To Which World Regions Does the Valence-Dominance Model of Social Perception Apply?

(Registered Report Stage 1)

Corresponding author: Benedict Jones (ben.jones@glasgow.ac.uk), Institute of Neuroscience & Psychology, University of Glasgow, Scotland, UK.

Benedict C Jones[1], Lisa M DeBruine[1], Jessica K Flake[2], Balazs Aczel[3], Matus Adamkovic[4], Ravin Alaei[5], Sinan Alper[6], Sara Álvarez Solas[7], Michael R Andreychik[8], Daniel Ansari[9], Jack D Arnal[10], Peter Babinčák[4], Benjamin Balas[11], Gabriel Baník[4], Krystian Barzykowski[12], Ernest Baskin[13], Carlota Batres[14], Jennifer L Beaudry[15], Khandis R Blake[16], Nicholas Bloxsom[17], Martha L Borras Guevara[18], Mark J Brandt[19], Debora I Burin[20], Dustin P Calvillo[21], Lilian Carvalho[22], Priyanka Chandel[23], Armand Chatard[24], Sau-Chin Chen[25], Coralie Chevallier[26], William J Chopik[27], Cody D Christopherson[28], Vinet Coetzee[29], Nicholas A Coles[30], Melissa F Colloff[31], Corey L Cook[32], Matthew T Crawford[33], Alexander F Danvers[34], Tiago JS de Lima[35], Barnaby JW Dixson[36], Tiantian Dong[37], Vilius Dranseika[38], Yarrow Dunham[39], Thomas R Evans[40], AnaMaria Fernandez[41], Paulo RS Ferreira[42], Heather D Flowe[43], Patrick S Forscher[44], Gwendolyn Gardiner[45], Eva Gilboa-Schechtman[46], Michael Gilead[47], Tripat Gill[48], Taylor D Gogan[15], Isaac González-Santoyo[49], Amanda C Hahn[50], Nándor Hajdú[51],Javad Hatami[52], Eric Hehman[2], Chuan-Peng Hu[53], Hans IJzerman[54], Michael Inzlicht[5], Natalia Irrazabal[55], Bastian Jaeger[19], Chaning Jang[56,57], Steve MJ Janssen[58], Zhongqing Jiang[59], Julia Jünger[60], Pavol Kačmár[61], Gwenaël Kaminski[62], Aycan Kapucu[63], Aslan Karaaslan[63], Monica A Koehn[64], Vanja Kovic[65], Pratibha Kujur[23], Chun-Chia Kung[66], Claus Lamm[67], Ai-Suan Lee[68], Nicole Legate[69], JuanDavid Leongómez[70], Carmel A Levitan[71], Hause Lin[5], Samuel Lins[72], Qinglan Liu[73], MarcoTullio Liuzza[74], Johannes Lutz[75], Harry Manley[76], Tara Marshall[77], Georgina W Mburu[56], Randy J McCarthy[78], Nicholas M Michalak[79], Vásquez-Amézquita Milena[70], Jeremy K Miller[80], JoseAntonio Muñoz Reyes[81], Erica D Musser[82], Lison Neyroud[83], Tonje K Nielsen[84], Erin M O'Mara[85], Ceylan Okan[86], Julian A Oldmeadow[15], Jerome Olsen[67], Asil A Özdoğru[87], Babita Pande[23], Marietta Papadatou-Pastou[88], Arti Parganiha[23], Noorshama Parveen[23], Gerit Pfuhl[89], Michael C Philipp[90], Isabel R Pinto[72], Pablo Polo[91], Sraddha Pradhan[23], John Protzko[92], Yue Qi[93], Dongning Ren[19], Ivan Ropovik[4], Nicholas O Rule[5], Waldir M Sampaio[94], Oscar R Sánchez[70], Diana R Santos[95], Elisa Sarda[54], S. Adil Saribay[96], Blair Saunders[97], Vidar Schei[98], Christoph Schild[99], Irina Schmid[100], Kathleen Schmidt[101], Martin Seehuus[102], Mohammad H Sharifian[103], Victor KM Shiramizu[104], Almog Simchon[47], Margaret M Singh[23], Miroslav Sirota[105], Guyan Sloane[106], Ian D Stephen[107], Stefan Stieger[100], Daniel Storage[108], Juncai Sun[109], Therese E Sverdrup[98], Peter Szecsi[51,] Christian K Tamnes[84], Chrystalle BY Tan[110], KokWei Tan[111], Martin Thirkettle[112], Enrique Turiegano[113], Kim Uittenhove[114], Heather L Urry[115], Miguel A Vadillo[116], Eugenio Valderrama[70], Jaroslava V Valentova[117], Nicolas Van der Linden[118], Wolf Vanpaemel[119], Marco AC Varella[117], Milena Vásquez-Amézquita[70], LeighAnn Vaughn[120], Evie Vergauwe[121], Michelangelo Vianello[122], Martin Voracek[67], David White[123], Megan L Willis[124], John Paul Wilson[125], Anna Wlodarczyk[126], Qi Wu[59], Sally Y Xie[2], Wen-Jing Yan[127], Xin Yang[39], Ilya Zakharov[128], Ingo Zettler[129], Janis H Zickfeld[84], Christopher R Chartier[17]

[1]Institute of Neuroscience & Psychology, University of Glasgow, UK, [2]Psychology, McGill University, Canada, [3]Institute of Psychology, ELTE, Eotvos Lorand University, Hungary, [4]Institute of Psychology, Faculty of Arts, University of Presov, Slovakia, [5]Psychology, University of Toronto, Canada, [6]Department of Psychology, Yasar University, Turkey, [7]Ecosystems, Universidad Regional Amazónica Ikiam, Ecuador, [8]Psychology, Fairfield University, USA, [9]Psychology, The University of Western Ontario, Canada, [10]Psychology Department, McDaniel College, USA, [11]Psychology, North Dakota State University, USA, [12]Institute of Psychology, Applied Memory Research Laboratory, Jagiellonian University, Poland, [13]Food Marketing, Saint Joseph's University, USA, [14]Department of Psychology, Franklin and Marshall College, USA, [15]Department of Psychological Sciences, Swinburne University of Technology, Australia, [16]Evolution & Ecology Research Centre,

School of Biological, Earth and Environmental Sciences, UNSW Sydney, Australia, [17]Department of Psychology, Ashland University, USA, [18]School of Psychology and Neuroscience, University of St Andrews, UK, [19]Department of Social Psychology, Tilburg University, Netherlands, [20]Instituto de Investigaciones, Facultad de Psicologia, Universidad de Buenos Aires - CONICET, Argentina, [21]Psychology, California State University San Marcos, USA, [22]Marketing, FGV/EAESP, Brazil, [23]School of Studies in Life Science, Raipur, Pt. Ravishankar Shukla University, India, [24]Psychologie, Université de Poitiers et Centre National de la Recherche Scientifique, France, [25]Department of Human Development and Psychology, Tzu-Chi University, Taiwan, [26]Departement d'Etudes Cognitives, Ecole Normale Supérieure, France, [27]Psychology, Michigan State University, USA, [28]Clinical Mental Health Counseling, Southern Oregon University, USA, [29]Psychology, University of Pretoria, South Africa, [30]Psychology, University of Tennessee-Knoxville, USA, [31]Psychology, University of Birmingham, UK, [32]Department of Psychology, Pacific Lutheran University, USA, [33]School of Psychology, Victoria University of Wellington, New Zealand, [34]Department of Psychology, University of Arizona, USA, [35]Psychology, Universidade de Fortaleza, Brazil, [36]School of Psychology, The University of Queensland, Australia, [37]Student, Qufu Normal University, China, [38]Institute of Philosophy, Vilnius University, Lithuania, [39]Psychology, Yale University, USA, [40]School of Psychological, Social and Behavioural Sciences, Coventry University, UK, [41]Laboratorio de Evolución y Relaciones Interpersonales, Universidad de Santiago de Chile, Chile, [42]Faculdade de Ciências Humanas, Universidade Federal da Grande Dourados, Brazil, [43]School of Psychology, University of Birmingham, UK,

[44]Department of Psychological Science, University of Arkansas, USA, [45]Psychology, University of California, Riverside, USA, [46]Psychology, Bar Ilan University, Israel, [47]Psychology, Ben-Gurion University, Israel, [48]Lazaridis School of Business & Economics, Wilfrid Laurier University, Canada, [49]Psychology, National Autonomous University of Mexico, México, [50]Department of Psychology, Humboldt State Universtiy, USA, [51]Department of Affective Psychology, Eötvös Loránd University, Hungary, [52]Faculty of Psychology and Educational Science, University of Tehran, Iran, [53]Deutsches Resilienz Zentrum (DRZ), Johannes Gutenberg University Medical Center, Germany, [54]LIP/PC2S, Université Grenoble Alpes, France, [55]Facultad de Ciencias Sociales, Universidad de Palermo - CONICET, Argentina, [56]Research, Busara Center for Behavioral Economics, Kenya, [57]The Busara Center for Behavioral Economics, Tanzania, [58]School of Psychology, University of Nottingham - Malaysia Campus, Malaysia, [59]Psychology, Liaoning Normal University, China, [60]Biological Personality Psychology, University of Göttingen, Germany, [61]Department of Psychology, Faculty of Arts, University of Pavol Jozef Šafárik in Košice, Slovakia, [62]CLLE (UMR 5263), Toulouse University, France, [63]Psychology, Ege University, Turkey, [64]School of Social Sciences and Psychology, Western Sydney University, Australia, [65]Laboratory for Neurocognition and Applied Cognition, Department of Psychology, University of Belgrade, Faculty of Philosophy, Serbia, [66]Psychology, National Cheng Kung University, Taiwan, [67]Department of Basic Psychological Research and Research Methods, Faculty of Psychology, University of Vienna, Austria, [68]Department of Psychology and Counselling, Universiti Tunku Abdul Rahman, Malaysia, [69]Psychology, Illinois Institute of

Technology, USA, [70]Faculty of Psychology, Universidad El Bosque, Colombia, [71]Cognitive Science, Occidental College, United States, [72]Departament of Psychology, University of Porto, Portugal, [73]Department of Psychology, Hubei University, China, [74]Surgical and Medical Sciences, "Magna Graecia" University of Catanzaro, Italy, [75]Department of Psychology, University of Potsdam, Germany, Deutschland, [76]Faculty of Psychology, Chulalongkorn University, Thailand, [77]Health, Aging & Society, McMaster University, Canada, [78]Department of Psychology, Northern Illinois University, USA, [79]Psychology, University of Michigan, USA, [80]Department of Psychology, Willamette University, USA, [81]Laboratorio de Comportamiento Animal y Humano, Universidad de Playa Ancha, Chile, [82]Psychology, Florida International University, USA, [83]Psychology, Universite Pierre Mendes, France, [84]Department of Psychology, University of Oslo, Norway, [85]Psychology, University of Dayton, USA, [86]School of Social Sciences and Psychology, Western Sydney University, Australia, [87]Psychology, Üsküdar University, Turkey, [88]School of Education, Department of Primary Education, National and Kapodistrian University of Athens, Greece, [89]Department of Psychology, UiT The Arctic University of Norway, Norway, [90]School of Psychology, Massey University, New Zealand, [91]Centro de Estudios Avanzados, Universidad de Playa Ancha, Chile. [92]Psychological & Brain Sciences, University of California, Santa Barbara, USA, [93]CAS Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, China, [94]Department of Psychology, Universidade Federal de São Carlos, Brazil, [95]Faculdade de Ciências Humanas, Universidade Federal da Grande Dourados, Brazil, [96]Psychology, Boğaziçi University, Turkey, [97]Psychology, School of Social Sciences, University of Dundee, UK, [98]Department of Strategy and Management, NHH Norwegian School of Economics, Norway, [99]Department of Psychology, University of Copenhagen, Denmark, [100]Department of Psychology and Psychodynamics, Karl Landsteiner University of Health Sciences, Austria, [101]Department of Psychology, Southern Illinois University Carbondale, USA, [102]Psychology, Middlebury College, USA, [103]Department of Psychology, University of Tehran, Iran, [104]Brain Institute, Universidade Federal do Rio Grande do Norte, Brazil, [105]Psychology, University of Essex, UK, [106]Psychology, University of Exeter, UK, [107]Department of Psychology, Macquarie University, Australia, [108]Psychology Department, University of Denver, USA, [109]Professor, Qufu Normal University, China, [110]Department of Community and Family Medicine, Universiti Malaysia Sabah, Malaysia, [111]School of Psychology and Clinical Language Sciences, University of Reading Malaysia, Malaysia, [112]Centre for Behavioural Science and Applied Psychology, Sheffield Hallam University, UK, [113]Psychology, Universidad Autnoma de Madrid, Spain, [114]Department of Developmental Cognitive Psychology, University of Geneva, Switzerland, [115]Psychology, Tufts University, USA, [116]Departamento de Psicología Básica, Universidad Autónoma de Madrid, Spain, [117]Department of Experimental Psychology, Institute of Psychology, University of Sao Paulo, Brazil, [118]Center for Social and Cultural Psychology, Université Libre de Bruxelles, Belgium, [119]Faculty of Psychology and Educational Sciences, University of Leuven, Belgium, [120]Psychology, Ithaca College, USA, [121]Faculty of Psychology and Educational Sciences, University of Geneva, Switzerland, [122]Department of Philosophy, Sociology, Education and Applied Psychology, University of

Padova, Italy, [123]School of Psychology, UNSW Sydney, Australia, [124]School of Behavioural and Health Sciences, Australian Catholic University, Australia, [125]Psychology, Montclair State University, USA, [126]Escuela de Psicología, Universidad Católica del Norte, Chile, [127]Applied Psychology, Wenzhou University, China, [128]Developmental Behavioral Genetics Lab, Psychological Institute of Russian Academy of Education, Russia, [129]Department of Psychology, University of Copenhagen, Denmark.

Abstract

Over the last ten years, Oosterhof and Todorov's valence-dominance model has emerged as the most prominent account of how people evaluate faces on social dimensions. In this model, two dimensions (valence and dominance) underpin social judgments of faces. To which world regions this model applies is a critical, yet unanswered, question. We will address this question by replicating Oosterhof and Todorov's methodology across multiple world regions.

To Which World Regions Does the Valence-Dominance Model of Social Perception Apply?

People quickly and involuntarily form impressions of others based on their facial appearance[1-3]. These impressions then influence important social outcomes[4,5]. For example, people are more likely to cooperate in socioeconomic interactions with individuals whose faces are evaluated as more trustworthy[6], vote for individuals whose faces are evaluated as more competent[7], and seek romantic relationships with individuals whose faces are evaluated as more attractive[8]. Facial appearance can even influence life-or-death outcomes. For example, untrustworthy-looking defendants are more likely to receive death sentences[9]. Given that such evaluations influence profound outcomes, understanding how people evaluate others' faces can provide insight into a potentially important route through which social stereotypes impact behavior[10,11].

Over the last decade, Oosterhof and Todorov's valence-dominance model[12] has emerged as the most prominent account of how we evaluate faces on social dimensions[5]. Oosterhof and Todorov identified 13 different traits (aggressiveness, attractiveness, caringness, confidence, dominance, emotional stability, unhappiness, intelligence, meanness, responsibility, sociability, trustworthiness, and weirdness) that perceivers spontaneously use to evaluate faces when forming trait impressions[12]. From these traits, they derived a two-dimensional model of perception: *valence* and *dominance*. *Valence,* best characterized by rated trustworthiness, was defined as the extent to which the target was perceived as having the *intention* to harm the viewer[12]. *Dominance,* best characterized by rated dominance, was defined as the extent to which the target was perceived as having the *ability* to inflict harm on the viewer[12]. Crucially, the model proposes that these two dimensions are sufficient to drive social evaluations of faces. As a consequence, the majority of research on the effects of social evaluations of faces has focused on one or both of these dimensions[4,5].

Successful replications of the valence-dominance model have only been conducted in Western samples[13,14]. This focus on the West is consistent with research on human behavior more broadly, which typically draws general assumptions from analyses of Western participants' responses[15]. Kline et al. recently termed this problematic practice the *Western centrality assumption* and argued that regional variation, rather than universality, is likely the default for human behavior[16].

Consistent with Kline's notion that human behavior is best characterized by regional variation, two recent studies of social evaluation of faces by Chinese participants indicate different factors underlie their impressions[17,18]. Both studies reported that Chinese participants' social evaluations of faces were underpinned by a valence dimension similar to that reported by Oosterhof and Todorov for Western participants, but not by a corresponding dominance dimension. Instead, both studies reported a second dimension, referred to as *capability*, which was best characterized by rated intelligence. Furthermore, the ethnicity of the faces rated only subtly affected perceptions[17]. Research into potential cultural differences in the effects of experimentally manipulated facial characteristics on social perceptions has also found little evidence that cultural differences in social perceptions of faces depend on the ethnicity of the faces presented[19-21]. Collectively, these results suggest that the Western centrality assumption may be an important barrier to understanding how people evaluate faces on social dimensions. Crucially, these studies also suggest that the valence-dominance model is not necessarily a universal account of social evaluations of faces and warrants further investigation in the broadest set of samples possible.

Although the studies described above demonstrate that the valence-dominance model is not perfectly universal, to which specific world regions it does and does not apply are open and important questions. Demonstrating differences between British and Chinese raters is evidence against the universality of the valence-dominance model, but it does not adequately address these questions. Social perception in China may be unique in not fitting the valence-dominance model because of the atypically high general importance placed on status-related traits, such as capability, during social interactions in China[22,23]. Indeed, Tan et al. demonstrated face-processing differences between Chinese participants living in mainland China and Chinese participants living in nearby countries, such as Malaysia[24]. Insights regarding the unique formation of social perceptions in other cultures and world regions are lacking. Only a large-scale study investigating social perceptions in many different world regions can provide such insights.

To establish the world regions to which the valence-dominance model applies, we will replicate Oosterhof and Todorov's methodology[12] in a wide range of world regions (Africa, Asia, Australia and New Zealand, Central America and Mexico, Eastern Europe, the Middle East, the USA and Canada, Scandinavia, South America, the UK, and Western Europe; see Table 1). Our study will be the most comprehensive test of social evaluations of faces to date, including more than 9,000 participants. Participating research groups were recruited via the Psychological Science Accelerator project[25-27]. Previous studies compared two cultures to demonstrate regional differences[17,18]. By contrast, the scale and scope of our study will allow us to generate the most comprehensive picture of the world regions to which the valence-dominance model does and does not apply.

We will test two specific competing predictions.

Prediction 1. The valence-dominance model will apply to all world regions.

Prediction 2. The valence-dominance model will apply in Western-world regions, but not other world regions.

Table 1

World Regions, Countries, and Localities of Planned Data Collection

World region	Countries and Localities
Africa	Kenya, South Africa
Asia	China, India, Malaysia, Taiwan, Thailand
Australia and New Zealand	Australia, New Zealand
Central America and Mexico	Ecuador, El Salvador, Mexico
Eastern Europe	Hungary, Lithuania, Poland, Russia, Serbia, Slovakia

The Middle East	Iran, Israel, Turkey
The USA and Canada	Canada, the USA
Scandinavia	Denmark, Norway
South America	Argentina, Brazil, Chile, Colombia
The UK	England, Scotland, Wales
Western Europe	Austria, Belgium, France, Germany, Italy, the Netherlands, Portugal, Spain, Switzerland

Note. We will collect data from a minimum of 350 raters per world region based on the simulations described in the Methods section below.

