Neuropsychology

ISSN: 0894-4105

2022, Vol. 36, No. 7, 664–682
https://doi.org/10.1037/neu0000817

Does Culture Shape Our Understanding of Others' Thoughts and Emotions? An Investigation Across 12 Countries

François Quesque[1, 2, 3], Antoine Coutrot[4], Sharon Cox[5], Leonardo Cruz de Souza[6], Sandra Baez[7], Juan Felipe Cardona[8], Hannah Mulet-Perreault[9], Emma Flanagan[10, 11], Alejandra Neely-Prado[12], Maria Florencia Clarens[13], Luciana Cassimiro[14], Gada Musa[15], Jennifer Kemp[16], Anne Botzung[16], Nathalie Philippi[16], Maura Cosseddu[17], Catalina Trujillo-Llano[8], Johan Sebastián Grisales-Cardenas[8], Sol Fittipaldi[18, 19, 20], Nahuel Magrath Guimet[13], Ismael Luis Calandri[13], Lucia Crivelli[13], Lucas Sedeno[20], Adolfo M. Garcia[18, 19, 20, 21, 22, 23], Fermin Moreno[24, 25], Begoña Indakoetxea[24, 25], Alberto Benussi[26], Millena Vieira Brandão Moura[6], Hernando Santamaria-Garcia[27], Diana Matallana[27], Galina Pryanishnikova[28], Anna Morozova[29], Olga Iakovleva[28], Nadezda Veryugina[28], Oleg Levin[28], Lina Zhao[30], Junhua Liang[30], Thomas Duning[30], Thibaud Lebouvier[1, 2, 3], Florence Pasquier[1, 2, 3], David Huepe[12], Myriam Barandiaran[24, 25], Andreas Johnen[31], Elena Lyashenko[28], Ricardo F. Allegri[13], Barbara Borroni[26], Frederic Blanc[16], Fen Wang[30], Mônica Sanches Yassuda[14], Patricia Lillo[15, 32], Antônio Lúcio Teixeira[33], Paulo Caramelli[6], Carol Hudon[9], Andrea Slachevsky[15, 32, 34, 35], Agustin Ibáñez[18, 19, 20, 22, 23, 36], Michael Hornberger[10, 11], and Maxime Bertoux[1, 2, 3, 10, 11]

[1] Lille Neuroscience and Cognition, Inserm, CHU Lille, Université de Lille
[2] Lille Center of Excellence for Neurodegenerative Diseases (LiCEND), France
[3] Development of Innovative Strategy for Transdisciplinary Approach on Alzheimer's Disease Laboratory, Labex DistALZ, Lille, France
[4] CNRS, LS2N, Université de Nantes
[5] Department of Behavioural Science and Health, Institute of Epidemiology and Healthcare, University College London
[6] Behavioral and Cognitive Neurology Research Group, Faculdade de Medicina, Universidade Federal de Minas Gerais
[7] Department of Psychology, Universidad de Los Andes
[8] Centro de Investigación en Neurociencia Clínica y Comportamental (CINCCO), Facultad de Psicología, Universidad del Valle
[9] CERVO Brain Research Centre, Université Laval
[10] Norwich Medical School, University of East Anglia
[11] Department of Clinical Neurosciences, University of Cambridge
[12] Center for Social and Cognitive Neuroscience, School of Psychology, Universidad Adolfo Ibáñez
[13] Department of Neurology, FLENI Foundation
[14] School of Arts, Sciences and Humanities, University of São Paulo
[15] Physiopathology Department (ICBM), Neuroscience and East Neuroscience Departments, Faculty of Medicine, Neuropsychology and Clinical Neuroscience Laboratory (LANNEC), University of Chile
[16] Centre Mémoire de Ressources et Recherche (CMRR) des Hôpitaux universitaires de Strasbourg, University of Strasbourg
[17] Neurology Unit, Spedali Civili Hospital, Brescia, Italy
[18] Latin American Brain Health Institute (BrainLat), Universidad Adolfo Ibáñez
[19] Centro de Neurociencias Cognitivas (CNC), Universidad de San Andrés
[20] National Scientific and Technical Research Council (CONICET), Buenos Aires, Argentina
[21] Departamento de Lingüística y Literatura, Facultad de Humanidades, Universidad de Santiago de Chile
[22] Global Brain Health Institute (GBHI), University of California-San Francisco (UCSF)
[23] Global Brain Health Institute, Trinity College Dublin, University of Dublin
[24] Department of Neurology, Unit of Cognitive Disorders, Hospital Universitario Donostia
[25] Neurosciences Area, Group of Neurodegenerative Diseases, Biodonostia Institute, University of the Basque Country
[26] Centre for Neurodegenerative Disorders, Department of Clinical and Experimental Sciences, University of Brescia
[27] School of Medicine, Neuroscience Doctorate, Aging Institute, Physiology and Psychiatry Department, Pontificia Universidad Javeriana
[28] Russian Medical Academy of Continuous Professional Education, Moscow, Russia
[29] Central Clinic No 1 of the Ministry of Internal Affairs, Moscow, Russia
[30] Innovation Center for Neurological Disorders, Department of Neurology, Xuan Wu Hospital, Capital Medical University
[31] Clinic of Neurology With Institute for Translational Neurology, University Hospital Münster
[32] Geroscience Center for Brain Health and Metabolism (GERO), Faculty of Medicine, University of Chile
[33] Faculdade Santa Casa BH, Belo Horizonte, Brazil

