0054); A. Ikeda was supported by JSPS KAKENHI (20J21976); K.M.K. was
 1085 supported by National Natural Science Foundation of China (NSFC) (grant number
 1086 31530032) and Key Technological Projects of Guangdong Province (grant number
 1087 2018B030335001); J.B.C. was supported by National Science Foundation Graduate Research
 1088 Fellowship (Grant No. DGE-1839285); M. Parzuchowski, K. Rybus, and N.M.S. were
 1089 supported by Polish National Science Center & DFG Beethoven grant
 1090 (2016/23/G/HS6/01775); A.C.S. was supported by Portuguese National Foundation for
 1091 Science and Technology [Grant number SFRH/BD/126304/2016]; L. Boncinelli was
 1092 supported by PRIN 2017, grant n. 20178293XT, Italian Ministry of Education and Research;
 1093 M.F.F.R. was supported by PSA 006 BRA 008 Data Collection in Support of PSADM 001
 1094 Measurement Invariance Project; M. Misiak was supported by scholarship of the Foundation
 1095 for Polish Science (START) and by the scholarship of the National Science Centre
 1096 (2020/36/T/HS6/00256); P.B. was supported by Slovak Research and Development Agency
 1097 [project no. APVV-18-0140]; M.A. was supported by Slovak Research and Development
 1098 Agency [project no. APVV-17-0418] & project PRIMUS/20/HUM/009; A. Findor and M.H.
 1099 were supported by Slovak Research and Development Agency under contract no. APVV-17-
 1100 0596; T.G. was supported by Social Sciences and Humanities Research Council of Canada (#
 1101 950-224884); P.P. was supported by Swedish Research Council (2016-06793); Y.L. was
 1102 supported by The Project of Philosophy and Social Science Research in Colleges and
 1103 Universities in Jiangsu Province (grant number: 2020SJA0017); M. Kowal was supported by
 1104 the scholarship of the National Science Centre (2019/33/N/HS6/00054); P.A. was supported
 1105 by UID/PSI/03125/2019 from the Portuguese National Foundation for Science and
 1106 Technology.

1107 The funders had no role in study design, data collection and analysis, decision to publish or
 1108 preparation of the manuscript.

1109

1110 **Author Contribution**

1111 **Conceptualization:** B. Bago and B.A.

1112 **Data curation:** B. Bago, M. Kovacs, and T.N.

- 1113 **Formal analysis:** B. Bago, M. Kovacs, T.N., Z.K., and B.P.
- 1114 **Funding acquisition:** P.A., P.M., K.Q., I. Zettler, and R. Hoekstra
- 1115 **Investigation:** B. Bago, M. Kovacs, M.A., S. Adamus, S. Albaloooshi, N.A.-A., S. Alper,
 1116 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.
 1117 Bashour, E. Baskin, L. Batalha, C.B., J. Bavolar, F.B., M. Becker, B. Becker, A.B., M.
 1118 Białek, E. Bilancini, D.B., L. Boncinelli, J. Boudesseul, B.T.B., E.M.B., M.M.B., D.P.C.,
 1119 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.,
 1120 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.-
 1121 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.,
 1122 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.,
 1123 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,
 1124 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.,
 1125 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.,
 1126 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.,
 1127 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.,
 1128 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.,
 1129 T.L.M., A. Mirisola, M. Misiak, P.M., M.M.-J., A. Monajem, D.M., E.D.M., E.N., I.N.,
 1130 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,
 1131 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.,
 1132 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.,
 1133 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.,
 1134 K.A.Š., M. Sharifian, J. Shi, Y.S., E.S., M. Sirota, M. Slipenkyj, Ç.S., A. Sorokowska, P.S.,
 1135 S. Söylemez, N.K.S., I.D.S., A. Sternisko, L.S.-W., S.L.K.S., S. Stieger, D.S., J. Strube,
 1136 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.
 1137 Trémolière, L.V.T., B.N.T., M. Twardawski, M.A. Vadillo, Z.V., L.A.V., B.V., D.V., M.V.,
 1138 M.A. Vranka, S. Wang, S.-L.W., S. Whyte, L.S.W., A.W., X.W., F.X., S. Yadanar, H.Y.,
 1139 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
 1140 B.A.
- 1141 **Methodology:** B. Bago, Z.K., B.P., R. Hoekstra, and B.A.
- 1142 **Project administration:** B. Bago, M. Kovacs, J.P., and B.A.
- 1143 **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.,
 1144 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
 1145 B.A.
- 1146 **Supervision:** B. Bago, J.P., P.A., A.A.Ö., P.S., M.V., and B.A.
- 1147 **Validation:** B. Bago, M. Kovacs, T.N., Z.K., B.P., and B.A.
- 1148 **Visualization:** B. Bago, M. Kovacs, T.N., and B.P.
- 1149 **Writing - original draft:** B. Bago and B.A.
- 1150 **Writing - review & editing:** B. Bago, M. Kovacs, J.P., B.P., M.A., S. Adamus, S.
 1151 Albaloooshi, N.A.-A., S. Alper, S.A.-S., S.G.A., S. Amaya, P.K.A., G.A., D.A., P.A.,
 1152 J.J.B.R.A., A.A., P.B., K.B., B. Bashour, E. Baskin, L. Batalha, C.B., J. Bavolar, F.B., M.
 1153 Becker, B. Becker, A.B., M. Białek, E. Bilancini, D.B., L. Boncinelli, J. Boudesseul, B.T.B.,