Methods

Ethics

Each research group has approval from their local Ethics Committee or IRB to conduct the study, has explicitly indicated that their institution does not require approval for the researchers to conduct this type of face-rating task, or has explicitly indicated that the current study is covered by a preexisting approval. Although the specifics of the consent procedure will differ across research groups, all participants will provide informed consent. All data will be stored centrally on University of Glasgow servers.

Procedure

Oosterhof and Todorov derived their valence-dominance model from a principal components analysis of ratings (by US raters) of 66 faces for 13 different traits (aggressiveness, attractiveness, caringness, confidence, dominance, emotional stability, intelligence, meanness, responsibility, sociability, trustworthiness, unhappiness, and weirdness)[12]. Using the criteria of the number of components with eigenvalues greater than 1.0, this analysis produced two principal components. The first component explained 63% of the variance in trait ratings, strongly correlated with rated trustworthiness (r = .94), and weakly correlated with rated dominance (r = -.24). The second component explained 18% of the variance in trait ratings, strongly correlated with rated dominance (r = .93), and weakly correlated with rated trustworthiness (r = -.06). We will replicate Oosterhof and Todorov's method[12] and primary analysis in each world region we examine.

Stimuli in our study will come from an open-access, full-color, face image set[28] consisting of 60 men and 60 women taken under standardized photographic conditions (M_{age} = 26.4 years, SD = 3.6 years, Range = 18 to 35 years). These 120 images will consist of 30 Black (15 male, 15 female), 30 White (15 male, 15 female), 30 Asian (15 male, 15 female), and 30 Latin faces (15 male, 15 female). As in Oosterhof and Todorov's study[12], the individuals photographed posed looking directly at the camera with a neutral expression, and all of background, lighting, and clothing (here, a grey t-shirt) are constant across images.

In our study, adult raters will be randomly assigned to rate the 13 adjectives tested by Oosterhof and Todorov using scales ranging from 1 (*Not at all*) to 9 (*Very*) for all 120 faces in a fully randomized order at their own pace. Because all researchers will collect data through an identical interface (except for differences in instruction language), data collection protocols will be highly standardized across labs. Each participant will complete the block of 120 face-rating trials twice so that we can report test-retest reliabilities of ratings; ratings from the first and second blocks will be averaged for all analyses (see CODE 1.5.5 in the Supplemental Materials).

Raters will also complete a short questionnaire requesting demographic information (sex, age, ethnicity). These variables were not considered in Oosterhof and Todorov's analyses but will be collected in our study so that other researchers can use them in secondary analyses of the published data. The data from this study will be the largest and most comprehensive open access set of face ratings from around the world with open stimuli by far, providing an invaluable resource for further research addressing the Western centrality assumption in person perception research.

Raters will complete the task in a language appropriate for their country (see below). To mitigate potential problems with translating single-word labels, dictionary definitions for each of the 13 traits will be provided. Twelve of these dictionary definitions have previously been used to test for effects of social impressions on the memorability of face photographs[19]. Dominance (not included in that study) will be defined as "strong, important."

Participants

Simulations determined that we should obtain at least 25 different raters for each of the 13 traits in every region (see https://osf.io/x7fus/ for code and data). We focused on ratings of attractiveness and intelligence for the simulations because they showed the highest and lowest agreement among the traits analyzed by Oosterhof and Todorov, respectively. First, we sampled from a population of 2,513 raters, each of whom had rated the attractiveness of 102 faces; these simulations showed that more than 99% of 1,000 random samples of 25 raters produced good or excellent interrater reliability coefficients (Cronbach's αs >.80). We then repeated these simulations sampling from a population of 37 raters, each of whom rated the intelligence of 100 faces, showing that 93% of 1,000 random samples of 25 raters produced good or excellent interrater reliability coefficients (Cronbach's αs >.80). Thus, averages of ratings from 25 or more raters will produce reliable dependent variables in our analyses; we plan to test at least 9,000 raters in total.

In addition to rating the faces for the 13 traits examined by Oosterhof and Todorov, 25 participants in each region will be randomly assigned to rate the targets' age in light of Sutherland et al.'s results showing that a youth/attractiveness dimension emerged from analyses of a sample of faces with a very diverse age range[30]. Age ratings will not be included in analyses relating to replications of Oosterhof and Todorov's valence-dominance model, but analyzed only in additional exploratory analyses.

Analysis Plan

The code to be used for these analyses is included in the Supplemental Materials and publicly available from the Open Science Framework (https://osf.io/87rbg/). To facilitate assessment of the Stage 1 Registered Report, the specific sections of code are cited below as (CODE x.x.x).

Ratings from each world region will be analyzed separately and anonymous raw data will be published on the Open Science Framework. Our analyses will directly replicate the principal component analysis reported by Oosterhof and Todorov to test their theoretical model in each region sampled (CODE 2.1). First, we will calculate the average rating for each face separately for each of the 13 traits (CODE 2.1.2). We will then subject these mean ratings to principal component analysis with orthogonal components and no rotation, as Oosterhof and Todorov did (CODE 2.1.3). Using the criteria reported they reported, we will retain and interpret components with eigenvalues greater than 1.0 (CODE 2.1.3.1).

Criteria for replicating Oosterhof and Todorov's valence-dominance model. We will use multiple sources of evidence to judge whether Oosterhof and Todorov's valence-dominance model replicated in a given world region. First, we will examine the solution from the principal components analysis conducted in each region and determine if Oosterhof and Todorov's primary pattern replicated according to three criteria: (i) the first two components have eigenvalues greater than 1.0, (ii) the first component (i.e., the one explaining more of the variance in ratings) correlates strongly with trustworthiness ($\lambda > .7$) and weakly with dominance ($\lambda < .5$), and (iii) the second component (i.e., the one explaining less of the variance in ratings) correlates strongly with dominance ($\lambda > .7$) and weakly with trustworthiness ($\lambda < .5$). If the solution in a world region meets all three of these criteria, we will conclude that the primary pattern of the model replicated in that region (CODE 2.1.3.3).

In addition to reporting whether the primary pattern was replicated in each region, we will also report Tucker's coefficient of congruence[31,32]. The congruence coefficient, ϕ, ranges from -1 to 1 and quantifies the similarity between two vectors of loadings[33]. It is:

$$\phi(x, y) = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2 \sum y_i^2}}$$

where x_i and y_i are the loadings of variable i (i = 1, …, n number of indicators in the analysis) onto factors x and y. For the purposes of the current research we will compare the vector of loadings from the first component from Oosterhof and Todorov to the vector of loadings from the first component estimated from each world region. We will repeat this analysis for the second component. This will produce a standardized measure of component similarity for each component in each world region that is not sensitive to the mean size of the loadings[34]. Further, this coefficient is fitting for the current study because it does not require an a priori specification of a factor structure for each group, as would be needed if we were to compare the factor structures in a multiple-group confirmatory factor analysis. Following previous guidelines[34], we will conclude that the components in Oosterhof and Todorov are not similar to those estimated in a given world region if the coefficient is less than .85, are fairly similar if it is between .85 - .94, and equal if it is greater than .95. (CODE 2.1.4.2).

Thus, we will report whether the solution has the same primary pattern that Oosterhof and Todorov found and quantify the degree of similarity between each component and the corresponding component from Oosterhof and Todorov's work. This connects to our competing predictions:

Prediction 1 (The valence-dominance model will apply to all world regions) will be supported if the solution from the principal components analysis conducted in each region satisfy *all* of the criteria described above. Specifically, the primary pattern is replicated and the components have at least a fair degree of similarity as quantified by a ϕ of .85 or greater.

Prediction 2 (The valence-dominance model will apply in Western-world regions, but not other world regions) will be supported if the solutions from the principal components analysis conducted in Australia and New Zealand, The USA and Canada, Scandinavia, The UK, and Western Europe, but not Africa, Asia, Central America and Mexico, Eastern Europe, The Middle East, or South America, satisfy the criteria described above.

Exclusions. Data from raters who fail to complete all 120 ratings in the first block of trials or who provide the same rating for 75% or more of the faces will be excluded from analysis (CODES 1.5.1,1.5.3, and 1.5.5).

Data-quality checks. Following previous research testing the valence-dominance model[12-14], data quality will be checked by separately calculating the interrater agreement (indicated by Cronbach's α and test-retest reliability) for each trait in every world region (CODE 2.1.1). A trait will only be included in the analysis for that region if the coefficient exceeds .70. Cases in which the coefficient does not exceed .70 will be reported and discussed. Test-retest reliability of traits will be reported but not used to exclude traits from analysis.

Power analysis. Simulations show we have more than 95% power to detect the key effect of interest (i.e., two components meeting the criteria for replicating Oosterhof and Todorov's work, as described above). We used the open data from Morrison et al.'s replication[13] of Oosterhof and Todorov's research to generate a variance-covariance matrix representative of typical interrelationships among the 13 traits that will be tested in our study. We then generated 1,000 samples of 120 faces from these distributions and ran our planned principal components analysis (which is identical to that reported by

Oosterhof & Todorov) on each sample (see https://osf.io/87rbg/ for code and data). Results of >99% of these analyses matched our criteria for replicating Oosterhof and Todorov's findings. This demonstrates that 120 faces will give us more than 95% power to replicate Oosterhof and Todorov's results.

Robustness analyses. Oosterhof and Todorov extracted and interpreted components with an eigenvalue greater than 1.0 using an unrotated principal components analysis. As described above, we will directly replicate their method in our main analyses but acknowledge that this type of analysis has been criticized.

First, it has been argued that exploratory factor analysis with rotation, rather than an unrotated principal components analysis, is more appropriate when one intends to measure correlated latent factors, as is the case in the current study[35,36]. Second, the extraction rule of eigenvalues greater than 1.0 has been criticized for not indicating the optimal number of components, as well as for producing unreliable components[37,38].

To address these limitations, we will repeat our main analyses using exploratory factor analysis with an oblimin rotation as the model and a parallel analysis to determine the number of factors to extract. We will also recalculate the congruence coefficient described above for these exploratory factor analysis results (CODE 2.2.1).

We will use parallel analysis to determine the number of factors to extract because it has been described as yielding the optimal number of components (or factors) across the largest array of scenarios[35,39,40] (CODE 2.2.1). In a parallel analysis, random data matrices are generated such that they have the same number of cases and variables as the real data. The mean eigenvalue from the components of the random data is compared to the eigenvalue for each component from the real data. Components are then retained if their eigenvalues exceed those from the randomly generated data[41].

The purpose of these additional analyses is twofold. First, to address potential methodological limitations in the original study and, second, to ensure that the results of our replication of Oosterhof and Todorov's study are robust to the implementation of those more rigorous analytic techniques. The same criteria for replicating Oosterhof and Todorov's model described above will be applied to this analysis (CODE 2.2.4-5).

References

1. Olivola, C. Y., & Todorov, A. (2010). Elected in 100 milliseconds: Appearance-based trait inferences and voting. Journal of Nonverbal Behavior, 34, 83-110.
2. Ritchie, K. L., Palermo, R., & Rhodes, G. (2017). Forming impressions of facial attractiveness is mandatory. Scientific Reports, 7, 469.
3. Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after 100 ms exposure to a face. Psychological Science, 17, 592-598.
4. Olivola, C. Y., Funk, F., & Todorov, A. (2014). Social attributions from faces bias human choices. Trends in Cognitive Sciences, 18, 566-570.
5. Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: Determinants, consequences, accuracy, and functional significance. Annual Review of Psychology, 66, 519-545.
6. Van 't Wout, M., & Sanfey, A. G. (2008). Friend or foe: The effect of implicit trustworthiness judgments in social decision-making. Cognition, 108, 796-803.
7. Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. Science, 308, 1623-1626.
8. Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., Hallam, M., & Smoot, M. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. Psychological Bulletin, 126, 390-423.
9. Wilson, J. P., & Rule, N. O. (2015). Facial trustworthiness predicts extreme criminal-sentencing outcomes. Psychological Science, 26, 1325-1331.

10. Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. Trends in Cognitive Sciences, 12, 455-460.

11. Jack, R. E., & Schyns, P. G. (2017). Toward a social psychophysics of face communication. Annual Review of Psychology, 68, 269-297.

12. Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. Proceedings of the National Academy of Sciences of the USA, 105, 11087-11092.

13. Morrison, D., Wang, H., Hahn, A. C., Jones, B. C., & DeBruine, L. M. (2017). Predicting the reward value of faces and bodies from social perception. PLoS ONE, 12, e0185093.

14. Wang, H., Hahn, A. C., DeBruine, L. M., & Jones, B. C. (2016). The motivational salience of faces is related to both their valence and dominance. PLoS ONE, 11, e0161114.

15. Henrich, J., Heine, S., & Norenzayan, A. (2010). The weirdest people in the world? Behavioral and Brain Sciences, 33, 61-83.

16. Kline, M. A., Shamsudheen, R., & Broesch, T. (2018). Variation is the universal: Making cultural evolution work in developmental psychology. Philosophical Transactions of the Royal Society B, 373, 20170059.

17. Sutherland, C. A. M., Liu, X., Zhang, L., Chu, Y., Oldmeadow, J. A., & Young, A. W. (2018). Facial first impressions across culture: Data-driven modeling of Chinese and British perceivers' unconstrained facial impressions. Personality and Social Psychology Bulletin, 44, 521-537.

18. Wang, H., Han, C., Hahn, A., Fasolt, V., Morrison, D.,... Jones, B. C. (2018). A data-driven study of Chinese participants' social judgments of Chinese faces. PsyArXiv.

19. Han, C., Wang, H., Hahn, A. C., Fisher, C. I., Kandrik, M., Fasolt, V., ... & Jones, B. C. (2018). Cultural differences in preferences for facial coloration. Evolution and Human Behavior, 39, 154-159.

20. Perrett, D. I., Lee, K. J., Penton-Voak, I., Rowland, D., Yoshikawa, S., Burt, D. M., ... & Akamatsu, S. (1998). Effects of sexual dimorphism on facial attractiveness. Nature, 394, 884.

21. Xie, S.Y., Flake, J.K., & Hehman, E. (in press). Perceiver and target characteristics contribute to impression formation differently across race and gender. Journal of Personality and Social Psychology.

22. Li, N. P., Valentine, K. A., & Patel, L. (2011). Mate preferences in the US and Singapore: A cross-cultural test of the mate preference priority model. Personality and Individual Differences, 50, 291-294.

23. Ting-Toomey, S. (1994). Face and facework: An introduction. In S. Ting-Toomey (Ed.),The Challenge of Facework: Cross-Cultural and Interpersonal Issues (pp. 1-14). Albany, NY: State University of New York Press.

24. Tan, C. B. Y., Stephen, I. D., Whitehead, R., Sheppard, E. (2012). You look familiar: How Malaysian Chinese recognize faces. PLoS ONE 7, e29714.

25. Chartier, C, McCarthy, R., & Urry, H. (2018). The Psychological Science Accelerator. APS Observer, 31, 30.

26. Chawla, D. S. (2017). A new 'accelerator' aims to bring big science to psychology. Science. doi:10.1126/science.aar4464

27. Moshontz, H., Campbell, L., Ebersole, C. R., IJzerman, H., Urry, H. L., Forscher, P. S., … Chartier, C. R. (in press). The Psychological Science Accelerator: Advancing psychology through a distributed collaborative network. Advances in Methods and Practices in Psychological Science.

28. Ma, Correll, & Wittenbrink (2015). The chicago face database: A free stimulus set of faces and norming data. Behavior Research Methods, 47, 1122-1135.

29. Bainbridge, W. A., Isola, P., & Oliva, A. (2013). The intrinsic memorability of face photographs. Journal of Experimental Psychology: General, 142, 1323-1334.

30. Sutherland, C. A., Oldmeadow, J. A., Santos, I. M., Towler, J., Burt, D. M., & Young, A. W. (2013). Social inferences from faces: Ambient images generate a three-dimensional model. Cognition, 127, 105-118.

31. Burt, C. (1948). The factorial study of temperament traits. British Journal of Psychology, Statistical Section, 1, 178-203.

32. Tucker, L. R. (1951). A method for synthesis of factor analysis studies (Personnel Research Section Report No. 984). Washington, DC: Department of the Army.

33. Davenport, E. C., Jr., (1990). Significance testing of congruence coefficients: A good idea? Educational and Psychological Measurement, 50, 289-296.

34. Lorenzo-Seva, U., & ten Berge, J. M. F. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. Methodology, 2, 57-64.

35. Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. Psychological Methods, 4, 272-299.

36. Park, H. S., Dailey, R., & Lemus, D. (2002). The use of exploratory factor analysis and principal components analysis in communication research. Human Communication Research, 28, 562-577.

37. Cliff, N. (1988). The Eigenvalues-greater-than-one rule and the reliability of components. Psychological Bulletin, 103, 276-279.

38. Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. Psychological Bulletin, 99, 432-442.

39. O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and velicer's MAP test. Behavior Research Methods, Instruments, & Computers, 32, 396-402.

40. Schmitt, T. A. (2011). Current methodological considerations in exploratory and confirmatory factor analysis. Journal of Psychoeducational Assessment, 29, 304-321.

41. Courtney, M. G. R. (2013). Determining the number of factors to retain in EFA_: Using the SPSS R-Menu v2.0 to make more judicious estimations. Practical Assessment, Research & Evaluation, 18, 1-14.

Author contributions. Benedict Jones, Lisa DeBruine and Jessica Flake proposed and designed the project, designed the analysis plan, drafted and revised the Stage 1 submission, and will carry out data collection. Christopher Chartier is the Director of the Psychological Science Accelerator, will carry out data collection, and drafted and revised Stage 1 submission. All other authors had input into the design of the project and analysis plan, revised the Stage 1 submission, and will carry out data collection.

Acknowledgments. This is the first empirical study selected to be run via the Psychological Science Accelerator, a new initiative for conducting large-scale psychological research (https://psysciacc.org/). The authors are grateful to Alex Todorov, Alex Jones, Chris Chambers for feedback on this project. Lisa M DeBruine is supported by the European Research Council (KINSHIP). Daniel Ansari is supported by the Natural Sciences and Engineering Council of Canada. Krystian Barzykowski is supported by the National Science Centre, Poland (2015/19/D/HS6/00641). Coralie Chevallier is supported by grants ANR-10-LABX-0087 IEC and ANR-10-IDEX-0001-02 PSL. Nicholas A Coles is supported by the National Science Foundation Graduate Research Fellowship #R010138018. Ana Maria Fernandez is supported by Fondecyt #1181114. Tripat Gill is supported by the Social Sciences and Humanities Research Council of Canada. Taylor D Gogan is supported by an Australian Government Research Training Program Scholarship. Isaac González-Santoyo is supported by grant CONACYT-CIENCIA BÁSICA #241744. Eric Hehman is supported by SSHRC 430-2016-00094. Hans IJzerman is supported by a French National Research Agency “Investissements d’avenir” program grant (ANR-15-IDEX-02). Chun-Chia Kung is supported by grant MOST-106-2410-H-006-036. Erica D Musser is supported by grant NIMH R03 MH110812 02. Yue Qi is supported by the Beijing Natural Science Foundation (5184035), the Scientific Foundation of the Institute of Psychology, Chinese Academy of Sciences (Y5CX122005). Nicholas O Rule is supported by the Social Sciences and Humanities Research Council of Canada. Waldir M Sampaio is supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES). Diana R Santos is supported by the Conselho Nacional de Desenvolvimento Cientifico e Tecnológico - Brasil (CNPq). Christoph Schild is supported by The Carlsberg Foundation (CF16-0444). Miguel A Vadillo is supported by grants 2016-T1/SOC-1395 from Comunidad de Madrid (Programa de Atracción de Talento Investigador) and PSI2017-85159-P from Agencia Estatal deInvestigación, Ministerio de Economía y Competitividad. Evie Vergauwe is supported by the Swiss National Science Foundation (PZ00P1_154911). Wen-Jing Yan is supported by the National Natural Science Foundation of China (31500875). Ingo Zettler is supported by The Carlsberg Foundation (CF16-0444). Christopher R Chartier is supported by the Society for Personality and Social Psychology. The funders had or will have no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests. The authors declare no competing interests.