1154 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.,
 1155 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.,
 1156 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.,
 1157 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,
 1158 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.,
 1159 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.,
 1160 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,
 1161 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.,
 1162 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.,
 1163 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,
 1164 J.P.M., N.M., M.C.M., L.M., T.L.M., A. Mirisola, M. Misiak, P.M., M.M.-J., A. Monajem,
 1165 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.,
 1166 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.,
 1167 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.
 1168 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,
 1169 A.C.S., N.S., C.S., K.S., K.A.Š., M. Sharifian, J. Shi, Y.S., E.S., M. Sirota, M. Slipenkyj,
 1170 Ç.S., A. Sorokowska, P.S., S. Söylemez, N.K.S., I.D.S., A. Sternisko, L.S.-W., S.L.K.S., S.
 1171 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.,
 1172 M. Tonković, B. Trémolière, L.V.T., B.N.T., M. Twardawski, M.A. Vadillo, Z.V., L.A.V.,
 1173 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.
 1174 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.,
 1175 J.Z., X.Z., and B.A.

1176 **Competing interests statement**

1177 The authors declare no competing interests.

1178

1179

1180 **Table 1**1181 *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 (Study 1a)	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				
Study 1a	381	622	566	1,569
Study 1b	327	553	546	1,426
Study 2a	323	690	2,971	3,984
Study 2b	277	576	2,660	3,513

1182 Note. Study 1b and Study 2b refers to the Speedboat dilemmas (recall, all of our subjects
 1183 answered to one trolley and a speedboat dilemmas)

1184

1185

1186

1187

1188 **Table 2**1189 *The effect of personal force on moral dilemma judgements*

1190

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 ⁻³ , 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 ⁻³ , 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 ³	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 ⁻⁵ , 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 ⁹	<1.00*10 ⁻³ , 5.50*10 ⁹	-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]

1191 Note. BF = Bayes Factor, RR = Robustness Region of the prior

1192

1193

1194

1195

1196 **Table 3**
 1197 *Interaction between personal force and intention on moral judgments*
 1198

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}	$<1.00 \times 10^{-5}$, 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

1199 Note. BF = Bayes Factor, RR = Robustness Region of the prior

1200

1201 **Table 4**
 1202 *Individualism/collectivism associations with the interaction between personal force and*
 1203 *intention on moral judgments (Trolley dilemmas)*
 1204

Dilemma	Variable	With familiarity exclusion				No familiarity exclusion			
		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

1205

1206 **Table 5**
 1207 *The effect of intention on moral dilemma judgements (Trolley dilemmas)*
 1208

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]

1209

1210

1211

1212

1213

1214 **Table 6**
 1215 *The effect of intention on moral dilemma judgements (Speedboat dilemmas)*
 1216

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]

1217
 1218
 1219
 1220
 1221
 1222

1223 **Figure 1**

1224 *Results on Trolley and Speedboat dilemmas in Study 1 (effect of personal force) with all*
 1225 *exclusion criteria applied (A,C) and including familiar participants (B,D). Error bars are*
 1226 *95% Confidence Intervals around the mean. Scale ranged from 1 (completely unacceptable)*
 1227 *to 9 (completely acceptable). In the trolley problem n = 1,569 when all exclusion criteria is*
 1228 *applied, and n = 3,524 when the familiarity exclusion not applied. In the speedboat dilemma,*
 1229 *n = 1,426 when all exclusion criteria is applied, and n = 3,295 when the familiarity exclusion*
 1230 *not applied.*

1231
 1232
 1233

1234 **Figure 2**

1235 *Results on Trolley and Speedboat dilemmas in Study 2 (personal force and intention*
 1236 *interaction) with all exclusion criteria applied (A,C) and including familiar participants*
 1237 *(B,D). Error bars represent 95% Confidence Intervals. Scale ranged from 1 (completely*
 1238 *unacceptable) to 9 (completely acceptable). In the trolley problem n = 3,984 when all*
 1239 *exclusion criteria is applied, and n = 9,844 when the familiarity exclusion not applied. In the*
 1240 *speedboat dilemma, n = 3,513 when all exclusion criteria is applied, and n = 9,006 when the*
 1241 *familiarity exclusion not applied.*

1242
 1243
 1244

1245 **Figure 3**

1246 *Correlation between country-level collectivism and the Eta squared effect size of the*
 1247 *interaction between personal force and intention with all exclusion criteria applied (left*
 1248 *panel) and including participants familiar with the trolley problem (right panel) . The size of*
 1249 *the circles indicate the size of the sample in a given country. Blue line is the weighted*
 1250 *regression line.*

1251
 1252

1253 **Figure 4**

1254 *Correlation between country-level collectivism and the Eta squared effect size of the*
 1255 *interaction between personal force and intention with all exclusion criteria applied (left) and*

1254 *including familiar participants (right) on the speedboat problem. The size of the circles*
1255 *indicate the size of the sample in a given country. Blue line is the weighted regression line.*
1256
1257
1258

1259 **References**

1260

- 1261 1. Mill, J. S. & Bentham, J. *Utilitarianism and other essays*. (Penguin, 1987).
- 1262 2. Kant, I. *Groundwork for the metaphysics of morals*. (Yale University Press, 1785).
- 1263 3. Greene, J. D. *Moral tribes: emotion, reason and the gap between us and them*. (Penguin
1264 Press, 2013).
- 1265 4. London, J. A. How should we model rare disease allocation decisions? *Hastings Cent.*
1266 *Rep.* **42**, 3 (2014).
- 1267 5. Bonnefon, J.-F., Shariff, A. & Rahwan, I. The social dilemma of autonomous vehicles.
1268 *Science* **352**, 1573–1576 (2016).
- 1269 6. Foot, P. The problem of abortion and the doctrine of double effect. *Oxf. Rev.* **5**, 5–15
1270 (1967).
- 1271 7. Baron, J. Utilitarian vs. deontological reasoning: method, results, and theory. in *Moral*
1272 *inferences* (eds. Bonnefon, J.-F. & Trémolière, B.) 137–151 (Psychology Press, 2017).
- 1273 8. Baron, J. & Gürçay, B. A meta-analysis of response-time tests of the sequential two-
1274 systems model of moral judgment. *Mem. Cognit.* **45**, 566–575 (2017).
- 1275 9. Cushman, F., Young, L. & Hauser, M. The role of conscious reasoning and intuition in
1276 moral judgment: Testing three principles of harm. *Psychol. Sci.* **17**, 1082–1089 (2006).
- 1277 10. Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M. & Cohen, J. D. The neural
1278 bases of cognitive conflict and control in moral judgment. *Neuron* **44**, 389–400 (2004).
- 1279 11. Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M. & Cohen, J. D. An fMRI
1280 investigation of emotional engagement in moral judgment. *Science* **293**, 2105–2108
1281 (2001).
- 1282 12. Gürçay, B. & Baron, J. Challenges for the sequential two-system model of moral
1283 judgement. *Think. Reason.* **23**, 49–80 (2017).
- 1284 13. Mikhail, J. Universal moral grammar: Theory, evidence and the future. *Trends Cogn. Sci.*
1285 **11**, 143–152 (2007).
- 1286 14. Boyle, J. Medical ethics and double effect: the case of terminal sedation. *Theor. Med.*

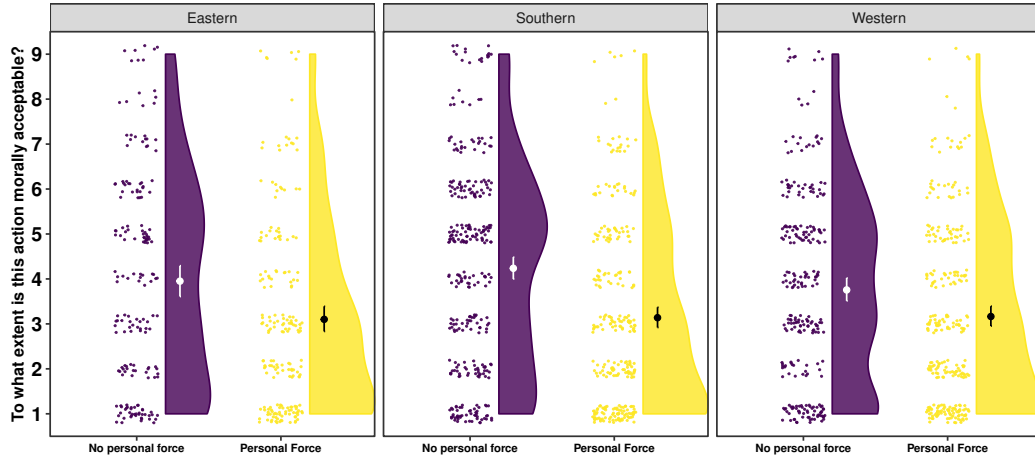
- 1287 *Bioeth.* **25**, 51–60 (2004).
- 1288 15. Gross, M. L. Bioethics and armed conflict: Mapping the moral dimensions of medicine
1289 and war. *Hastings Cent. Rep.* **34**, 22–30 (2004).
- 1290 16. Gross, M. L. Killing civilians intentionally: double effect, reprisal, and necessity in the
1291 Middle East. *Polit. Sci. Q.* **120**, 555–579 (2005).
- 1292 17. Tully, P. A. The doctrine of double effect and the question of constraints on business
1293 decisions. *J. Bus. Ethics* **58**, 51–63 (2005).
- 1294 18. Greene, J. D. *et al.* Pushing moral buttons: The interaction between personal force and
1295 intention in moral judgment. *Cognition* **111**, 364–371 (2009).
- 1296 19. Cushman, F. Action, outcome, and value: A dual-system framework for morality.
1297 *Personal. Soc. Psychol. Rev.* **17**, 273–292 (2013).
- 1298 20. Cushman, F., Gray, K., Gaffey, A. & Mendes, W. B. Simulating murder: The aversion to
1299 harmful action. *Emotion* **12**, 2 (2012).
- 1300 21. Ellsworth, R. M. & Walker, R. S. Sociobiology of lethal violence in small-scale societies.
1301 in *The Routledge International Handbook of Biosocial Criminology* 85–102 (Routledge,
1302 2014).
- 1303 22. Abarbanell, L. & Hauser, M. D. Mayan morality: An exploration of permissible harms.
1304 *Cognition* **115**, 207–224 (2010).
- 1305 23. Barrett, H. C. *et al.* Small-scale societies exhibit fundamental variation in the role of
1306 intentions in moral judgment. *Proc. Natl. Acad. Sci.* **113**, 4688–4693 (2016).
- 1307 24. Cushman, F., Young, L. & Greene, J. D. Our multi-system moral psychology: Towards a
1308 consensus view. in *The Oxford handbook of moral psychology* 47–71 (2010).
- 1309 25. Koenigs, M. *et al.* Damage to the prefrontal cortex increases utilitarian moral
1310 judgements. *Nature* **446**, 908–911 (2007).
- 1311 26. Perkins, A. M. *et al.* A dose of ruthlessness: Interpersonal moral judgment is hardened
1312 by the anti-anxiety drug lorazepam. *J. Exp. Psychol. Gen.* **142**, 612–620 (2013).
- 1313 27. Szekely, R. D. & Miu, A. C. Incidental emotions in moral dilemmas: The influence of
1314 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