Translation Guidelines

This section describes the procedure we will use to translate instructions, trait labels, and trait definitions from English to the testing languages used in each region. This process reflects and extends best practice in translating for cross-cultural research, as described in Brislin (1970).

Translation personnel.

***Language coordinator*.** Will coordinate translation process and discuss final version with translators.

A translators. Will translate from English to target language and discuss final version with coordinator and B translator (*N* = 2, both bilingual).

***B translators*.** Will translate from target language to English and discuss final version with coordinator and A translator (*N* = 2, both bilingual).

***External readers*.** Will read materials for final clarity check (*N* = 2, both nonacademics).

Individual researchers (or research groups) carrying out data collection. Will provide final checks and suggest any necessary cultural adjustments.

Translation process.

Step 1 (translation). Original document is translated from English to target language by A translators, resulting in document Version A.

Step 2 (back-translation). Version A is translated back from target language to English by B translators, independently resulting in Version B.

Step 3 (discussion). Version A and B are discussed among translators and the language coordinator, discrepancies in Version A and B are detected and solutions discussed. Version C is created.

Step 4 (external readings). Version C is tested on two nonacademics fluent in the target language. Members of the fluent group are asked how they perceive and understand the translation. Possible misunderstandings are noted and again discussed as in Step 3.

Step 5 (possible cultural adjustments). Data collection labs read materials and identify any adjustments for their local participant sample. Adjustments are discussed with the language coordinator, who makes any necessary changes, resulting in the final version for each site.

This process will produce the Final Translated Document, containing the instructions that will be used in the study.

Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology, 1*, 185-216.

PSA1: Planned Analyses

Code

1 Load Data

This script requires up-to-date packages (especially tidyverse 1.2.1).

Code

```
## Loading required package: viridisLite
```

Code

```
## ── Attaching packages ───────────────────────────────────────────────────────────────────────── tidyverse 1.2.1 ──
```

```
## ✔ ggplot2 2.2.1     ✔ purrr   0.2.4
## ✔ tibble  1.4.2     ✔ dplyr   0.7.4
## ✔ tidyr   0.8.0     ✔ stringr 1.2.0
## ✔ readr   1.1.1     ✔ forcats 0.2.0
```

```
## ── Conflicts ──────────────────────────────────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ ggplot2::%+%()   masks psych::%+%()
## ✖ ggplot2::alpha() masks psych::alpha()
## ✖ dplyr::filter()  masks stats::filter()
## ✖ dplyr::lag()     masks stats::lag()
## ✖ dplyr::select()  masks MASS::select()
```

Code

```
## [1] "R version 3.4.3 (2017-11-30)"
```

1.1 Simulate Study Data (for Stage 1 RR)

1.1.1 Load Data for Simulations (from OSF)

Code

1.1.2 Simulate multivariate distribution

Code

1.1.3 Simulate one region

Code

1.2 Load Study Data (for Stage 2 RR)

Load study data and demographic questionnaires from the data folder.

Code

1.3 Load Auxillary Data

Data on regions and stimuli.

1.3.1 Load Region Data

Code

1.3.2 Load Stimulus Info

Code

ethnicity	gender	n	mean_age	sd_age
asian	female	15	26.15	3.33
asian	male	15	26.40	3.21
black	female	15	27.00	3.51
black	male	15	28.07	4.27
latinx	female	15	25.27	2.42
latinx	male	15	26.31	4.00
white	female	15	25.77	3.03
white	male	15	26.06	4.46

Code

Stimuli in our study will be an open-access, full-color, face image set consisting of 60 men and 60 women (mean age=26.38 years, SD=3.57 years, range=18.7307692 to 34.9310345 years), taken under standardized photographic conditions (Ma et al., 2015).

1.3.3 Load O&T 2008 Loadings

Factor loadings from Oosterhof & Todorov (2008) will be used below for Factor Congruence analyses.

Code

	P1	P2
aggressive	-0.71	0.66
attractive	0.81	0.32
caring	0.90	0.29
confident	0.68	0.65
dominant	-0.24	0.93
emostable	0.93	0.19
intelligent	0.72	0.13
mean	-0.76	0.55

responsible	0.91	0.11
sociable	0.91	0.20
trustworthy	0.94	0.06
unhappy	-0.71	0.01
weird	-0.87	0.22

1.4 Data Processing

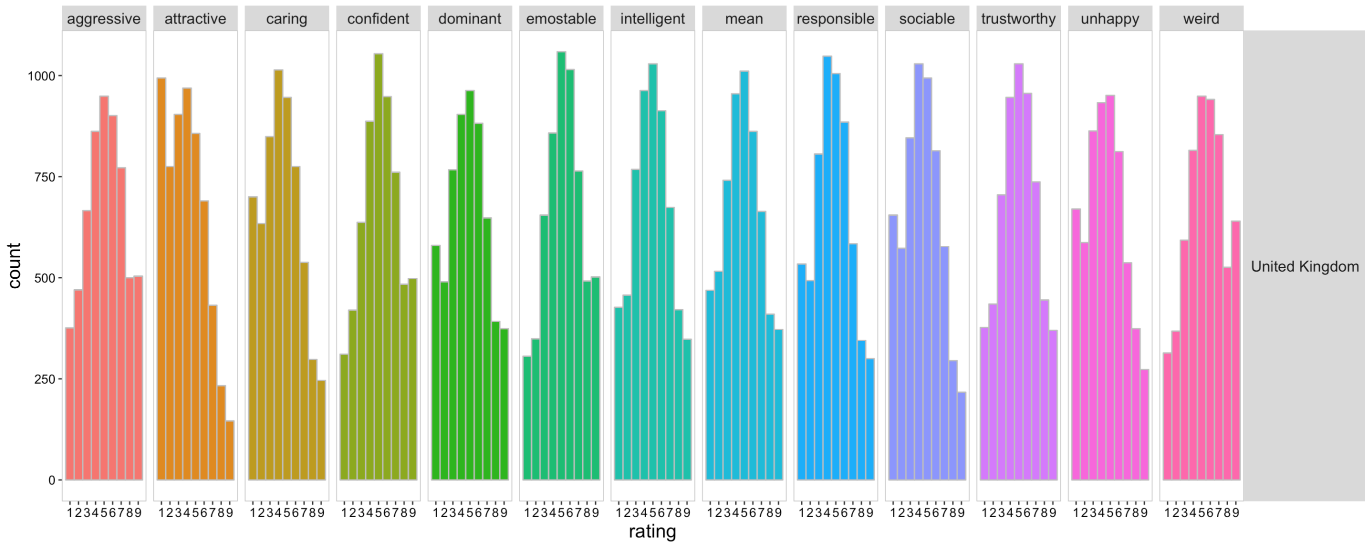
1.4.1 Join Data

Code

1.4.2 Graph distributions for trait by region

Code

Code



1.5 Data checks

Code

1.5.1 Participants who did not complete at least one rating for each of 120 stimuli

Code

user_id	sex	age	country	language	trait	region	lab	trials	stim_n

1.5.2 Participants who did not complete exactly 240 trials

Code

user_id	sex	age	country	language	trait	region	lab	trials	stim_n

1.5.3 Participants with low-variance responses in block 1

Code

trait	TOTAL
TOTAL	0

1.5.4 Participants with no region

Code

```
## # A tibble: 0 x 3
## # ... with 3 variables: user_id <int>, country <chr>, lab <chr>
```

Code

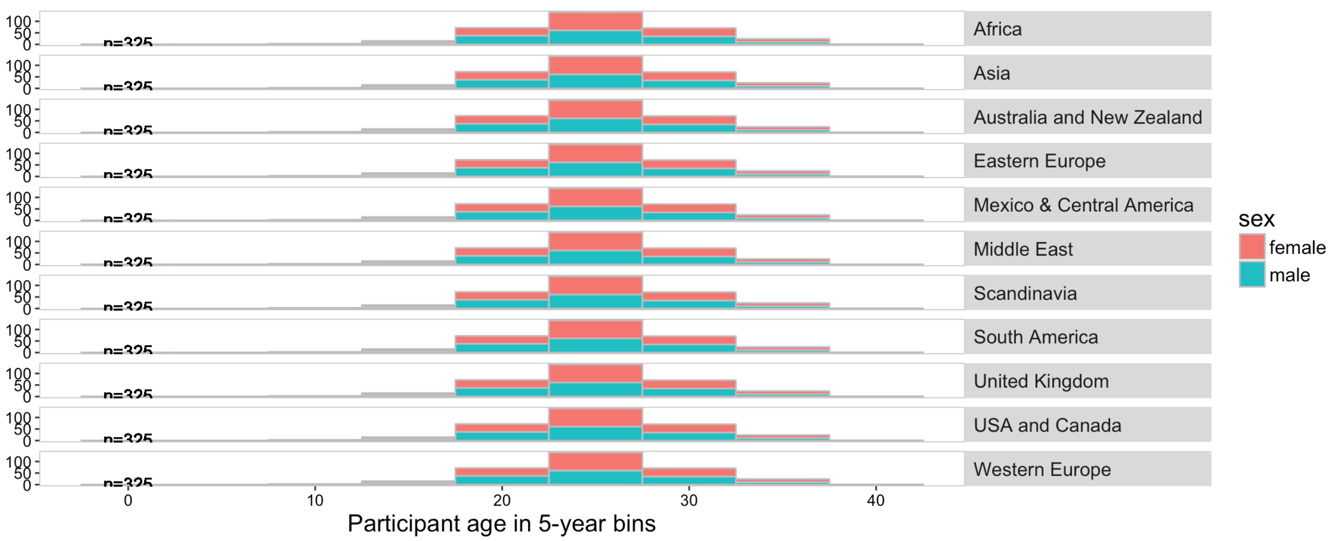
1.5.5 Remove excluded data and average ratings

Code

1.6 Participant Demographics

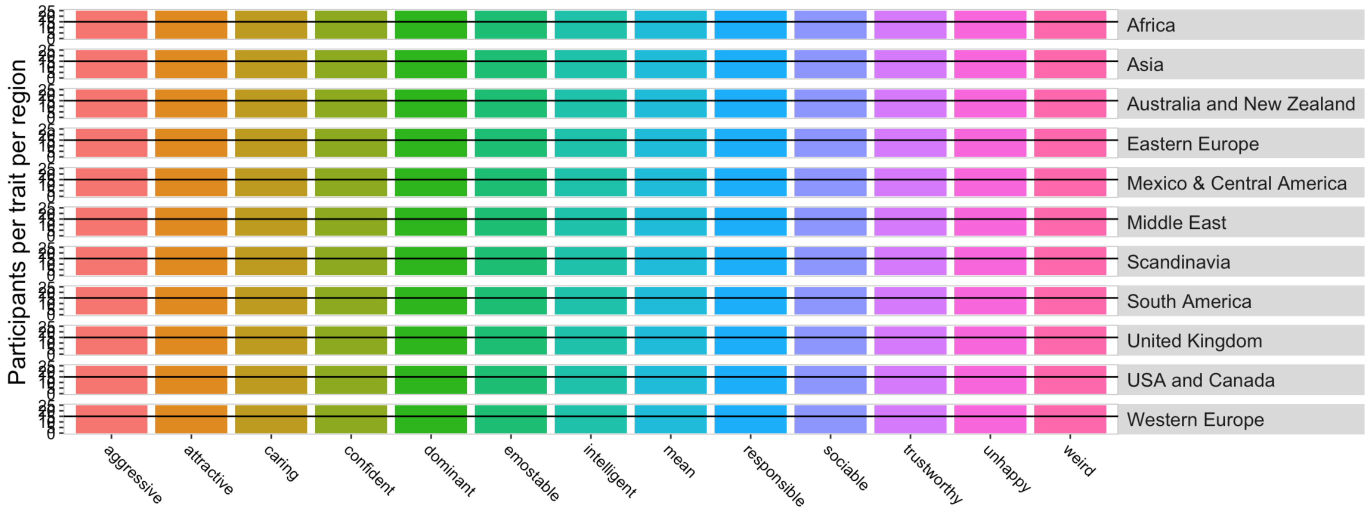
1.6.1 Age and sex distribution per region

Code



1.6.2 Participants per trait per region

Code



2 Analyses

2.1 Main Analysis

First, we will calculate the average rating for each face separately for each of the 13 traits. Like Oosterhof and Todorov (2008), we will then subject these mean ratings to principal component analysis with orthogonal components and no rotation. Using the criteria reported in Oosterhof and Todorov's (2008) paper, we will retain and interpret the components with an Eigenvalue > 1.

2.1.1 Calculate Alphas

Code

Code

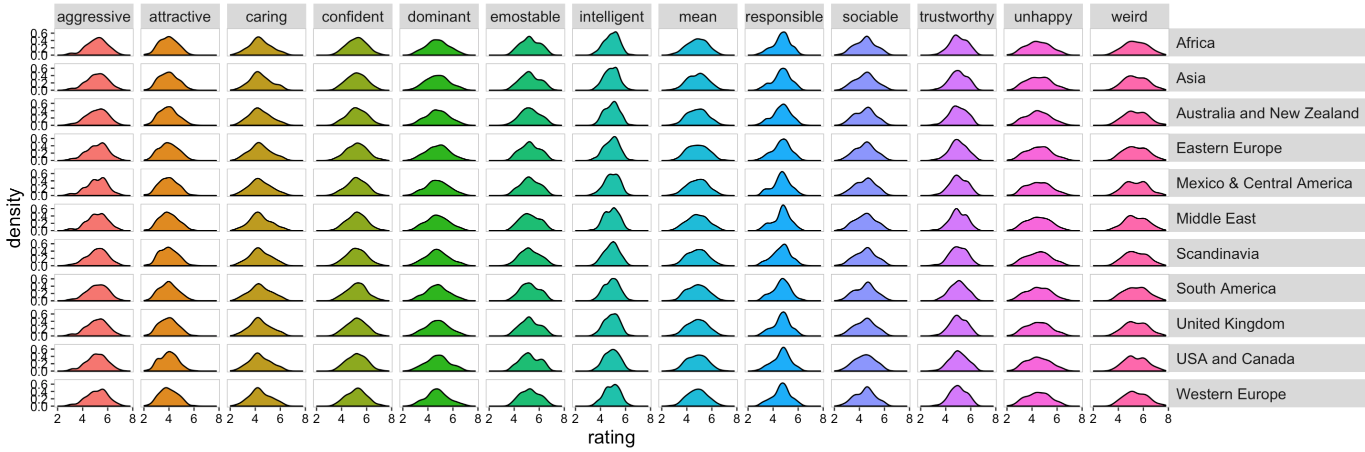
Cronbach's Alphas

	Africa	Asia	Australia & New Zealand	Eastern Europe	Mexico & Central America	Middle East	Scandinavia	South America	United Kingdom	USA & Canada	Western Europe
aggressive	0.879	0.883	0.873	0.878	0.88	0.877	0.872	0.877	0.898	0.873	0.881
attractive	0.841	0.847	0.846	0.854	0.849	0.852	0.84	0.855	0.875	0.841	0.844
caring	0.888	0.879	0.883	0.883	0.887	0.886	0.876	0.893	0.904	0.885	0.884
confident	0.853	0.856	0.857	0.868	0.866	0.854	0.861	0.863	0.885	0.858	0.864
dominant	0.893	0.895	0.897	0.892	0.9	0.894	0.895	0.891	0.912	0.896	0.894
emostable	0.852	0.864	0.861	0.853	0.857	0.86	0.864	0.853	0.884	0.863	0.856
intelligent	0.761	0.739	0.749	0.762	0.765	0.746	0.749	0.763	0.797	0.765	0.777
mean	0.872	0.868	0.869	0.869	0.875	0.88	0.868	0.874	0.894	0.871	0.867
responsible	0.822	0.837	0.832	0.827	0.83	0.83	0.834	0.824	0.857	0.833	0.835
sociable	0.875	0.877	0.871	0.863	0.871	0.878	0.875	0.875	0.895	0.872	0.87
trustworthy	0.816	0.812	0.812	0.811	0.806	0.799	0.805	0.811	0.834	0.82	0.799
unhappy	0.904	0.908	0.904	0.907	0.908	0.903	0.907	0.908	0.921	0.902	0.902
weird	0.904	0.897	0.904	0.901	0.898	0.9	0.899	0.904	0.918	0.899	0.901