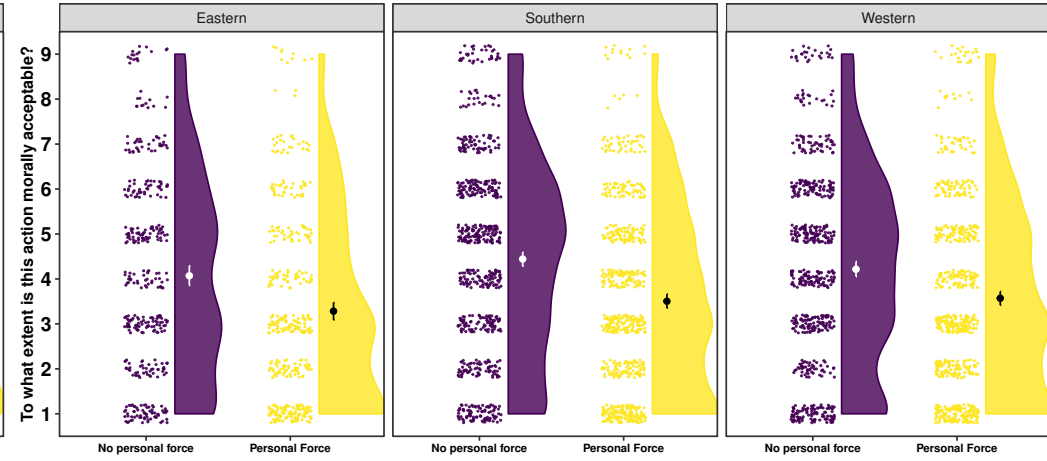
- 1343 and the judgments of philosophers. *Cognition* **179**, 241–265 (2018).
- 1344 41. Hauser, M. *Moral minds: How nature designed our universal sense of right and wrong*.
1345 (Ecco/HarperCollins Publishers, 2006).
- 1346 42. Hauser, M. D., Young, L. & Cushman, F. Reviving Rawls' linguistic analogy. *Moral*
1347 *Psychol.* **2**, 107–143 (2008).
- 1348 43. Gelfand, M. J., Nishii, L. H. & Raver, J. L. On the nature and importance of cultural
1349 tightness-looseness. *J. Appl. Psychol.* **91**, 1225 (2006).
- 1350 44. Awad, E., Dsouza, S., Shariff, A., Rahwan, I. & Bonnefon, J.-F. Universals and variations
1351 in moral decisions made in 42 countries by 70,000 participants. *Proc. Natl. Acad. Sci.*
1352 **117**, 2332–2337 (2020).
- 1353 45. Gawronski, B., Armstrong, J., Conway, P., Friesdorf, R. & Hütter, M. Consequences,
1354 norms, and generalized inaction in moral dilemmas: The CNI model of moral decision-
1355 making. *J. Pers. Soc. Psychol.* **113**, 343 (2017).
- 1356 46. Conway, P. & Gawronski, B. Deontological and utilitarian inclinations in moral decision
1357 making: a process dissociation approach. *J. Pers. Soc. Psychol.* **104**, 216 (2013).
- 1358 47. Moshontz, H. *et al.* The Psychological Science Accelerator: Advancing psychology
1359 through a distributed collaborative network. *Adv. Methods Pract. Psychol. Sci.* **1**, 501–
1360 515 (2018).
- 1361 48. Bond, M. H. & van de Vijver, F. J. R. Making scientific sense of cultural differences in
1362 psychological outcomes: Unpackaging the Magnum Mysterium. in *Culture and*
1363 *psychology. Cross-cultural research methods in psychology* (eds. Matsumoto, D. & van
1364 de Vijver, F. J. R.) 75–100 (Cambridge University Press, 2011).
- 1365 49. Costa, A. *et al.* Your morals depend on language. *PloS One* **9**, e94842 (2014).
- 1366 50. Meade, A. W. & Craig, S. B. Identifying careless responses in survey data. *Psychol.*
1367 *Methods* **17**, 437–455 (2012).
- 1368 51. Hannikainen, I. R., Machery, E. & Cushman, F. A. Is utilitarian sacrifice becoming more
1369 morally permissible? *Cognition* **170**, 95–101 (2018).
- 1370 52. Dienes, Z. Using Bayes to get the most out of non-significant results. *Front. Psychol.* **5**,

- 1371 781 (2014).
- 1372 53. Schönbrodt, F. D. & Wagenmakers, E.-J. Bayes factor design analysis: Planning for
1373 compelling evidence. *Psychon. Bull. Rev.* **25**, 128–142 (2018).
- 1374 54. Morey, R. D., Rouder, J. N. & Jamil, T. *BayesFactor: Computation of Bayes factors for*
1375 *common designs*. (2015). [Computer Software]
- 1376 55. R Core Team. *R: A language and environment for statistical computing*. R Foundation
1377 *for Statistical Computing, Vienna, Austria*. (2018). [Computer Software]
- 1378 56. Dienes, Z. How do I know what my theory predicts? Accessed via:
1379 <https://psyarxiv.com/yqaj4>. *Adv. Methods Pract. Psychol. Sci.* **2**, 364–377 (2019).
- 1380 57. Muthukrishna, M. *et al.* Beyond WEIRD psychology: measuring and mapping scales of
1381 cultural and psychological distance. *Psychol. Sci.* **31**. 678-701 (2018).
- 1382

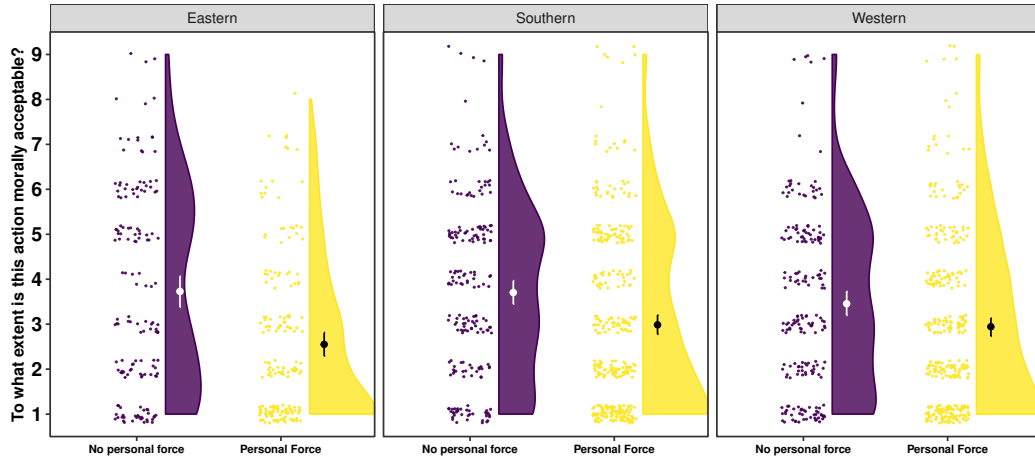
A Trolley problem – all exclusion criteria applied



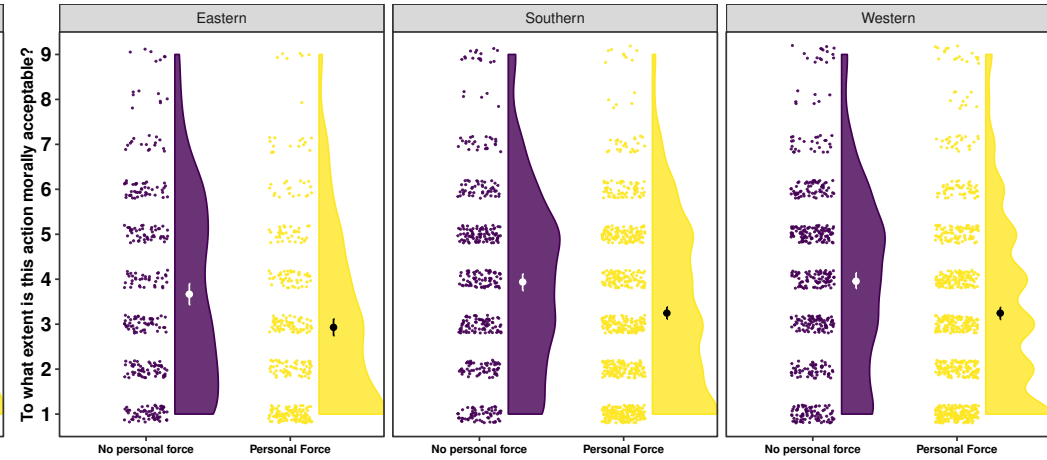
B Trolley problem – familiarity exclusion not applied