2.1.2 Calculate Aggregate Scores

Code

Code



2.1.3 PCA

Code

2.1.3.1 Number of PCs (and proportion variance) by region

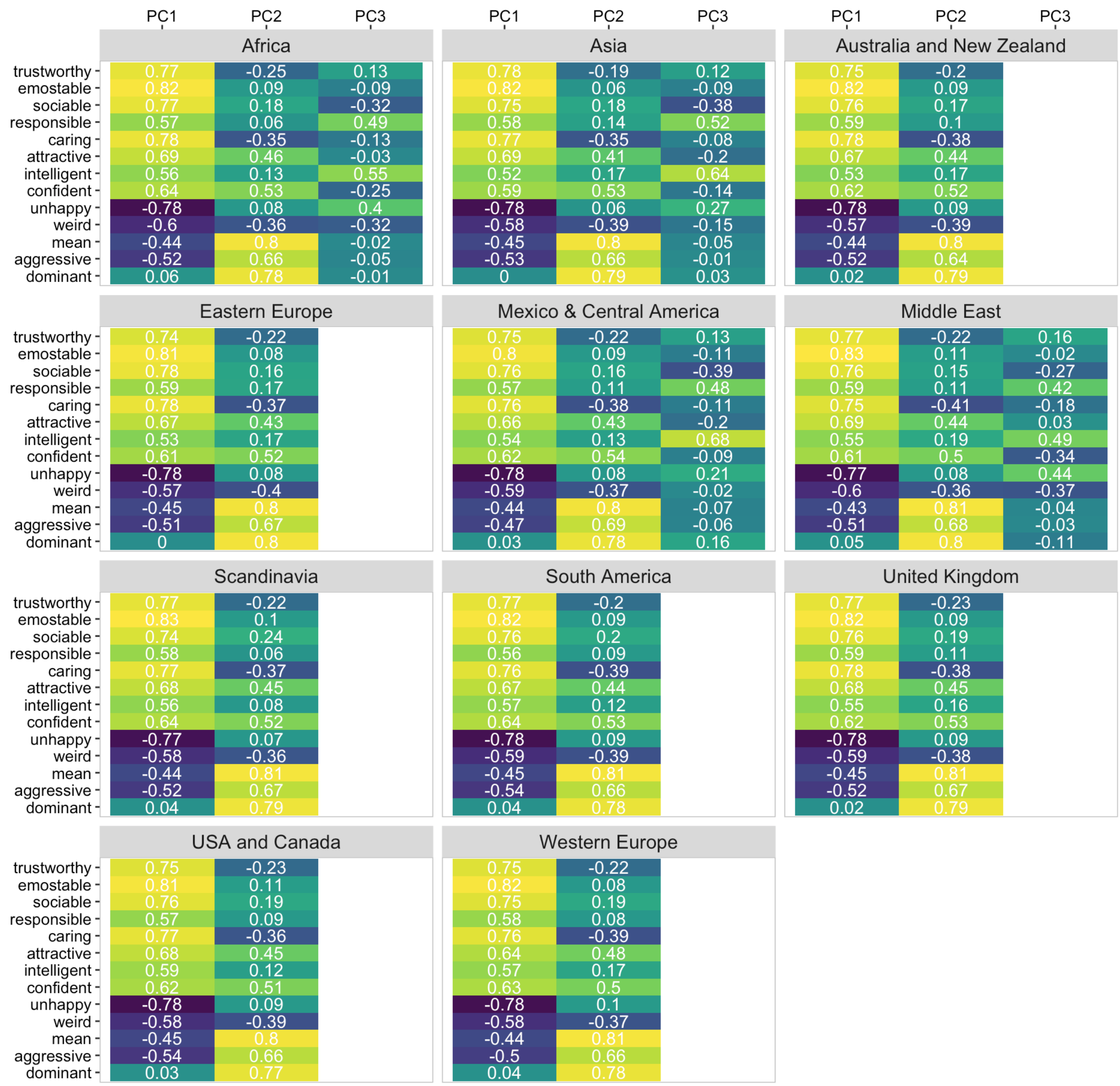
Code

region	nPCs	PC1	PC2	PC3
Africa	3	0.417	0.197	0.079
Asia	3	0.408	0.196	0.079
Australia and New Zealand	2	0.405	0.197	NA
Eastern Europe	2	0.404	0.201	NA
Mexico & Central America	3	0.400	0.201	0.078
Middle East	3	0.412	0.202	0.077
Scandinavia	2	0.410	0.201	NA
South America	2	0.411	0.199	NA
United Kingdom	2	0.414	0.203	NA
USA and Canada	2	0.410	0.196	NA
Western Europe	2	0.405	0.201	NA

2.1.3.2 Trait Loadings by Region and PC

Code

Trait Loadings



2.1.3.3 PCA Replication criteria

Oosterhof and Todorov's valence-dominance model will be judged to have been replicated in a given world region if the first two components both have Eigenvalues > 1, the first component (i.e., the one explaining more of the variance in ratings) is correlated strongly (loading > .7) with trustworthiness and weakly (loading < .5) with dominance, and the second component (i.e., the one explaining less of the variance in ratings) is correlated strongly (loading > .7) with dominance and weakly (loading < .5) with trustworthiness. All three criteria need to be met to conclude that the model was replicated in a given world region.

Code

region	PC1 dominant	PC1 trustworthy	PC2 dominant	PC2 trustworthy	replicated
Africa	0.06	0.77	0.78	-0.25	TRUE
Asia	0.00	0.78	0.79	-0.19	TRUE

Australia and New Zealand	0.02	0.75	0.79	-0.20	TRUE
Eastern Europe	0.00	0.74	0.80	-0.22	TRUE
Mexico & Central America	0.03	0.75	0.78	-0.22	TRUE
Middle East	0.05	0.77	0.80	-0.22	TRUE
Scandinavia	0.04	0.77	0.79	-0.22	TRUE
South America	0.04	0.77	0.78	-0.20	TRUE
United Kingdom	0.02	0.77	0.79	-0.23	TRUE
USA and Canada	0.03	0.75	0.77	-0.23	TRUE
Western Europe	0.04	0.75	0.78	-0.22	TRUE

2.1.4 Factor Congruence

Code

region	origP1_regionPC1	origP1_regionPC2	origP2_regionPC1	origP2_regionPC2
Africa	0.98 **	-0.1	0.19	0.8
Asia	0.98 **	-0.07	0.16	0.8
Australia and New Zealand	0.98 **	-0.08	0.18	0.79
Eastern Europe	0.98 **	-0.07	0.17	0.79
Mexico & Central America	0.98 **	-0.09	0.19	0.8
Middle East	0.98 **	-0.09	0.18	0.79
Scandinavia	0.98 **	-0.09	0.19	0.8
South America	0.98 **	-0.08	0.18	0.79
United Kingdom	0.98 **	-0.08	0.18	0.8

USA and Canada	0.98 **	-0.08	0.18	0.8
Western Europe	0.98 **	-0.09	0.19	0.79

factor congruence < .85 = not similar
* factor congruence >= .85 and < .95 = fairly similar
** factor congruence > .95 = equal

2.2 Robustness Checks

2.2.1 Exploratory Factor Analysis (EFA)

Code

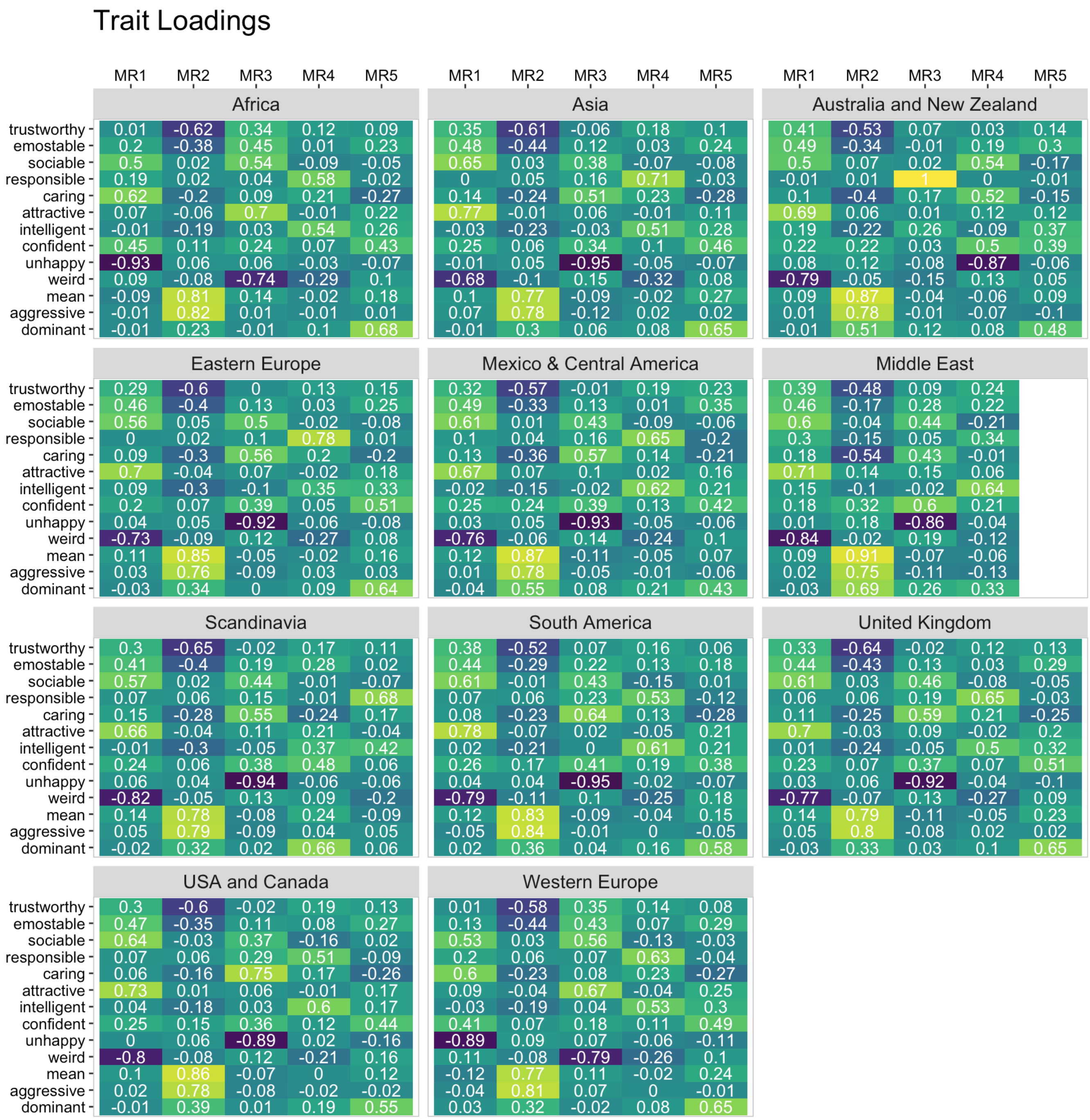
2.2.2 Number of MRs (and proportion variance) by region

Code

region	nMRs	MR1	MR2	MR3	MR4	MR5
Africa	5	0.170	0.176	0.168	0.081	0.088
Asia	5	0.179	0.177	0.145	0.097	0.088
Australia and New Zealand	5	0.173	0.191	0.103	0.158	0.069
Eastern Europe	5	0.162	0.185	0.154	0.087	0.091
Mexico & Central America	5	0.174	0.189	0.154	0.096	0.068
Middle East	4	0.199	0.207	0.163	0.087	NA
Scandinavia	5	0.169	0.185	0.157	0.099	0.076
South America	5	0.185	0.176	0.169	0.086	0.070
United Kingdom	5	0.174	0.184	0.158	0.084	0.097
USA and Canada	5	0.182	0.178	0.163	0.080	0.075
Western Europe	5	0.162	0.176	0.164	0.086	0.097

2.2.3 Trait Loadings by Region and MR

Code



2.2.4 Replication criteria

Code

region	MR1 dominant	MR1 trustworthy	MR2 dominant	MR2 trustworthy	replicated
Africa	-0.01	0.01	0.23	-0.62	FALSE
Asia	-0.01	0.35	0.30	-0.61	FALSE
Australia and New Zealand	-0.01	0.41	0.51	-0.53	FALSE
Eastern Europe	-0.03	0.29	0.34	-0.60	FALSE
Mexico & Central	-0.04	0.32	0.55	-0.57	FALSE

America					
Middle East	-0.03	0.39	0.69	-0.48	FALSE
Scandinavia	-0.02	0.30	0.32	-0.65	FALSE
South America	0.02	0.38	0.36	-0.52	FALSE
United Kingdom	-0.03	0.33	0.33	-0.64	FALSE
USA and Canada	-0.01	0.30	0.39	-0.60	FALSE
Western Europe	0.03	0.01	0.32	-0.58	FALSE

2.2.5 Factor Congruence

Code

region	origP1_regionMR1	origP1_regionMR2	origP2_regionMR1	origP2_regionMR2
Africa	0.62	-0.59	0.3	0.48
Asia	0.69	-0.6	0.3	0.47
Australia and New Zealand	0.71	-0.56	0.24	0.6
Eastern Europe	0.68	-0.61	0.25	0.48
Mexico & Central America	0.71	-0.53	0.25	0.62
Middle East	0.77	-0.54	0.24	0.66
Scandinavia	0.68	-0.62	0.26	0.46
South America	0.7	-0.55	0.25	0.56
United Kingdom	0.68	-0.61	0.26	0.49
USA and Canada	0.7	-0.56	0.25	0.57
Western Europe	0.62	-0.6	0.28	0.5

factor congruence < .85 = not similar
* factor congruence >= .85 and < .95 = fairly similar
** factor congruence > .95 = equal

~~~~
---
title: 'PSA1: Planned Analyses'
output:
  html_document:
    code_folding: hide
    number_sections: true
    toc: yes
    toc_depth: 5
---

# Load Data

This script requires up-to-date packages (especially tidyverse 1.2.1).

```{r libraries, messages = FALSE}
library(psych) # for SPSS-style PCA
library(MASS) # for simulating multivariate normal distributions
library(paran) # for parallel analyses
library(GPArotation) # for robustness checks
library(viridis) # for nice colours
library(kableExtra) # for nice tables
library(tidyverse) # for data cleaning
R.version.string
```


## Simulate Study Data (for Stage 1 RR)

### Load Data for Simulations (from OSF)

```{r message=FALSE, eval = T}

data_sim_agg <- rbind(
    read_csv("https://osf.io/6sz8k/download") %>% # male ratings
      gather(stim_id, rating, andrej:vladislav),
    read_csv("https://osf.io/375ag/download") %>% # female ratings
      gather(stim_id, rating, alexandra:zlata)
  ) %>%
  filter(type == "faces") %>%
  rename(
    rater_id = user_id,
    stim_sex = stimulus_sex,
    trait = judgment
  ) %>%
  # reverse code happy to unhappy
  mutate(
    rating = ifelse(trait == "happy", 7 - rating, rating),
    trait = ifelse(trait == "happy", "unhappy", trait),
    trait = ifelse(trait == "emotionally_stable", "emostable", trait)
  ) %>%
~~~~

````
  group_by(stim_id, trait) %>%
  summarise(rating = mean(rating)) %>%
  ungroup() %>%
  mutate(rating = rating / 7 * 9) %>% # convert to 9-point scale
  spread(trait, rating)
```


### Simulate multivariate distribution

```{r sim-intercept-matrix, eval = T}
# function to generate n stimulus intercepts from rating data (only rating
columns)
sim_agg_ratings <- function(data, n) {
  mu <- data %>% # the means of each rating
    summarise_all("mean") %>%
    t() %>%
    as.vector()

  stdevs <- data %>% # the SDs of each rating
    summarise_all("sd") %>%
    t() %>%
    as.vector()

  cor_mat <- cor(data)

  sigma <- (stdevs %*% t(stdevs)) * cor_mat

  mvrnorm(n, mu, sigma)
}

# generate a sample intercept matrix for 120 faces
dat_sim_test <- data_sim_agg %>%
    select(-stim_id) %>%
    sim_agg_ratings(120) %>%
    t() %>%
    as.data.frame() %>%
    magrittr::set_colnames( paste0("t", 1:120)) %>%
    rownames_to_column(var = "trait")

```


### Simulate one region

```{r sim-data, eval = T}
# simulate data for Stage 1 RR, load it below for actual study

sim_n <- 25
user_ids <- 1:(13*sim_n)
````

~~~~
ratings_sim <- purrr::map_df(1:(2*sim_n), function(x) {
  dat_sim_test %>%
    mutate(user_i = rnorm(nrow(.))) %>% # add a user intercept to simulate
rater variation
    mutate_at(vars(t1:t120), funs(
      pmax.int(1, pmin.int(9, round(rnorm(13, ., 2) + user_i)))
    )) %>%
    select(-user_i)
})

ratings_raw <- ratings_sim %>%
  mutate(side1 = NA,  # include some vars so code below works
         order120 = NA,
         endtime = NA,
         starttime = NA,
         language = "English",
         id = row_number(),
         user_id = c(user_ids, user_ids))

demog <- tibble(
  user_id = user_ids,
  language = "English",
  country = "GB",
  endtime = NA,
  starttime = NA,
  sex = sample(0:1, 13*sim_n, replace = T),
  age = rpois(13*sim_n, 25),
  ethnicity = "",
  lab = "simdata"
)

```


## Load Study Data (for Stage 2 RR)

Load study data and demographic questionnaires from the data folder.

```{r, message=FALSE, eval = F}

# read all the files into a list
files <- tibble(
  filename = list.files("data", full.names = TRUE)
) %>%
  filter(filename != "data/regions.csv") %>%
  filter(filename != "data/psa_cfd_faces.csv") %>%
  separate(filename, c("data", "type", "language", "csv"), remove = F) %>%
  select(-data, -csv)

demog <- files %>%
~~~~

```
  filter(type == "demog") %>%
  mutate(data = map(filename, read_csv)) %>%
  unnest() %>%
  select(-filename, -type) %>%
  # select only first time through demog
  arrange(user_id, endtime) %>%
  group_by(user_id) %>%
  filter(row_number() == 1) %>%
  ungroup()

ratings_raw <- files %>%
  filter(type == "ratings") %>%
  mutate(data = map(filename, read_csv)) %>%
  unnest() %>%
  select(-filename, -type)
```

Load Auxillary Data

Data on regions and stimuli.

Load Region Data

```{r load-region, message=FALSE}
# https://raw.githubusercontent.com/lukes/ISO-3166-Countries-with-Regional-
Codes/master/all/all.csv
# na = "" because otherwise Namibia changes to NA
regions <- read_csv("data/regions.csv", na = "") %>%
  rename(country = `alpha-2`) %>%
  select(name, country, psa_region) %>%
  rename(region = psa_region)
```

Load Stimulus Info

```{r load-stim-info, message=FALSE}
stim_info <- read_csv("data/psa_cfd_faces.csv") %>%
  mutate(ethnicity = recode(Race, "A" = "asian", "B" = "black", "L" =
"latinx", "W" = "white"),
         gender = recode(Gender, "M" = "male", "F" = "female")
  )

stim_info %>%
  group_by(ethnicity, gender) %>%
  summarise(
    n = n(),
    mean_age = round(mean(Age), 2),
    sd_age = round(sd(Age), 2)
```

````
  ) %>%
  knitr::kable("html") %>%
  kable_styling("striped")

stim_n_male <- sum(stim_info$gender == "male")
stim_n_female <- sum(stim_info$gender == "female")
mean_age <- mean(stim_info$Age) %>% round(2)
sd_age <- sd(stim_info$Age) %>% round(2)
min_age <- min(stim_info$Age)
max_age <- max(stim_info$Age)
```

Stimuli in our study will be an open-access, full-color, face image set
consisting of `r stim_n_male` men and `r stim_n_female` women (mean age=`r
mean_age` years, SD=`r sd_age` years, range=`r min_age` to `r max_age`
years), taken under standardized photographic conditions (Ma et al., 2015).

### Load O&T 2008 Loadings

Factor loadings from Oosterhof & Todorov (2008) will be used below for
Factor Congruence analyses.