C Speedboat problem – all exclusion criteria applied

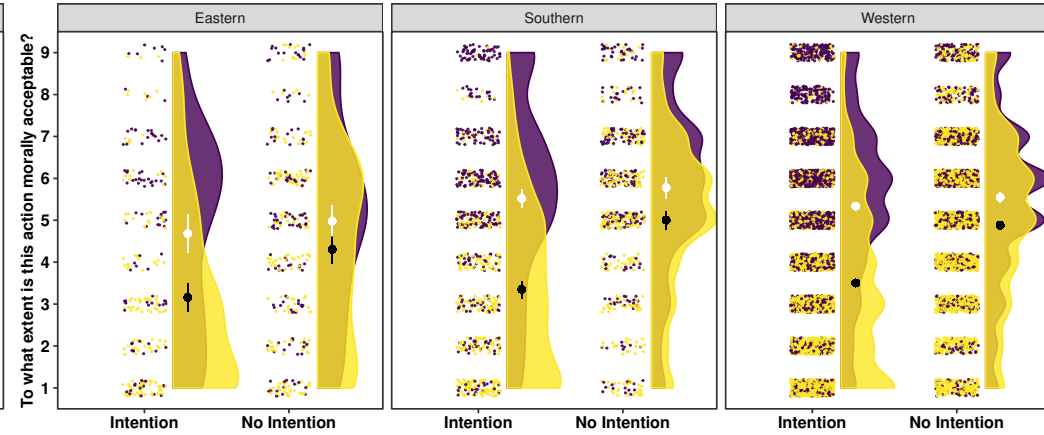
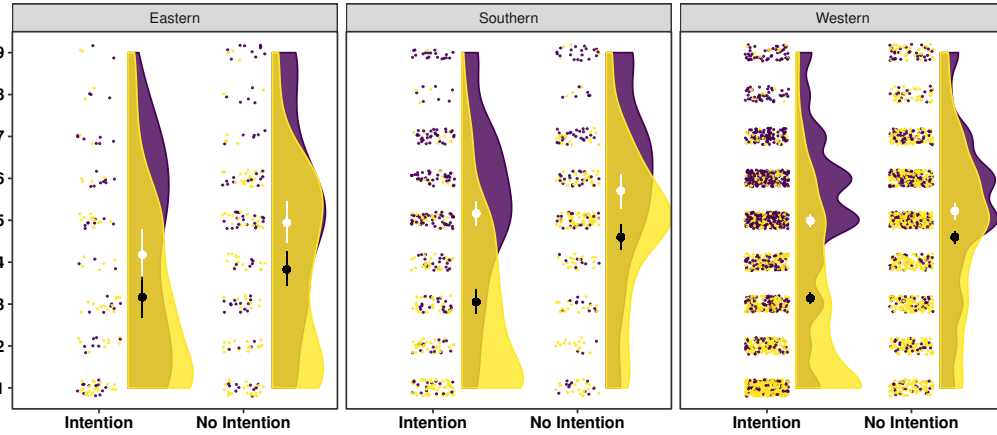


D Speedboat problem – familiarity exclusion not applied



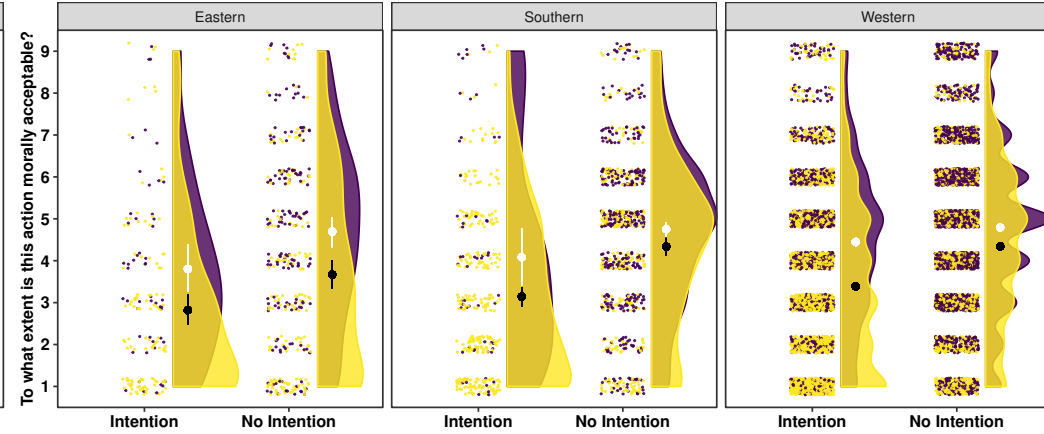
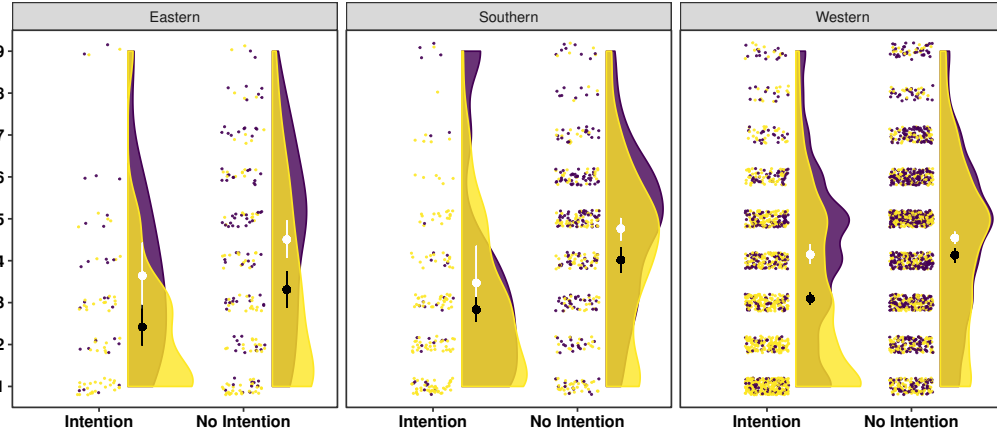
A Trolley problem – all exclusion criteria applied

B Trolley problem – familiarity exclusion not applied



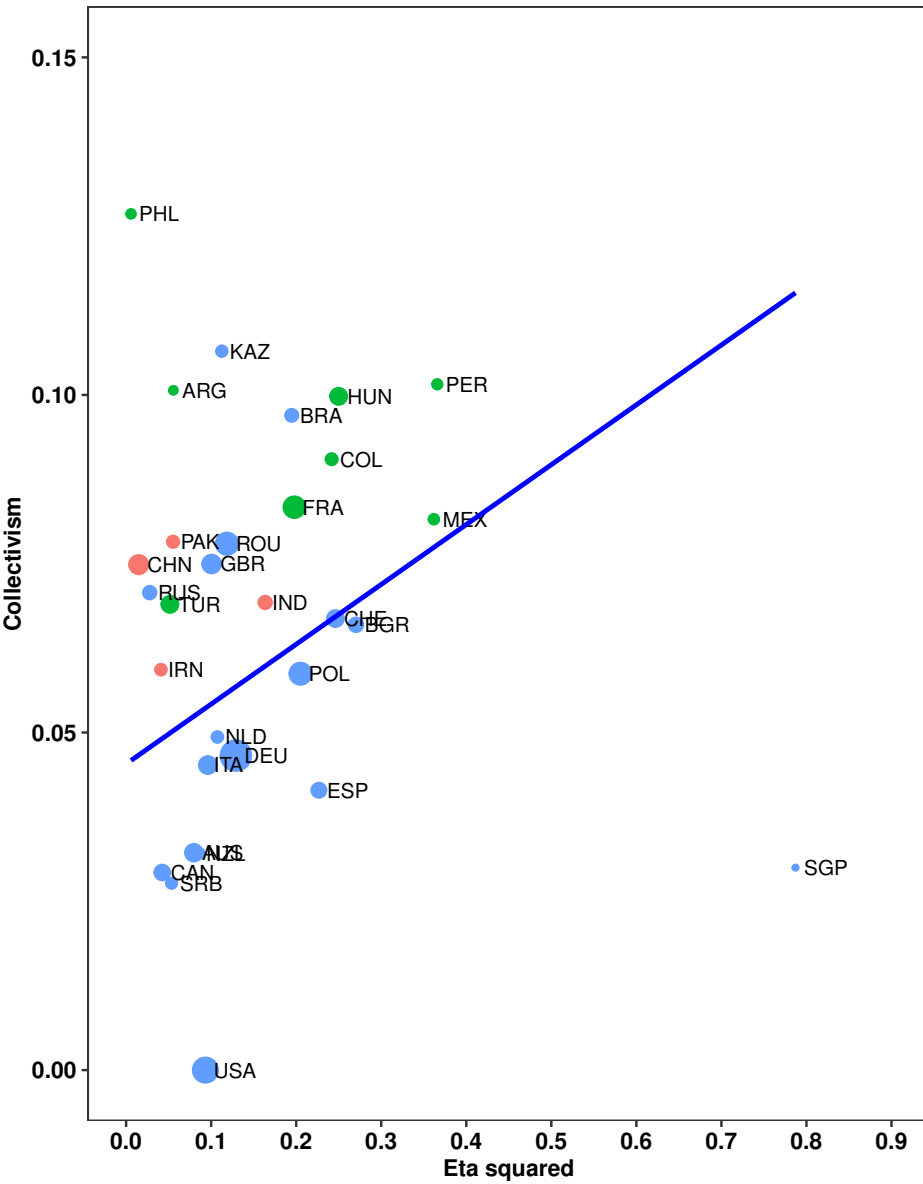
C Speedboat problem – all exclusion criteria applied