```{r otloadings}
ot2008_pca_loadings <- read_csv(
"trait,P1,P2
trustworthy,0.94,0.06
emostable,0.93,0.19
responsible,0.91,0.11
sociable,0.91,0.20
caring,0.90,0.29
weird,-.87,0.22
attractive,0.81,0.32
mean,-.76,0.55
intelligent,0.72,0.13
aggressive,-.71,0.66
unhappy,-.71,0.01
confident,0.68,0.65
dominant,-.24,0.93"
) %>%
  arrange(trait) %>%
  as.data.frame() %>%
  column_to_rownames(var = "trait")

ot2008_efa_loadings <- ot2008_pca_loadings # get or calculate EFA loadings
from Todorov

ot2008_pca_loadings %>%
  knitr::kable("html") %>%
  kable_styling("striped")
```
````
````

## Data Processing

### Join Data

```{r}
ratings <- ratings_raw %>%
  arrange(user_id, endtime) %>%
  group_by(user_id) %>%
  mutate(block = row_number()) %>%
  ungroup() %>%
  gather("var", "val", t1:order120) %>%
  mutate(
    "stim_id" = str_replace(var, "[a-z]+", ""),
    "var" = str_replace(var, "\\d+", "")
  ) %>%
  spread(var, val, convert = TRUE) %>%
  rename("rating" = t) %>%
  filter(!is.na(rating)) %>% # remove missing ratings
  select(-side, -id, -starttime, -endtime) %>% # side is irrelevant for 1-
item ratings
  left_join(demog, by = c("user_id", "language")) %>%
  select(-starttime, -endtime) %>%
  mutate(sex = recode(sex, "0" = "male", "1" = "female", "2" = "other", "3"
= "no answer")) %>%
  left_join(regions, by = "country")

```

### Graph distributions for trait by region
```{r}
# plot styles
bgcolor <- "white"
textcolor <- "black"
PSA_theme <- theme(
    plot.background = element_rect(fill = bgcolor, color = NA),
    panel.background = element_rect(fill = NA, color = "grey"),
    legend.background = element_rect(fill = NA),
    panel.grid.major = element_blank(),
    panel.grid.minor = element_blank(),
    text = element_text(color = textcolor, size=15),
    axis.text = element_text(color = textcolor, size=10),
    strip.text.y = element_text(angle = 0, hjust = 0)
  )
```

```{r, fig.width = 15, fig.height=6}

ggplot(ratings, aes(rating, fill = trait)) +
```

  geom_histogram(binwidth = 1, color = "grey", show.legend = F) +
  facet_grid(region~trait) +
  scale_x_continuous(breaks = 1:9) +
  PSA_theme

```

## Data checks

```{r}
part <- ratings %>%
  group_by(user_id, sex, age, country, language, trait, region, lab) %>%
  summarise(trials = n(),
            stim_n = n_distinct(stim_id)) %>%
  ungroup()
```

### Participants who did not complete at least one rating for each of 120 stimuli
```{r}
part %>%
  filter(stim_n != 120) %>%
  knitr::kable("html") %>%
  kable_styling("striped")
```

### Participants who did not complete exactly 240 trials
```{r}
part %>%
  filter(trials != 240) %>%
  knitr::kable("html") %>%
  kable_styling("striped")
```

### Participants with low-variance responses in block 1
```{r}

identical_rating_threshold <- 0.75 * 120 # use this for registered analyses

inv_participants <- ratings %>%
  filter(block == 1) %>%
  count(user_id, region, trait, rating) %>%
  group_by(user_id, region, trait) %>%
  filter(n == max(n)) %>% # find most common rating for each P
  ungroup() %>%
  filter(n >= identical_rating_threshold) # select Ps who gave the same rating to >= 75% of stimuli

inv <- inv_participants %>%
  count(region, trait) %>%

```
  spread(region, nn, fill = 0) %>%
  mutate(TOTAL = rowSums(select_if(., is.numeric), na.rm = T))

inv_total <-  group_by(inv) %>%
  summarise_if(is.numeric, sum, na.rm = T) %>%
  mutate(trait = "TOTAL")

bind_rows(inv,inv_total) %>%
  knitr::kable("html") %>%
  kable_styling("striped")
```

### Participants with no region
```{r}
part %>%
  filter(is.na(region)) %>%
  select(user_id, country, lab)
```

```{r}
# try to fix this from the lab data
fixed_ratings <- ratings %>%
  mutate(region = case_when(
    # user_id == 545816 ~ "South America",
    # user_id == 545830 ~ "Western Europe",
    TRUE ~ region
  )) %>%
  mutate(region = ifelse(is.na(region), "None", region))
```

### Remove excluded data and average ratings

```{r}
data <- fixed_ratings %>%
  group_by(user_id, trait) %>%
  filter(
    # did not complete 1+ ratings for each of 120 stimuli
    dplyr::n_distinct(stim_id) == 120,
    !is.na(region)   # did not specify region (none expected)
  ) %>%
  anti_join(inv_participants, by = "user_id") %>% # exclude Ps with low
variance
  ungroup() %>%
  group_by(user_id, age, sex, ethnicity, language, lab, country, region,
trait, stim_id) %>%
  summarise(rating = mean(rating)) %>% # average ratings across 2
  ungroup()
```

```

```{r, echo = F}
## Duplicate data for each region - take out for real analysis!

uk <- data %>%
  mutate(region = "United Kingdom")

africa <- data %>%
  mutate(
    region = "Africa",
    user_id = user_id+0.1,
    rating = rating + sample(-1:1, nrow(.), replace = T, prob = c(.2, .6, .
2))
  )

asia <- data %>%
  mutate(
    region = "Asia",
    user_id = user_id+0.2,
    rating = rating + sample(-1:1, nrow(.), replace = T, prob = c(.2, .6, .
2))
  )

me <- data %>%
  mutate(
    region = "Middle East",
    user_id = user_id+0.3,
    rating = rating + sample(-1:1, nrow(.), replace = T, prob = c(.2, .6, .
2))
  )

cam <- data %>%
  mutate(
    region = "Mexico & Central America",
    user_id = user_id+0.4,
    rating = rating + sample(-1:1, nrow(.), replace = T, prob = c(.2, .6, .
2))
  )

usa <- data %>%
  mutate(
    region = "USA and Canada",
    user_id = user_id+0.5,
    rating = rating + sample(-1:1, nrow(.), replace = T, prob = c(.2, .6, .
2))
  )

ee <- data %>%

```
  mutate(
    region = "Eastern Europe",
    user_id = user_id+0.6,
    rating = rating + sample(-1:1, nrow(.), replace = T, prob = c(.2, .6, .
2))
  )

we <- data %>%
  mutate(
    region = "Western Europe",
    user_id = user_id+0.7,
    rating = rating + sample(-1:1, nrow(.), replace = T, prob = c(.2, .6, .
2))
  )

oz <- data %>%
  mutate(
    region = "Australia and New Zealand",
    user_id = user_id+0.8,
    rating = rating + sample(-1:1, nrow(.), replace = T, prob = c(.2, .6, .
2))
  )

scand <- data %>%
  mutate(
    region = "Scandinavia",
    user_id = user_id+0.9,
    rating = rating + sample(-1:1, nrow(.), replace = T, prob = c(.2, .6, .
2))
  )

sa <- data %>%
  mutate(
    region = "South America",
    user_id = user_id+0.95,
    rating = rating + sample(-1:1, nrow(.), replace = T, prob = c(.2, .6, .
2))
  )

data <- uk %>%
  bind_rows(africa) %>%
  bind_rows(asia) %>%
  bind_rows(me) %>%
  bind_rows(cam) %>%
  bind_rows(usa) %>%
  bind_rows(ee) %>%
  bind_rows(we) %>%
  bind_rows(oz) %>%
  bind_rows(scand) %>%
  bind_rows(sa) %>%
```

```
  mutate(rating = rating %>% pmax.int(1) %>% pmin.int(9))
```

## Participant Demographics

### Age and sex distribution per region
```{r, fig.width=12, fig.height=5}
data %>%
  group_by(user_id, sex, age, region) %>%
  summarise() %>%
  ungroup() %>%
  group_by(region) %>%
  mutate(n = n()) %>%
  ungroup() %>%
  ggplot(aes(age, fill = sex)) +
  geom_histogram(binwidth = 5, color = "grey") +
  geom_text(aes(x=0, y=5, label = paste0("n=",n)), color = "black") +
  labs(title="", y="", x="Participant age in 5-year bins") +
  facet_grid(region~.) +
  PSA_theme
```

### Participants per trait per region

```{r, fig.width=12, fig.height=5}
data %>%
  group_by(trait, region) %>%
  summarise(n = n_distinct(user_id)) %>%
  ggplot(aes(trait, n)) +
  geom_col(aes(fill = trait), show.legend = F) +
  geom_hline(yintercept = 15) +
  facet_grid(region~., scale = "free") +
  labs(title="", x="", y="Participants per trait per region") +
  theme( axis.text.x = element_text(angle = -45, hjust = 0) ) +
  PSA_theme
```

# Analyses

## Main Analysis

First, we will calculate the average rating for each face separately for each of the 13 traits. Like Oosterhof and Todorov (2008), we will then subject these mean ratings to principal component analysis with orthogonal components and no rotation. Using the criteria reported in Oosterhof and Todorov's (2008) paper, we will retain and interpret the components with an Eigenvalue > 1.

````
Calculate Alphas

```{r}

data_alpha <- data %>%
  select(user_id, region, stim_id, rating, trait) %>%
  spread(stim_id, rating, sep = "_") %>%
  group_by(trait, region) %>%
  nest(.key = d) %>%
  mutate(alpha = map(d, function(d) {
    if (dim(d)[1] > 2) {
      # calculate cronbach's alpha
      subdata <- d %>%
        as_tibble() %>%
        select(-user_id) %>%
        t()

      capture.output(suppressWarnings(a <- psych::alpha(subdata)))
      a$total["std.alpha"] %>% pluck(1) %>% round(3)
    } else {
      NA
    }
  })) %>%
  select(-d) %>%
  unnest(alpha)
```

```{r, warning=F, fig.width=15, fig.height=4}
data_alpha %>%
  mutate(
    trait = as.factor(trait),
    trait = factor(trait, levels = rev(levels(trait)))
  ) %>%
  filter(!is.na(alpha)) %>%
  mutate(region = str_replace(region, " (and|&) ", " &\n")) %>%
  ggplot() +
  geom_tile(aes(region, trait, fill=alpha >=.7), show.legend = F) +
  geom_text(aes(region, trait, label=alpha), color = "black") +
  scale_y_discrete(drop=FALSE) +
  scale_x_discrete(position = "top") +
  labs(x="", y="", title="Cronbach's Alphas") +
  PSA_theme

```


Calculate Aggregate Scores

```{r}
data_agg <- data %>%
````
````

````
  group_by(region, trait, stim_id) %>%
  summarise(rating = mean(rating)) %>%
  ungroup() %>%
  spread(trait, rating)
```

```{r, fig.width=15, fig.height = 5}
data_agg %>%
  gather("trait", "rating", aggressive:weird) %>%
  ggplot(aes(rating, fill = trait)) +
  geom_density(show.legend = F) +
  facet_grid(region~trait) +
  PSA_theme
```


PCA

```{r}

pca_analyses <- data_agg %>%
  group_by(region) %>%
  nest(.key = "d") %>%
  mutate(pca = map(d, function(d) {
    traits <- select(d, -stim_id) %>%
      select_if(colSums(!is.na(.)) > 0) # omits missing traits

    # principal components analysis (SPSS-style, following Oosterhof &
Todorov)
    ev <- eigen(cor(traits))$values
    nfactors <- sum(ev > 1)

    pca <- principal(
      traits,
      nfactors=nfactors,
      rotate="none"
    )

    stats <- pca$Vaccounted %>%
      as.data.frame() %>%
      rownames_to_column() %>%
      mutate(type = "stat")

    unclass(pca$loadings) %>%
      as.data.frame() %>%
      rownames_to_column() %>%
      mutate(type = "trait") %>%
      bind_rows(stats) %>%
      gather("pc", "loading", 2:(ncol(.)-1))
  })) %>%
````

````
  select(-d) %>%
  unnest(pca)

```

#### Number of PCs (and proportion variance) by region
```{r}
pca_analyses %>%
  filter(rowname == "Proportion Var") %>%
  group_by(region) %>%
  mutate(nPCs = n()) %>%
  ungroup() %>%
  spread(pc, loading) %>%
  select(-rowname, -type) %>%
  mutate_if(is.numeric, round, 3) %>%
  knitr::kable("html") %>%
  kable_styling("striped")
```

#### Trait Loadings by Region and PC
```{r, fig.height=10, fig.width=10}

# order traits by P1 loading if loads positively on P1, or by -P2 loading
otherwise
trait_order <- ot2008_pca_loadings %>%
  rownames_to_column() %>%
  arrange(ifelse(P1>0,P1,-P2)) %>%
  pull(rowname)

pca_analyses %>%
  filter(type == "trait") %>%
  select(-type) %>%
  mutate(
    trait = as.factor(rowname),
    trait = factor(trait, levels = trait_order),
    loading = round(loading, 2)
  ) %>%
  ggplot() +
  geom_tile(aes(pc, trait, fill=loading), show.legend = F) +
  geom_text(aes(pc, trait, label=loading), color = "white") +
  scale_y_discrete(drop=FALSE) +
  scale_x_discrete(position = "top") +
  scale_fill_viridis() +
  facet_wrap(~region, scales = "fixed", ncol = 3) +
  labs(x = "", y = "", title="Trait Loadings") +
  PSA_theme
```

#### PCA Replication criteria
````
````

Oosterhof and Todorov’s valence-dominance model will be judged to have been replicated in a given world region if the first two components both have Eigenvalues > 1, the first component (i.e., the one explaining more of the variance in ratings) is correlated strongly (loading > .7) with trustworthiness and weakly (loading < .5) with dominance, and the second component (i.e., the one explaining less of the variance in ratings) is correlated strongly (loading > .7) with dominance and weakly (loading < .5) with trustworthiness. All three criteria need to be met to conclude that the model was replicated in a given world region.

````
```{r}

pca_analyses %>%
  filter(
    type == "trait",
    rowname %in% c("trustworthy", "dominant"),
    pc %in% c("PC1", "PC2")
  ) %>%
  select(-type) %>%
  mutate(rowname = paste(pc, rowname)) %>%
  select(-pc) %>%
  spread(rowname, loading) %>%
  mutate(replicated = ifelse(
    `PC1 dominant` < .5 & `PC1 trustworthy` > .7 &
    `PC2 dominant` > .7 & `PC2 trustworthy` < .5,
    TRUE, FALSE
  )) %>%
  mutate_if(is.numeric, round, 2) %>%
  knitr::kable("html") %>%
  kable_styling("striped")

```
````

### Factor Congruence

````
```{r region-fc-pca}
fc_pca <- pca_analyses %>%
  filter(type == "trait") %>%
  select(-type) %>%
  spread(pc, loading) %>%
  group_by(region) %>%
  nest(.key = "d") %>%
  mutate(fc = map(d, function(d) {
    loadings <- d %>%
      as.data.frame() %>%
      select(rowname, PC1, PC2) %>%
      arrange(rowname) %>%
      column_to_rownames()

    psych::factor.congruence(loadings, ot2008_pca_loadings) %>%
      as.data.frame() %>%
````
````

```
      rownames_to_column(var = "regionPC")
  })) %>%
  select(-d) %>%
  unnest(fc)

fc_pca %>%
  gather(origPC, congruence, P1:P2) %>%
  mutate(regionPC = paste0("region", regionPC),
         origPC = paste0("orig", origPC),
         congruence = case_when(
           congruence < .85 ~ paste(congruence, " "),
           congruence < .95 ~ paste(congruence, " *"),
           congruence > .95 ~ paste(congruence, " **")
         )) %>%
  unite(PC, origPC, regionPC, remove = T) %>%
  spread(PC, congruence) %>%
  knitr::kable("html") %>%
  kable_styling("striped")
```

factor congruence < .85 = not similar
* factor congruence >= .85 and < .95 = fairly similar
** factor congruence > .95 = equal

## Robustness Checks

### Exploratory Factor Analysis (EFA)

```{r efa, message=FALSE, warning=FALSE, results="hide"}

efa_analyses <- data_agg %>%
  group_by(region) %>%
  nest(.key = "d") %>%
  mutate(efa = map(d, function(d) {
    traits <- select(d, -stim_id) %>%
      select_if(colSums(!is.na(.)) > 0) # omits missing traits

    # Parallel Analysis with Dino's 'paran' package.
    nfactors <- paran(traits, iterations = 5000,
                      centile = 0, quietly = TRUE,
                      status = FALSE, all = TRUE,
                      cfa = TRUE, graph = FALSE)

    efa <- psych::fa(traits, nfactors$Retained)

    stats <- efa$Vaccounted %>%
      as.data.frame() %>%
      rownames_to_column() %>%
      mutate(type = "stat")
```

```
    unclass(efa$loadings) %>%
      as.data.frame() %>%
      rownames_to_column() %>%
      mutate(type = "trait") %>%
      bind_rows(stats) %>%
      gather("mr", "loading", 2:(ncol(.)-1))
  })) %>%
  select(-d) %>%
  unnest(efa)

```

### Number of MRs (and proportion variance) by region
```{r}
efa_analyses %>%
  filter(rowname == "Proportion Var") %>%
  group_by(region) %>%
  mutate(nMRs = n()) %>%
  ungroup() %>%
  spread(mr, loading) %>%
  select(-rowname, -type) %>%
  mutate_if(is.numeric, round, 3) %>%
  knitr::kable("html") %>%
  kable_styling("striped")
```

### Trait Loadings by Region and MR
```{r, fig.height=10, fig.width=10}

# order traits by P1 loading if loads positively on P1, or by -P2 loading
otherwise
trait_order <- ot2008_efa_loadings %>%
  rownames_to_column() %>%
  arrange(ifelse(P1>0,P1,-P2)) %>%
  pull(rowname)

efa_analyses %>%
  filter(type == "trait") %>%
  select(-type) %>%
  mutate(
    trait = as.factor(rowname),
    trait = factor(trait, levels = trait_order),
    loading = round(loading, 2)
  ) %>%
  ggplot() +
  geom_tile(aes(mr, trait, fill=loading), show.legend = F) +
  geom_text(aes(mr, trait, label=loading), color = "white") +
  scale_y_discrete(drop=FALSE) +
  scale_x_discrete(position = "top") +
```

````
  scale_fill_viridis() +
  facet_wrap(~region, scales = "fixed", ncol = 3) +
  labs(x = "", y = "", title="Trait Loadings") +
  PSA_theme
```

### Replication criteria

```{r efa-replication-criteria}

efa_analyses %>%
  filter(
    type == "trait",
    rowname %in% c("trustworthy", "dominant"),
    mr %in% c("MR1", "MR2")
  ) %>%
  select(-type) %>%
  mutate(rowname = paste(mr, rowname)) %>%
  select(-mr) %>%
  spread(rowname, loading) %>%
  mutate(replicated = ifelse(
    `MR1 dominant` < .5 & `MR1 trustworthy` > .7 &
    `MR2 dominant` > .7 & `MR2 trustworthy` < .5,
    TRUE, FALSE
  )) %>%
  mutate_if(is.numeric, round, 2) %>%
  knitr::kable("html") %>%
  kable_styling("striped")

```

Factor Congruence
```{r fc-efa}

# order traits by P1 loading if loads positively on P1, or by -P2 loading
otherwise
trait_order <- ot2008_efa_loadings %>%
  rownames_to_column() %>%
  arrange(ifelse(P1>0,P1,-P2)) %>%
  pull(rowname)

fc_efa <- efa_analyses %>%
  filter(type == "trait") %>%
  select(-type) %>%
  spread(mr, loading) %>%
  group_by(region) %>%
  nest(.key = "d") %>%
  mutate(fc = map(d, function(d) {
    loadings <- d %>%
      as.data.frame() %>%
````