D Speedboat problem – familiarity exclusion not applied

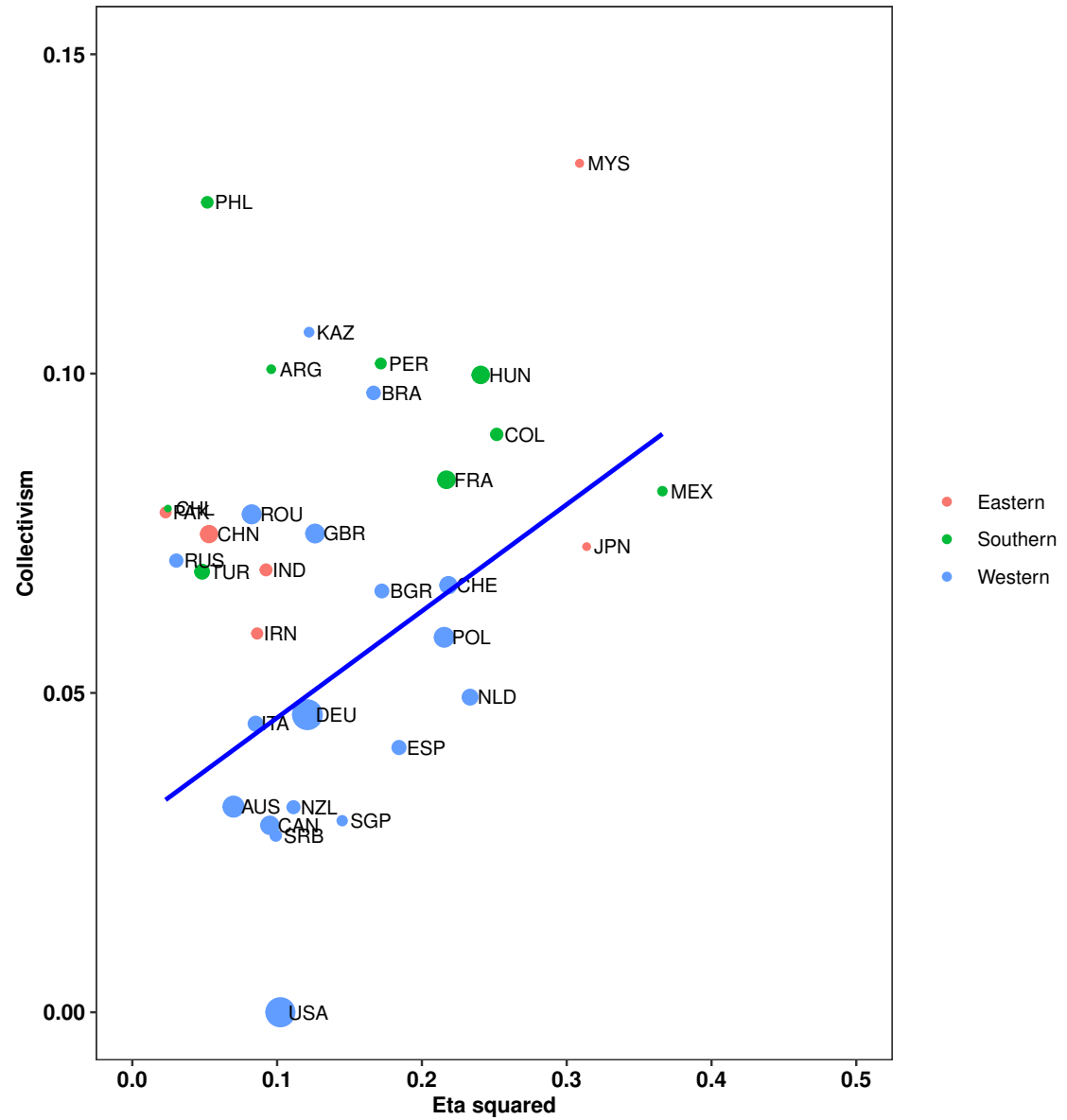


◆ No personal force
 ◆ Personal Force

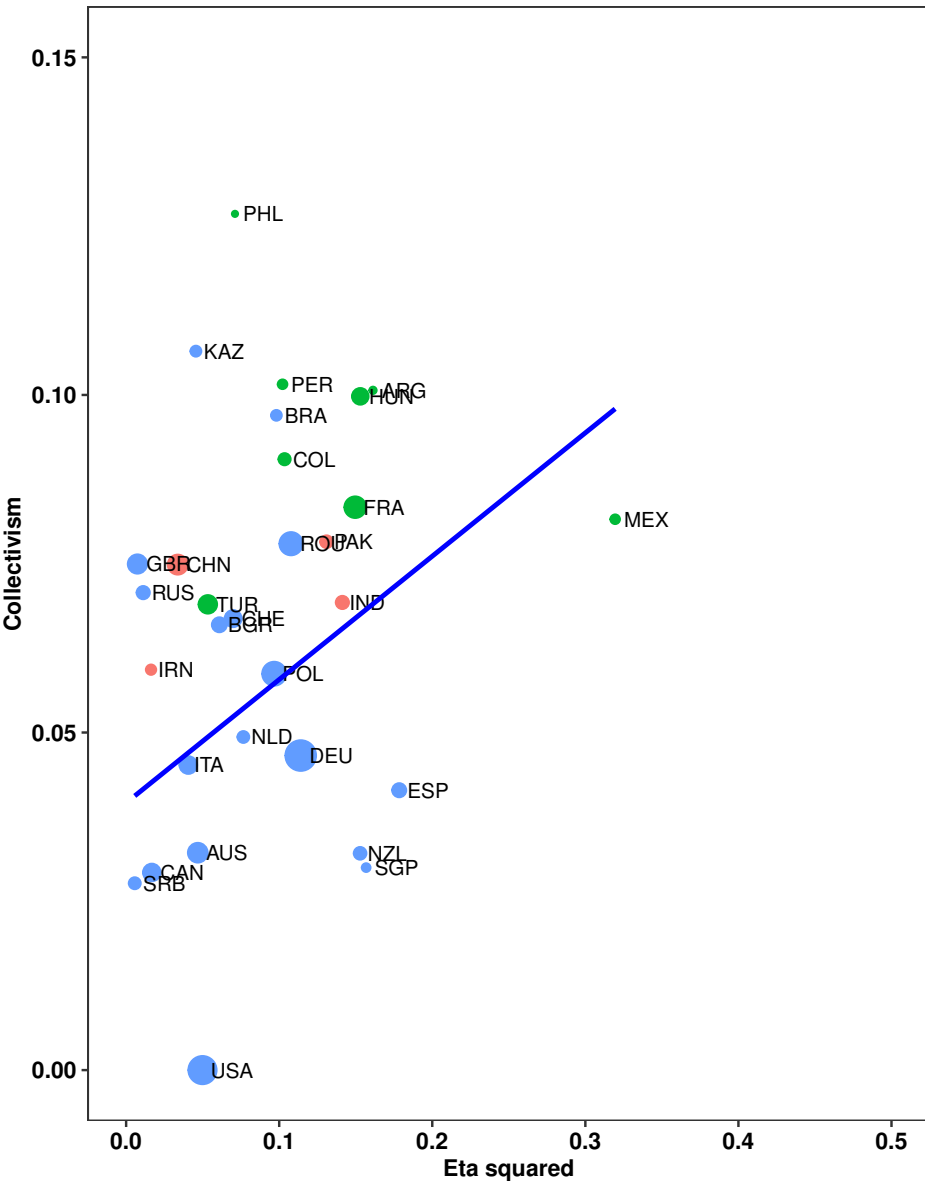
A Trolley problem (all exclusions applied)



B Trolley problem (familiarity exclusion not applied)



A Speedboat problem (all exclusions applied)



B Speedboat problem (familiarity exclusion not applied)