````
      select(rowname, MR1, MR2) %>%
      arrange(rowname) %>%
      column_to_rownames()

    psych::factor.congruence(loadings, ot2008_efa_loadings) %>%
  as.data.frame() %>%
  rownames_to_column(var = "regionMR")
  })) %>%
  select(-d) %>%
  unnest(fc)

fc_efa %>%
  gather(origPC, congruence, P1:P2) %>%
  mutate(regionMR = paste0("region", regionMR),
         origPC = paste0("orig", origPC),
         congruence = case_when(
           congruence < .85 ~ paste(congruence, " "),
           congruence < .95 ~ paste(congruence, " *"),
           congruence > .95 ~ paste(congruence, " **")
         )) %>%
  unite(PC, origPC, regionMR, remove = T) %>%
  spread(PC, congruence) %>%
  knitr::kable("html") %>%
  kable_styling("striped")
```

factor congruence < .85 = not similar
* factor congruence >= .85 and < .95 = fairly similar
** factor congruence > .95 = equal
````
````

Code

# PSA1 Factor Analysis Power Simulation

*Lisa DeBruine*



# Setup

## Load Libraries

Hide

```
library(MASS) # for simulating multivariate normal distributions
library(psych) # for SPSS-style PCA
library(viridis) # for nice colours
```

```
Loading required package: viridisLite
```

Hide

```
library(tidyverse) # for data cleaning
```

```
[37m── [1mAttaching packages[22m ─────────────────────────────────── tidyverse 1.2.1 ──[39m
[37m[32m✔[37m [34mggplot2[37m 3.0.0     [32m✔[37m [34mpurrr  [37m 0.2.5
[32m✔[37m [34mtibble [37m 1.4.2     [32m✔[37m [34mdplyr  [37m 0.7.6
[32m✔[37m [34mtidyr  [37m 0.8.1     [32m✔[37m [34mstringr[37m 1.3.1
[32m✔[37m [34mreadr  [37m 1.1.1     [32m✔[37m [34mforcats[37m 0.3.0[39m
[37m── [1mConflicts[22m ─────────────────────────────────────── tidyverse_conflicts() ──
[31m✖[37m [34mggplot2[37m::[32m%+%()[37m   masks [34mpsych[37m::%+%()
[31m✖[37m [34mggplot2[37m::[32malpha()[37m masks [34mpsych[37m::alpha()
[31m✖[37m [34mdplyr[37m::[32mfilter()[37m  masks [34mstats[37m::filter()
[31m✖[37m [34mdplyr[37m::[32mlag()[37m     masks [34mstats[37m::lag()
[31m✖[37m [34mdplyr[37m::[32mselect()[37m  masks [34mMASS[37m::select()[39m
```

# Load Data (from OSF)

Hide

```
data <- rbind(
    read_csv("https://osf.io/6sz8k/download") %>% # male ratings
      gather(stim_id, rating, andrej:vladislav),
    read_csv("https://osf.io/375ag/download") %>% # female ratings
      gather(stim_id, rating, alexandra:zlata)
  ) %>%
  filter(type == "faces") %>%
  rename(
    rater_id = user_id,
    stim_sex = stimulus_sex,
    trait = judgment
  ) %>%
  # reverse code happy to unhappy
  mutate(
    rating = ifelse(trait == "happy", 7 - rating, rating),
    trait = ifelse(trait == "happy", "unhappy", trait),
    trait = ifelse(trait == "emotionally_stable", "emostable", trait)
  )
```

```
Parsed with column specification:
cols(
  .default = col_integer(),
  judgment = col_character(),
  stimulus_sex = col_character(),
  type = col_character(),
  rater_sex = col_character(),
  sexpref = col_character(),
  rater_age = col_double()
)
See spec(...) for full column specifications.
Parsed with column specification:
cols(
  .default = col_integer(),
  judgment = col_character(),
  stimulus_sex = col_character(),
  type = col_character(),
  rater_sex = col_character(),
  sexpref = col_character(),
  rater_age = col_double()
)
See spec(...) for full column specifications.
```

# Visualise Data

Hide

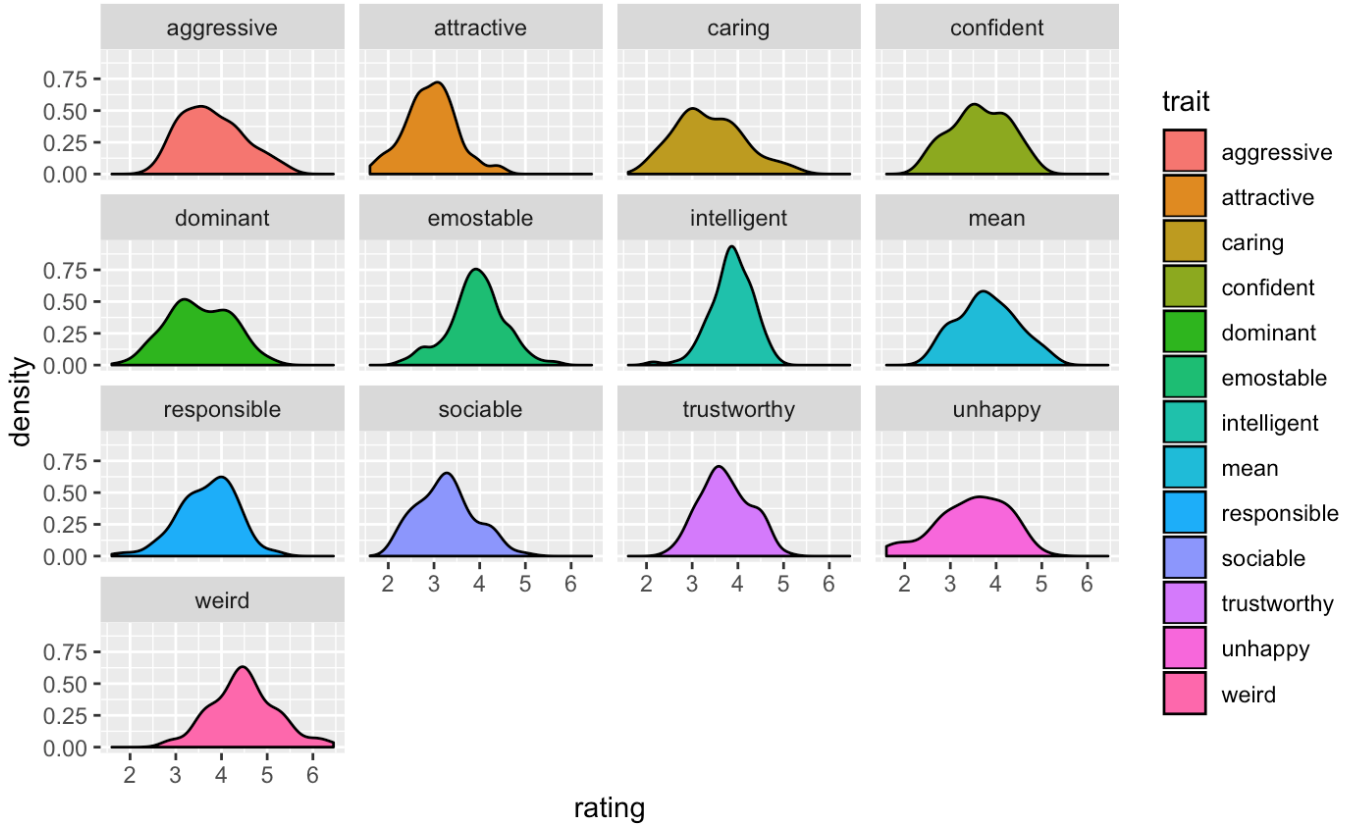
```
ggplot(data, aes(rating, fill = trait)) +
  geom_histogram(binwidth = 1, color = "grey") +
  facet_wrap(~trait)
```


# Data Processing

## Aggregate Ratings

Hide

```
data_agg <- data %>%
  group_by(stim_id, trait) %>%
  summarise(rating = mean(rating)) %>%
  ungroup() %>%
  spread(trait, rating)
data_agg %>%
  gather("trait", "rating", aggressive:weird) %>%
  ggplot(aes(rating, fill = trait)) +
  geom_density() +
  facet_wrap(~trait)
```




# Calculate correlations

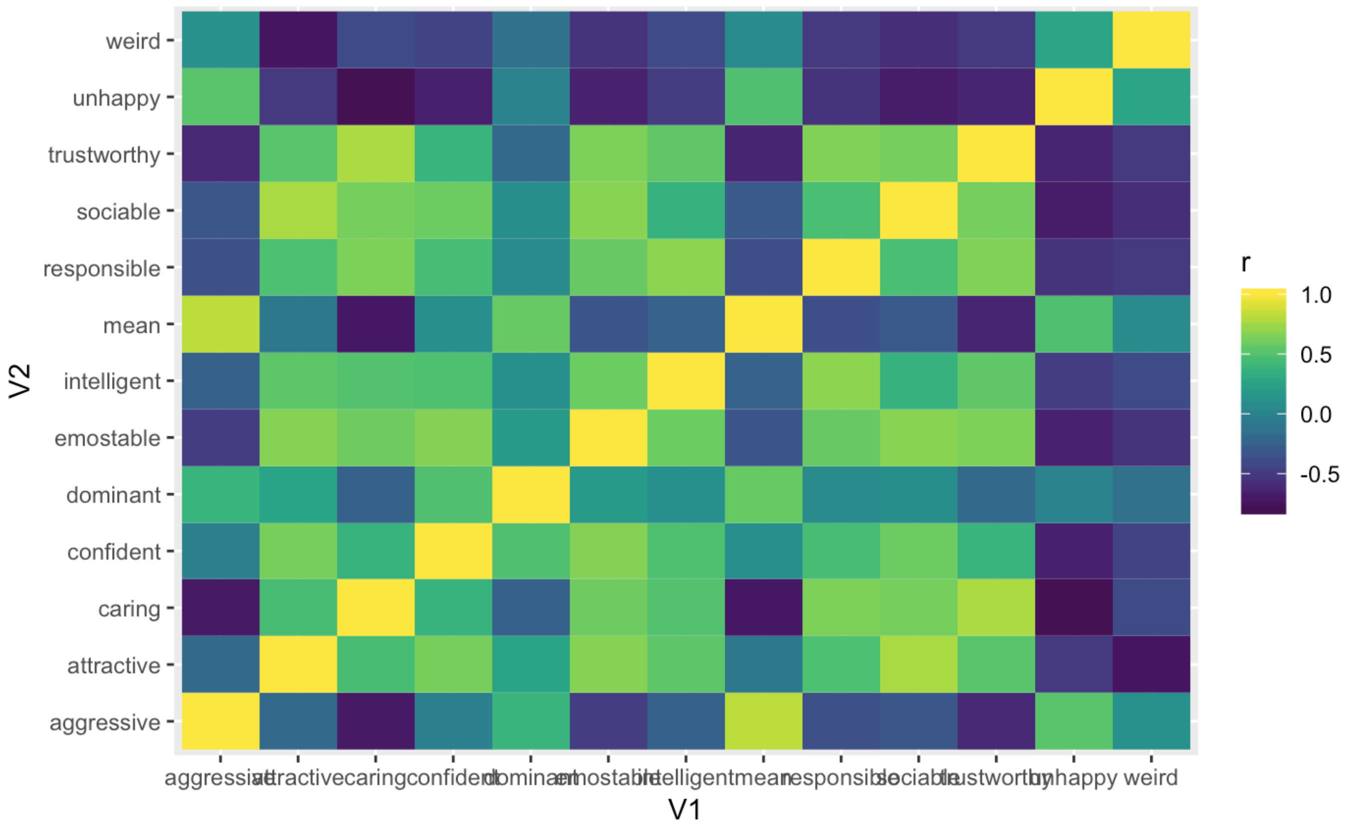
Hide

```
data_cor <- data_agg %>%
  select(-stim_id) %>%
  cor()
```

# Visualise Correlations

Hide

```
data_cor %>%
  as.data.frame() %>% # make it a data frame
  rownames_to_column(var = "V1") %>% # set rownames as V1
  gather("V2", "r", aggressive:weird) %>%
  ggplot(aes(V1, V2, fill=r)) +
  geom_tile() +
  scale_fill_viridis()
```




# Analysis

## How many PCs?

Hide

```
get_nfactors <- function(data, method="broken-stick", total.var = .95) {
  ev <- eigen(cor(data))$values
  if (method == "broken-stick") {
    # compare variance explained to null model
    n.ev <- length(ev)
    bsm <- data.frame(j=seq(1:n.ev), p=0)

    bsm$p[1] <- 1/n.ev
    for (i in 2:n.ev) {
      bsm$p[i] <- bsm$p[i-1] + (1/(n.ev + 1 - i))
    }
    bsm$p <- 100*bsm$p/n.ev

    my_ev <- 100*ev/sum(ev)
    null_ev <- bsm$p[n.ev:1]

    n_factors <- sum(my_ev >= null_ev)
  } else if (method == "Kaiser-Guttman") {
    # return PCs with eigenvalues greater than the mean eigenvalue
    n_factors <- sum(ev >= mean(ev));
  } else if (method == "total variance") {
    # return PCs explaining at least total.var variance
    cumvar <- cumsum(ev / sum(ev));
    n_factors <- sum(cumvar < total.var) + 1;
  } else if (method == "SPSS") {
    n_factors <- sum(ev > 1)
  }
  n_factors
}
```

# PCA

Hide

```
canon_traits <- data_agg %>% select(-stim_id)
nfactors <- get_nfactors(canon_traits, method = "SPSS");
# principal components analysis (SPSS-style, following Oosterhof & Todorov)
canon_pca <- principal(
  canon_traits,
  nfactors=nfactors,
  rotate="none"
)
canon_pca
```

```
Principal Components Analysis
Call: principal(r = canon_traits, nfactors = nfactors, rotate = "none")
Standardized loadings (pattern matrix) based upon correlation matrix
```

| | PC1 <S3: AsIs> | PC2 <S3: AsIs> | h2 <dbl> | u2 <dbl> | com <dbl> |
|---|---|---|---|---|---|
| aggressive | -0.60 | 0.65 | 0.7715944 | 0.2284056 | 1.987797 |

| | | | | | |
|---|---|---|---|---|---|
| aggressive | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] |
| attractive | 0.75 | 0.43 | 0.7487249 | 0.2512751 | 1.601561 |
| caring | 0.86 | -0.34 | 0.8546064 | 0.1453936 | 1.314196 |
| confident | 0.65 | 0.59 | 0.7610324 | 0.2389676 | 1.980491 |
| dominant | 0.00 | 0.82 | 0.6762027 | 0.3237973 | 1.000014 |
| emostable | 0.85 | 0.18 | 0.7469544 | 0.2530456 | 1.087099 |
| intelligent | 0.69 | 0.16 | 0.5022182 | 0.4977818 | 1.102179 |
| mean | -0.55 | 0.77 | 0.8988391 | 0.1011609 | 1.800288 |
| responsible | 0.77 | 0.04 | 0.5902089 | 0.4097911 | 1.004184 |
| sociable | 0.80 | 0.19 | 0.6754514 | 0.3245486 | 1.109518 |

1-10 of 13 rows  Previous **1** 2 Next

```
                       PC1  PC2
SS loadings           6.59 2.61
Proportion Var        0.51 0.20
Cumulative Var        0.51 0.71
Proportion Explained  0.72 0.28
Cumulative Proportion 0.72 1.00

Mean item complexity =  1.4
Test of the hypothesis that 2 components are sufficient.

The root mean square of the residuals (RMSR) is  0.07
 with the empirical chi square  83.05  with prob <  0.0052

Fit based upon off diagonal values = 0.98
```

# Get loadings from PCA output

Hide

```
canon_loadings <- unclass(canon_pca$loadings)
```

# Simulation

## Simulate multivariate distribution

Hide

```
# function to generate n datasets from rating data (only rating columns)
sim_agg_ratings <- function(data, n) {
  mu <- data %>% # the means of each rating
    summarise_all("mean") %>%
    t() %>%
    as.vector()

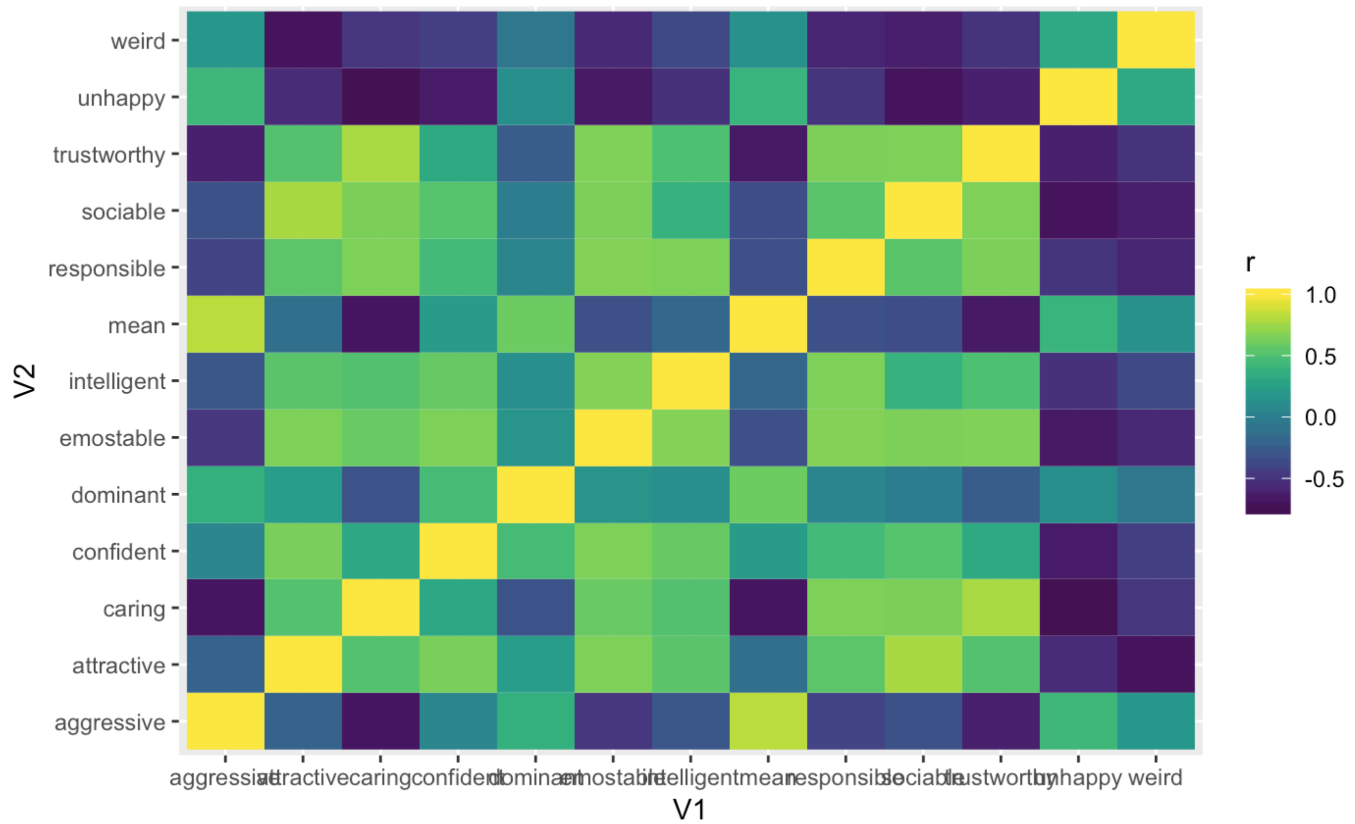
  stdevs <- data %>% # the SDs of each rating
    summarise_all("sd") %>%
    t() %>%
    as.vector()

  cor_mat <- cor(data)
  sigma <- (stdevs %*% t(stdevs)) * cor_mat

  mvrnorm(n, mu, sigma)
}
```

Hide

```
# test the function
set.seed(900)
data_test_sim <- data_agg %>%
  select(-stim_id) %>%
  sim_agg_ratings(120)
data_test_sim %>%
  as.data.frame() %>%
  write_csv("data_test_sim.csv")
cor(data_test_sim) %>%
  as.data.frame() %>% # make it a data frame
  rownames_to_column(var = "V1") %>% # set rownames as V1
  gather("V2", "r", aggressive:weird) %>%
  ggplot(aes(V1, V2, fill=r)) +
  geom_tile() +
  scale_fill_viridis()
```




# PCA Function

Hide

```
# function to generate PCA stats from data
sim_pca <- function(data) {
  nfactors <- get_nfactors(data, method = "SPSS");

  # principal components analysis (SPSS-style, following Oosterhof & Todorov)
  sim_pca <- principal(
    data,
    nfactors=nfactors,
    rotate="none"
  )
  unclass(sim_pca$loadings)
}
```

Hide

```
test_sim_loadings <- sim_pca(data_test_sim)
test_sim_loadings
```

```
                    PC1          PC2
aggressive  -0.58948856  0.6267395519
attractive   0.76944683  0.3890716128
caring       0.85830517 -0.3275039871
confident    0.60423390  0.6703399391
dominant    -0.06492687  0.8157540486
emostable    0.85873558  0.1886084742
intelligent  0.69812732  0.2418114309
mean        -0.52390652  0.7941215592
responsible  0.79294000  0.0776241780
sociable     0.81945369  0.1076499987
trustworthy  0.84507449 -0.2760149360
unhappy     -0.80335730  0.0004462812
weird       -0.67395813 -0.2495973655
```

Difference between canon and first two PC sim loadings

Hide

```
(canon_loadings[,1:2] - test_sim_loadings[,1:2])
```

```
                      PC1          PC2
aggressive  -0.0068073983  0.018260866
attractive  -0.0203893357  0.044100110
caring      -0.0004115545 -0.016915663
confident    0.0425140329 -0.084891847
dominant     0.0671147324  0.006558515
emostable   -0.0127298094 -0.011892141
intelligent -0.0069353147 -0.085376734
mean        -0.0235720025 -0.020101919
responsible -0.0254918400 -0.042521916
sociable    -0.0192759357  0.079878601
trustworthy  0.0070741730  0.043611267
unhappy     -0.0218835497  0.068079730
weird        0.0463622180 -0.099204821
```

# Replicate

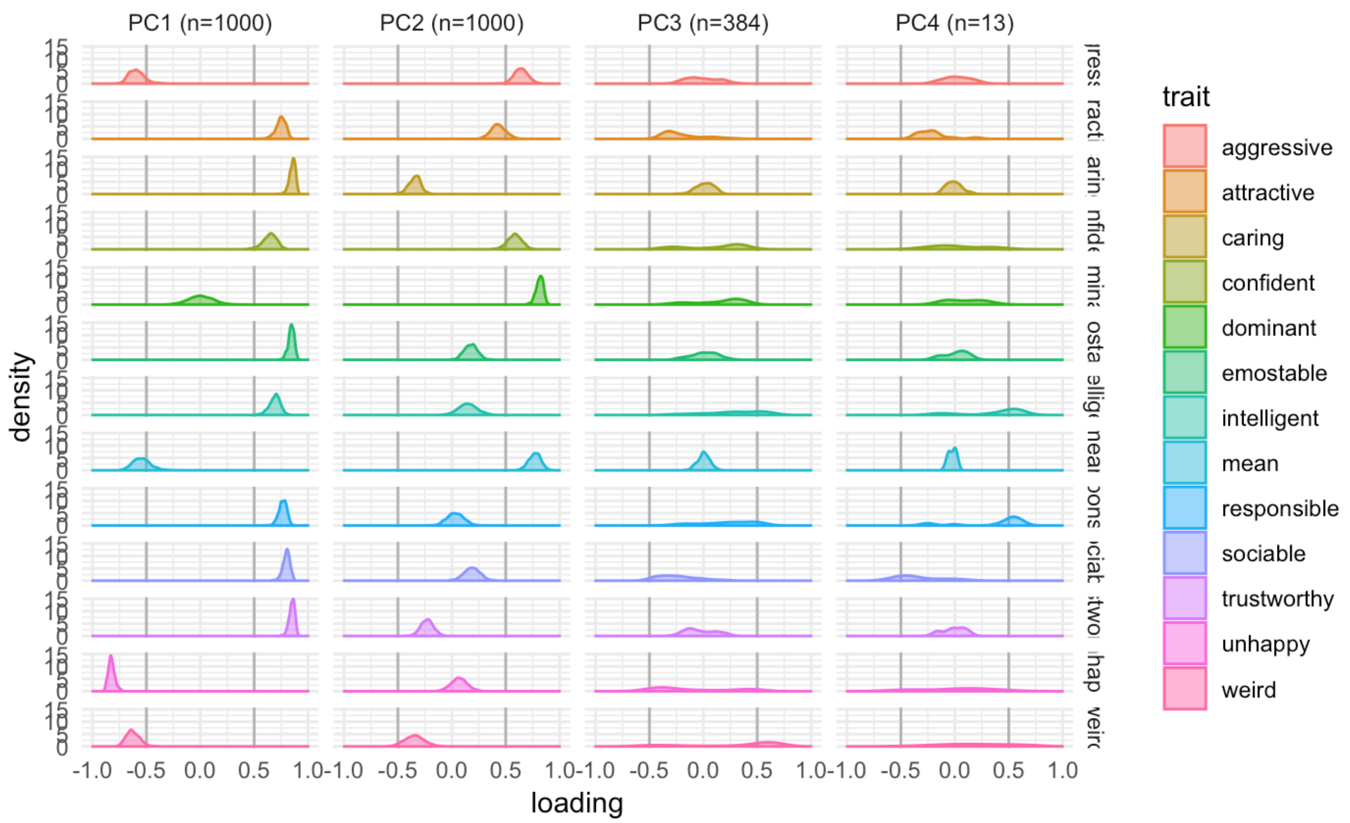
Hide

```
reps <- 1:1000 %>%
  purrr::map_df(function(x) {
  data_agg %>%
    select(-stim_id) %>%
    sim_agg_ratings(120)  %>%
    sim_pca()%>%
    as.data.frame() %>%
    rownames_to_column(var = "trait") %>%
    gather("PC", "loading", 2:ncol(.)) %>%
    mutate(rep = x)
})
```

# Plot loading distributions

Hide

```
reps %>%
  group_by(PC) %>%
  mutate(PC_n = paste0(PC, " (n=",n_distinct(rep), ")")) %>%
  ungroup() %>%
  ggplot() +
  geom_vline(xintercept = 0.5, color = "grey70") +
  geom_vline(xintercept = -0.5, color = "grey70") +
  geom_density(aes(loading, color = trait, fill = trait), alpha = 0.5) +
  xlim(-1,1) +
  facet_grid(trait~PC_n) +
  theme_minimal()
ggsave("loading_dist.png", width = 10, height = 12)
```




# What percent of sims meet each criterion?

Hide

```
strong_loading <- .7
weak_loading <- .5
meet_criteria <- reps %>%
  filter(PC %in% c("PC1", "PC2")) %>%
  spread(trait, loading) %>%
  mutate(
    trustworthy__c = ifelse(PC == "PC1",
      trustworthy > strong_loading,
      trustworthy < weak_loading),
   dominant__c = ifelse(PC == "PC1",
      abs(dominant) < weak_loading,
      dominant > strong_loading)
  ) %>%
  select(PC, rep, trustworthy__c, dominant__c) %>%
  gather("key", "val", trustworthy__c:dominant__c) %>%
  separate(key, c("trait", "type"), sep="__") %>%
  mutate(type = ifelse(is.na(type), "loading", "criterion")) %>%
  spread(type, val)
```

Hide

```
meet_criteria %>%
  unite(PCtrait, PC, trait) %>%
  spread(PCtrait, criterion) %>%
  count(PC1_dominant, PC1_trustworthy, PC2_dominant, PC2_trustworthy)
```

| PC1_dominant <lgl> | PC1_trustworthy <lgl> | PC2_dominant <lgl> | PC2_trustworthy <lgl> | n <int> |
|---|---|---|---|---|
| TRUE | TRUE | FALSE | TRUE | 3 |
| TRUE | TRUE | TRUE | TRUE | 997 |

2 rows

# Test correlations with trust and dom

Hide

```
sim_pca_trudom <- function(data) {
  data2 <- data %>%
    as_tibble() %>%
    #select(-dominant, - trustworthy) %>%
    as.matrix()

  nfactors <- get_nfactors(data2, method = "SPSS");

  # principal components analysis (SPSS-style, following Oosterhof & Todorov)
  sim_pca <- principal(
    data2,
    nfactors=nfactors,
    rotate="none",
    scores = T
  )

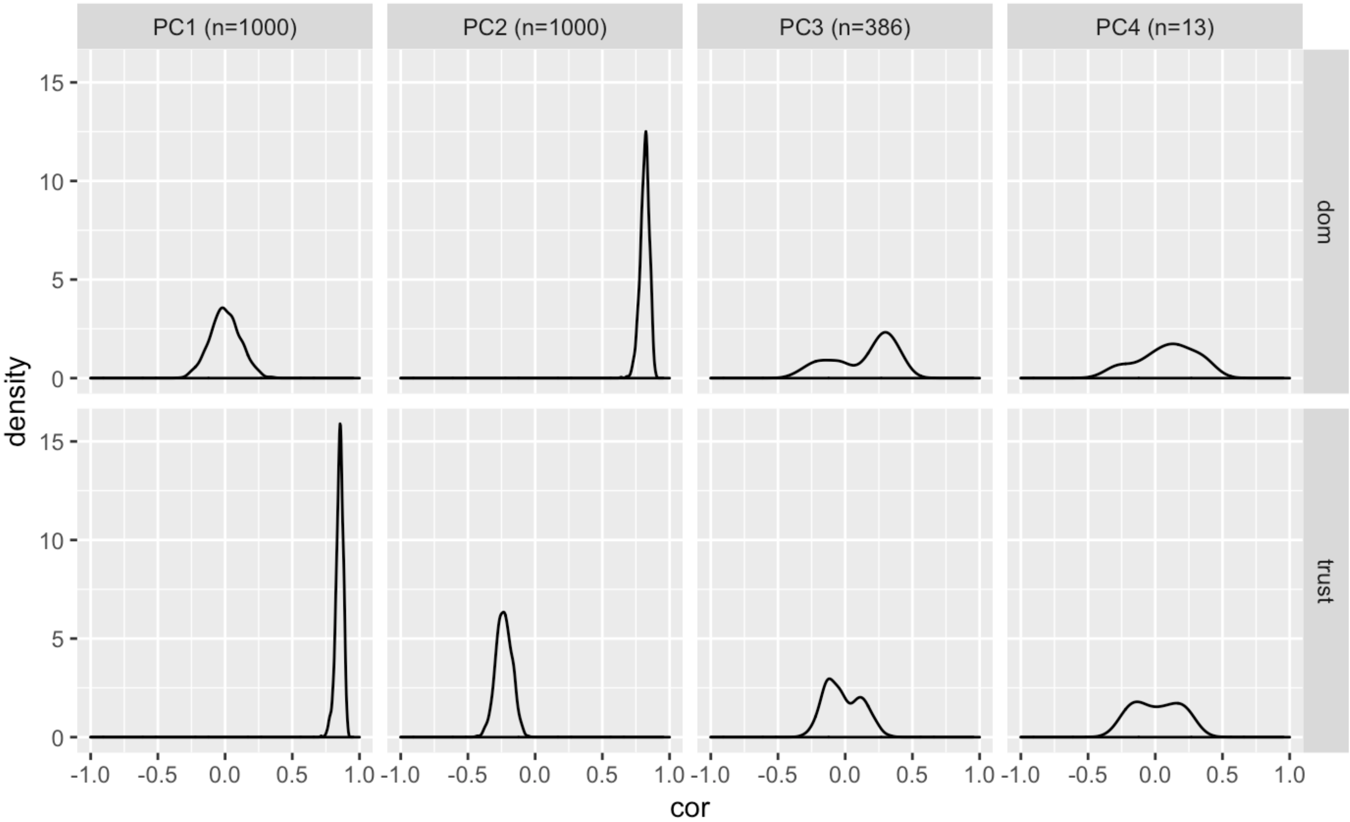
  sim_pca$scores %>%
    as.data.frame() %>%
    mutate(
      dom = as.data.frame(data)$dominant,
      trust = as.data.frame(data)$trustworthy
    )
}
```

Hide

```
reps2 <- 1:1000 %>%
  purrr::map_df(function(x) {
  data_agg %>%
    select(-stim_id) %>%
    sim_agg_ratings(120) %>%
    sim_pca_trudom() %>%
    cor() %>%
    as.data.frame() %>%
    rownames_to_column(var = "PC") %>%
    select(PC, dom, trust) %>%
    filter(PC != "dom", PC != "trust") %>%
    mutate(rep = x)
})
```

Hide

```
reps2 %>%
  group_by(PC) %>%
  mutate(PC_n = paste0(PC, " (n=",n_distinct(rep), ")")) %>%
  ungroup() %>%
  gather("trait", "cor", dom:trust) %>%
  ggplot(aes(cor)) +
  geom_density() +
  xlim(-1, 1) +
  facet_grid(trait~PC_n)
```


PC1 (n=1000)
PC2 (n=1000)
PC3 (n=386)
PC4 (n=13)
dom
trust
density
cor
0
5
10
15
-1.0
-0.5
0.0
0.5
1.0

````
---
title: "PSA1 Factor Analysis Power Simulation"
author: "Lisa DeBruine"
output:
  html_notebook:
    toc: true

Setup

Load Libraries
```{r}
library(MASS) # for simulating multivariate normal distributions
library(psych) # for SPSS-style PCA
library(viridis) # for nice colours
library(tidyverse) # for data cleaning
```

Load Data (from OSF)

```{r}

data <- rbind(
    read_csv("https://osf.io/6sz8k/download") %>% # male ratings
      gather(stim_id, rating, andrej:vladislav),
    read_csv("https://osf.io/375ag/download") %>% # female ratings
      gather(stim_id, rating, alexandra:zlata)
  ) %>%
  filter(type == "faces") %>%
  rename(
    rater_id = user_id,
    stim_sex = stimulus_sex,
    trait = judgment
  ) %>%
  # reverse code happy to unhappy
  mutate(
    rating = ifelse(trait == "happy", 7 - rating, rating),
    trait = ifelse(trait == "happy", "unhappy", trait),
    trait = ifelse(trait == "emotionally_stable", "emostable", trait)
  )

```


Visualise Data

```{r}

ggplot(data, aes(rating, fill = trait)) +
  geom_histogram(binwidth = 1, color = "grey") +
````
````

```
  facet_wrap(~trait)
```

```{r, eval=FALSE, echo = FALSE}

## calculate ICC for each
data_icc <- data %>%
  select(trait, stim_sex, rater_id, stim_id, rating) %>%
  group_by(trait, stim_sex) %>%
  nest(.key = "d") %>%
  mutate(icc = map(d, function(d) {
    d2 <- d %>%
      spread(stim_id, rating) %>%
      select(-rater_id) %>%
      t()

    icc(d2, model = "twoway", type = "consistency", unit = "single")$value
    })
  ) %>%
  ungroup() %>%
  select(-d) %>%
  unnest()

```

# Data Processing

## Aggregate Ratings

```{r}
data_agg <- data %>%
  group_by(stim_id, trait) %>%
  summarise(rating = mean(rating)) %>%
  ungroup() %>%
  spread(trait, rating)

data_agg %>%
  gather("trait", "rating", aggressive:weird) %>%
  ggplot(aes(rating, fill = trait)) +
  geom_density() +
  facet_wrap(~trait)
```


## Calculate correlations

```{r}
data_cor <- data_agg %>%
  select(-stim_id) %>%
```

````
  cor()
```

## Visualise Correlations

```{r}
data_cor %>%
  as.data.frame() %>% # make it a data frame
  rownames_to_column(var = "V1") %>% # set rownames as V1
  gather("V2", "r", aggressive:weird) %>%
  ggplot(aes(V1, V2, fill=r)) +
  geom_tile() +
  scale_fill_viridis()
```

Analysis

How many PCs?

```{r}
get_nfactors <- function(data, method="broken-stick", total.var = .95) {
  ev <- eigen(cor(data))$values

  if (method == "broken-stick") {
    # compare variance explained to null model
    n.ev <- length(ev)
    bsm <- data.frame(j=seq(1:n.ev), p=0)

    bsm$p[1] <- 1/n.ev
    for (i in 2:n.ev) {
      bsm$p[i] <- bsm$p[i-1] + (1/(n.ev + 1 - i))
    }
    bsm$p <- 100*bsm$p/n.ev

    my_ev <- 100*ev/sum(ev)
    null_ev <- bsm$p[n.ev:1]

    n_factors <- sum(my_ev >= null_ev)
  } else if (method == "Kaiser-Guttman") {
    # return PCs with eigenvalues greater than the mean eigenvalue
    n_factors <- sum(ev >= mean(ev));
  } else if (method == "total variance") {
    # return PCs explaining at least total.var variance
    cumvar <- cumsum(ev / sum(ev));
    n_factors <- sum(cumvar < total.var) + 1;
  } else if (method == "SPSS") {
    n_factors <- sum(ev > 1)
  }

  n_factors
````

}

```

## PCA

```{r}

canon_traits <- data_agg %>% select(-stim_id)

nfactors <- get_nfactors(canon_traits, method = "SPSS");

# principal components analysis (SPSS-style, following Oosterhof & Todorov)
canon_pca <- principal(
  canon_traits,
  nfactors=nfactors,
  rotate="none"
)

canon_pca

```

Get loadings from PCA output
```{r}
canon_loadings <- unclass(canon_pca$loadings)
```


Simulation

Simulate multivariate distribution

```{r}
# function to generate n datasets from rating data (only rating columns)

sim_agg_ratings <- function(data, n) {
  mu <- data %>% # the means of each rating
    summarise_all("mean") %>%
    t() %>%
    as.vector()

  stdevs <- data %>% # the SDs of each rating
    summarise_all("sd") %>%
    t() %>%
    as.vector()

  cor_mat <- cor(data)

  sigma <- (stdevs %*% t(stdevs)) * cor_mat

````
  mvrnorm(n, mu, sigma)
}
```


```{r}
# test the function
set.seed(900)

data_test_sim <- data_agg %>%
  select(-stim_id) %>%
  sim_agg_ratings(120)

data_test_sim %>%
  as.data.frame() %>%
  write_csv("data_test_sim.csv")

cor(data_test_sim) %>%
  as.data.frame() %>% # make it a data frame
  rownames_to_column(var = "V1") %>% # set rownames as V1
  gather("V2", "r", aggressive:weird) %>%
  ggplot(aes(V1, V2, fill=r)) +
  geom_tile() +
  scale_fill_viridis()

```

PCA Function
```{r}
# function to generate PCA stats from data

sim_pca <- function(data) {
  nfactors <- get_nfactors(data, method = "SPSS");

  # principal components analysis (SPSS-style, following Oosterhof &
Todorov)
  sim_pca <- principal(
    data,
    nfactors=nfactors,
    rotate="none"
  )

  unclass(sim_pca$loadings)
}

```

```{r}
````

````
test_sim_loadings <- sim_pca(data_test_sim)

test_sim_loadings

```


Difference between canon and first two PC sim loadings
```{r}
(canon_loadings[,1:2] - test_sim_loadings[,1:2])
```

## Replicate

```{r}

reps <- 1:1000 %>%
  purrr::map_df(function(x) {
  data_agg %>%
    select(-stim_id) %>%
    sim_agg_ratings(120)  %>%
    sim_pca()%>%
    as.data.frame() %>%
    rownames_to_column(var = "trait") %>%
    gather("PC", "loading", 2:ncol(.)) %>%
    mutate(rep = x)
})

```

## Plot loading distributions

```{r}
reps %>%
  group_by(PC) %>%
  mutate(PC_n = paste0(PC, " (n=",n_distinct(rep), ")")) %>%
  ungroup() %>%
  ggplot() +
  geom_vline(xintercept = 0.5, color = "grey70") +
  geom_vline(xintercept = -0.5, color = "grey70") +
  geom_density(aes(loading, color = trait, fill = trait), alpha = 0.5) +
  xlim(-1,1) +
  facet_grid(trait~PC_n) +
  theme_minimal()

ggsave("loading_dist.png", width = 10, height = 12)
```

## What percent of sims meet each criterion?
````
````

```{r}
strong_loading <- .7
weak_loading <- .5

meet_criteria <- reps %>%
  filter(PC %in% c("PC1", "PC2")) %>%
  spread(trait, loading) %>%
  mutate(
    trustworthy__c = ifelse(PC == "PC1",
      trustworthy > strong_loading,
      trustworthy < weak_loading),
   dominant__c = ifelse(PC == "PC1",
      abs(dominant) < weak_loading,
      dominant > strong_loading)
  ) %>%
  select(PC, rep, trustworthy__c, dominant__c) %>%
  gather("key", "val", trustworthy__c:dominant__c) %>%
  separate(key, c("trait", "type"), sep="__") %>%
  mutate(type = ifelse(is.na(type), "loading", "criterion")) %>%
  spread(type, val)

```

```{r}
meet_criteria %>%
  unite(PCtrait, PC, trait) %>%
  spread(PCtrait, criterion) %>%
  count(PC1_dominant, PC1_trustworthy, PC2_dominant, PC2_trustworthy)
```

## Test correlations with trust and dom

```{r}
sim_pca_trudom <- function(data) {
  data2 <- data %>%
    as_tibble() %>%
    #select(-dominant, - trustworthy) %>%
    as.matrix()

  nfactors <- get_nfactors(data2, method = "SPSS");

  # principal components analysis (SPSS-style, following Oosterhof &
Todorov)
  sim_pca <- principal(
    data2,
    nfactors=nfactors,
    rotate="none",
    scores = T
```

````
)

  sim_pca$scores %>%
    as.data.frame() %>%
    mutate(
      dom = as.data.frame(data)$dominant,
      trust = as.data.frame(data)$trustworthy
)
}
```

```{r}

reps2 <- 1:1000 %>%
  purrr::map_df(function(x) {
  data_agg %>%
    select(-stim_id) %>%
    sim_agg_ratings(120) %>%
    sim_pca_trudom() %>%
    cor() %>%
    as.data.frame() %>%
    rownames_to_column(var = "PC") %>%
    select(PC, dom, trust) %>%
    filter(PC != "dom", PC != "trust") %>%
    mutate(rep = x)
})

```



```{r}
reps2 %>%
  group_by(PC) %>%
  mutate(PC_n = paste0(PC, " (n=",n_distinct(rep), ")")) %>%
  ungroup() %>%
  gather("trait", "cor", dom:trust) %>%
  ggplot(aes(cor)) +
  geom_density() +
  xlim(-1, 1) +
  facet_grid(trait~PC_n)

```
````
````

Determining the number of raters for reliable mean ratings

Code

Lisa DeBruine & Benedict Jones

How many raters you need to sample to get reliable stimulus ratings will obviously depend on the raters, stimuli, and what they're being rated for. For example, if there is a lot of inter-rater variation or very little inter-stimulus variation, you will need more raters to generate mean ratings with any reliability.

If you have a large set of ratings of a type of stimulus, population of rater, and type of rating you're interested in, you can use the script below to figure out how many raters you need to sample to get mean stimulus ratings that are well-correlated with mean ratings from the full sample.

The example below is for attractiveness ratings using an open-access image set from our lab.

Hide

```
library(tidyverse)
library(purrr)
library(psych)
```

Read data from DeBruine, Lisa; Jones, Benedict (2017): Face Research Lab London Set. figshare. doi: 10.6084/m9.figshare.5047666 (https://doi.org/10.6084/m9.figshare.5047666)

Hide

```
data <- read_csv("https://ndownloader.figshare.com/files/8542045")
```

Calculate canonical mean ratings (average of all available ratings)

Hide

```
canon <- data %>%
  select(X001:X173) %>%
  group_by() %>%
  summarise_all(mean) %>%
  t()
```

Sample n raters from the set and calculate Cronbach's `alpha` and `r` from the Pearson's correlation with the canonical ratings.

Hide

```
get_alpha <- function(data, n) {
  # sample your full dataset
  data_sample <- data %>%
    sample_n(n) %>%
    select(X001:X173) # select only columns with your stimuli

  # calculate cronbach's alpha
  capture.output(suppressWarnings(a <- psych::alpha(t(data_sample))))
  alpha <- a$total["std.alpha"] %>% pluck(1)
  # calculate mean sample ratings
  sample_means <- data_sample %>%
    group_by() %>%
    summarise_all(mean) %>%
    t()

  # calculate correlation between sample mean ratings and canon
  r <- cor(sample_means, canon)[[1,1]]

  # return relevant data
  tibble(
    n = n,
    alpha = alpha,
    r = r
  )
}
```

Generate 1000 samples for 5 to 50 raters.

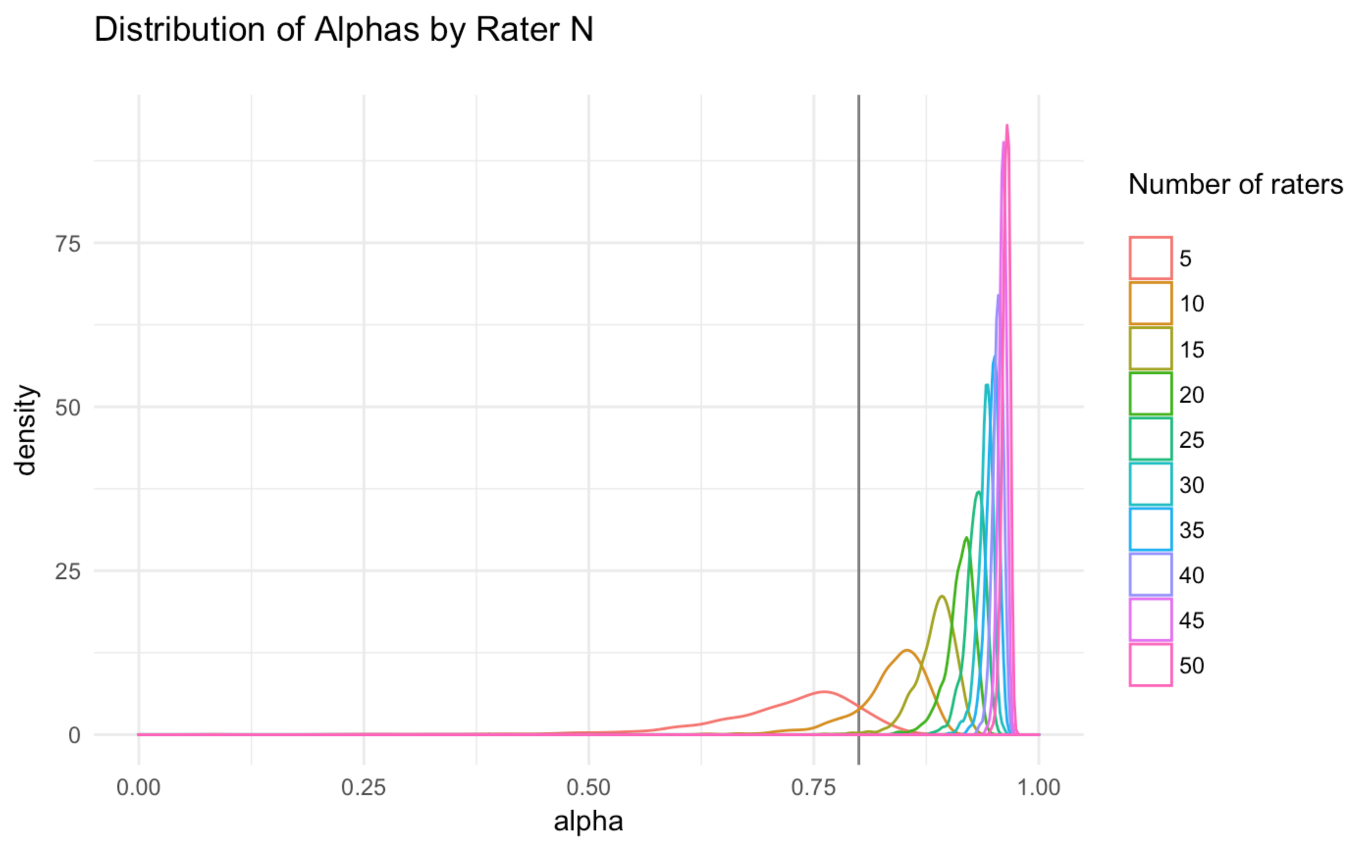
Hide

```
set.seed(909)
n_samples <- 1000
n_raters <- seq(5, 50, by = 5)
sim <- rep(n_raters, each = n_samples) %>%
  purrr::map_df( function(n) {
    get_alpha(data, n)
  })
```

Graph distribution of alphas

Hide

```
ggplot(sim) +
  geom_vline(xintercept = 0.8, color = "grey50" ) +
  geom_density(aes(alpha, colour = as.factor(n))) +
  xlim(0, 1) +
  labs(title = "Distribution of Alphas by Rater N\n",
       color = "Number of raters\n") +
  theme_minimal()
ggsave("alphas.png", width = 7, height = 5)
```

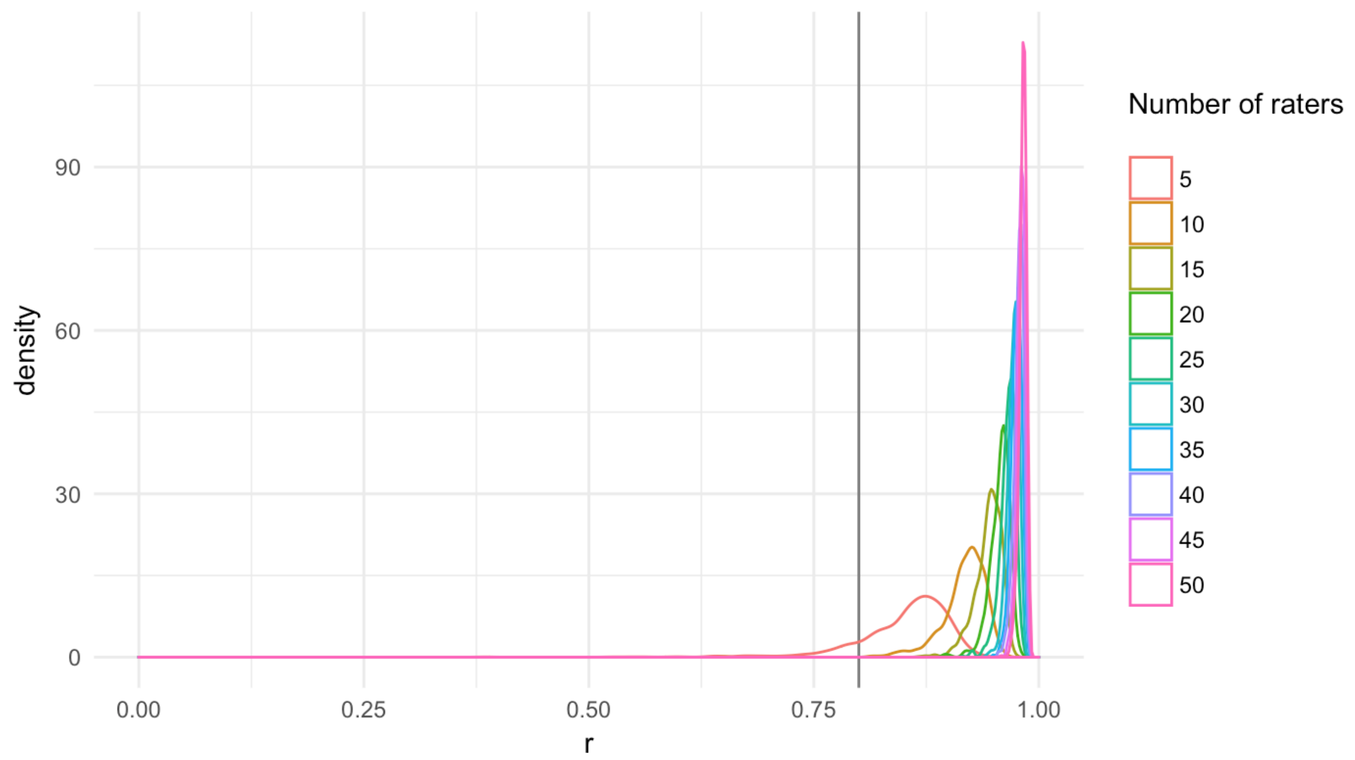


Graph distribution of correlations between sample means and canonical mean ratings.

Hide

```
ggplot(sim) +
  geom_vline(xintercept = 0.8, color = "grey50" ) +
  geom_density(aes(r, colour = as.factor(n))) +
  xlim(0, 1) +
  labs(title = "Distribution of Rating Correlations by Rater N\n",
       color = "Number of raters\n") +
  theme_minimal()
ggsave("correlations.png", width = 7, height = 5)
```

Distribution of Rating Correlations by Rater N



This table gives the median and 10th percentiles for `alpha` and `r`, as well as the proportion of `alpha`s over 0.8 (typically considered high).

Hide

```
sim %>%
  group_by(n) %>%
  summarise(
    `median alpha` = round(quantile(alpha, .5), 2),
    `90% alpha >` = round(quantile(alpha, .1), 2),
    `alpha >= 0.8` = round(mean(alpha >= 0.8), 2),
    `median r` = round(quantile(r, .5), 2),
    `90% r >` = round(quantile(r, .1), 2)
  )
```

n <dbl>	median alpha <dbl>	90% alpha > <dbl>	alpha >= 0.8 <dbl>	median r <dbl>	90% r > <dbl>
5	0.74	0.61	0.12	0.86	0.79
10	0.84	0.78	0.84	0.92	0.89
15	0.89	0.86	0.99	0.95	0.92
20	0.91	0.89	1.00	0.96	0.94
25	0.93	0.91	1.00	0.97	0.95
30	0.94	0.93	1.00	0.97	0.96
35	0.95	0.94	1.00	0.98	0.97
40	0.95	0.95	1.00	0.98	0.97

45	0.96	0.95	1.00	0.98	0.97
50	0.96	0.96	1.00	0.98	0.98

1-10 of 10 rows

---
title: "Determining the number of raters for reliable mean ratings"
author: "Lisa DeBruine & Benedict Jones"
output: html_notebook

How many raters you need to sample to get reliable stimulus ratings will obviously depend on the raters, stimuli, and what they're being rated for. For example, if there is a lot of inter-rater variation or very little inter-stimulus variation, you will need more raters to generate mean ratings with any reliability.

If you have a large set of ratings of a type of stimulus, population of rater, and type of rating you're interested in, you can use the script below to figure out how many raters you need to sample to get mean stimulus ratings that are well-correlated with mean ratings from the full sample.

The example below is for attractiveness ratings using an open-access image set from our lab.

```{r}
library(tidyverse)
library(purrr)
library(psych)
```

Read data from DeBruine, Lisa; Jones, Benedict (2017): Face Research Lab London Set. figshare. [doi: 10.6084/m9.figshare.5047666](https://doi.org/10.6084/m9.figshare.5047666)

```{r}
data <- read_csv("https://ndownloader.figshare.com/files/8542045")
```

Calculate canonical mean ratings (average of all available ratings)

```{r calc-canon}
canon <- data %>%
  select(X001:X173) %>% # select only columns with your stimuli
  group_by() %>%
  summarise_all(mean) %>%
  t()
```

Sample n raters from the set and calculate Cronbach's `alpha` and `r` from the Pearson's correlation with the canonical ratings.

```{r sim_function}
get_alpha <- function(data, n) {
  # sample your full dataset
```

````
  data_sample <- data %>%
    sample_n(n) %>%
    select(X001:X173) # select only columns with your stimuli

  # calculate cronbach's alpha
  capture.output(suppressWarnings(a <- psych::alpha(t(data_sample))))
  alpha <- a$total["std.alpha"] %>% pluck(1)

  # calculate mean sample ratings
  sample_means <- data_sample %>%
    group_by() %>%
    summarise_all(mean) %>%
    t()

  # calculate correlation between sample mean ratings and canon
  r <- cor(sample_means, canon)[[1,1]]

  # return relevant data
  tibble(
    n = n,
    alpha = alpha,
    r = r
  )
}
```

Generate 1000 samples for 5 to 50 raters.

```{r}
set.seed(909)

n_samples <- 1000
n_raters <- seq(5, 50, by = 5)

sim <- rep(n_raters, each = n_samples) %>%
  purrr::map_df( function(n) {
    get_alpha(data, n)
  })
```


Graph distribution of alphas

```{r}
ggplot(sim) +
  geom_vline(xintercept = 0.8, color = "grey50" ) +
  geom_density(aes(alpha, colour = as.factor(n))) +
  xlim(0, 1) +
  labs(title = "Distribution of Alphas by Rater N\n",
       color = "Number of raters\n") +
````

```
  theme_minimal()

ggsave("alphas.png", width = 7, height = 5)
```

Graph distribution of correlations between sample means and canonical mean ratings.

```{r}
ggplot(sim) +
  geom_vline(xintercept = 0.8, color = "grey50" ) +
  geom_density(aes(r, colour = as.factor(n))) +
  xlim(0, 1) +
  labs(title = "Distribution of Rating Correlations by Rater N\n",
       color = "Number of raters\n") +
  theme_minimal()

ggsave("correlations.png", width = 7, height = 5)
```

This table gives the median and 10th percentiles for `alpha` and `r`, as well as the proportion of `alpha`s over 0.8 (typically considered high).

```{r}
sim %>%
  group_by(n) %>%
  summarise(
    `median alpha` = round(quantile(alpha, .5), 2),
    `90% alpha >` = round(quantile(alpha, .1), 2),
    `alpha >= 0.8` = round(mean(alpha >= 0.8), 2),
    `median r` = round(quantile(r, .5), 2),
    `90% r >` = round(quantile(r, .1), 2)
  )
```