[34] Memory and Neuropsychiatric Clinic (CMYN) Neurology Department, Hospital del Salvador and Faculty of Medicine, University of Chile
[35] Servicio de Neurología, Departamento de Medicina, Clínica Alemana-Universidad del Desarrollo
[36] Trinity College Institute of Neuroscience (TCIN), Trinity College Dublin, University of Dublin

Measures of social cognition have now become central in neuropsychology, being essential for early and differential diagnoses, follow-up, and rehabilitation in a wide range of conditions. With the scientific world becoming increasingly interconnected, international neuropsychological and medical collaborations are burgeoning to tackle the global challenges that are mental health conditions. These initiatives commonly merge data across a diversity of populations and countries, while ignoring their specificity. ***Objective:*** In this context, we aimed to estimate the influence of participants' nationality on social cognition evaluation. This issue is of particular importance as most cognitive tasks are developed in highly specific contexts, not representative of that encountered by the world's population. ***Method:*** Through a large international study across 18 sites, neuropsychologists assessed core aspects of social cognition in 587 participants from 12 countries using traditional and widely used tasks. ***Results:*** Age, gender, and education were found to impact measures of mentalizing and emotion recognition. After controlling for these factors, differences between countries accounted for more than 20% of the variance on both measures. Importantly, it was possible to isolate participants' nationality from potential translation issues, which classically constitute a major limitation. ***Conclusions:*** Overall, these findings highlight the need for important methodological shifts to better represent social cognition in both fundamental research and clinical practice, especially within emerging international networks and consortia.

Key Points
Question: We estimate the influence of cultural factors on social cognition assessment. ***Findings:*** Participants' nationality accounted for more than 20% of the variance of social cognition scores. ***Importance:*** Cognitive tasks are developed in highly specific contexts and should not be used in other cultures without adaptations. ***Next Steps:*** Finer-grain analyses of cross-cultural variations coupled with neural correlates of performances' convergences and divergences.

Keywords: social cognition, emotion recognition, mentalizing, theory of mind, culture

Supplemental materials: https://doi.org/10.1037/neu0000817.supp

This article was published Online First July 14, 2022.
Antoine Coutrot https://orcid.org/0000-0001-9569-3548
Sharon Cox https://orcid.org/0000-0001-8494-5105
Leonardo Cruz de Souza https://orcid.org/0000-0001-5027-9722
Emma Flanagan https://orcid.org/0000-0003-1492-7061
Alejandra Neely-Prado https://orcid.org/0000-0002-5565-9676
Maria Florencia Clarens https://orcid.org/0000-0003-4334-4404
Maura Cosseddu https://orcid.org/0000-0003-3718-9153
Johan Sebastián Grisales-Cardenas https://orcid.org/0000-0002-7460-2923
Sol Fittipaldi https://orcid.org/0000-0001-7704-3349
Ismael Luis Calandri https://orcid.org/0000-0002-6983-1430
Lucia Crivelli https://orcid.org/0000-0002-0083-9389
Adolfo M. Garcia https://orcid.org/0000-0002-6936-0114
Florence Pasquier https://orcid.org/0000-0001-9880-9788
Elena Lyashenko https://orcid.org/0000-0002-4483-6147
Mônica Sanches Yassuda https://orcid.org/0000-0002-9182-2450
Antônio Lúcio Teixeira https://orcid.org/0000-0002-9621-5422
Paulo Caramelli https://orcid.org/0000-0002-4786-6990
Agustin Ibáñez https://orcid.org/0000-0001-6758-5101
Michael Hornberger https://orcid.org/0000-0002-2214-3788
Maxime Bertoux https://orcid.org/0000-0002-5283-503X
Agustin Ibáñez and Michael Hornberger contributed equally.
The authors would like to thank Prof. Stephanie D. Preston and Prof. Skye McDonald. The authors also deeply appreciated the help of Prof. McDonald for English revision. For their administrative or clinical support related to the acquisition of data, the authors would like to thank María Goldfeder, Agostina Galiani (INECO Foundation, Argentina), Ma Eugenia Martin, Cecilia Cervino, Ma Eugenia Tabernero, Micaela Arruabarena, Carlos Martinez Canyazo, Yanina Varela (FLENI Foundation, Argentina), Sophia Reul, Sabine Bruchmann (University Hospital Muenster, Germany), Aurélie Funkiewiez, Bruno Dubois, Elisabeth Ryckaert (Pitié-Salpêtrière Hospital, France), Gara Tomé, María de Arriba (Hospital Universitario Donostia, Spain), Clélie Philipps, Thimotée Albasser (University of Strasbourg, France), Alicia Wilcox, James Rowe (University of Cambridge, U.K.), Stephen Jeffs (University of East Anglia, U.K.), Ana Carolina Ventura, Andrezza Ferreira dos Santos, Caroline Folli, Emanuel Nery, Gabrielle Oliveira (University of São Paulo, Brazil), and Ângelo Ribeiro Vaz de Faria (Federal University of Minas Gerais, Belo Horizonte, Brazil). The following funding sources have been involved in the study: Administrative Department of Science, Technology and Innovation from Colombia (COLCIENCIAS 1106-744-55314), Takeda Grant (CW2680521), CONICET, FONCYT-PICT (2017-1818, 2017-1820), ANID/FONDAP (15150012), ANID/PIA/ANILLOS (ACT210096) ANID/FONDECYT (1140423 & 1100975), GBHI

continued

Humans are a highly social species, characterized by a unique level of cooperation among exceptionally large and genetically heterogeneous groups. In the last decades, cognitive tasks have been validated to quantify human social cognition (i.e., the set abilities allowing us to interact efficiently with others); however, the ability of these tests to be suitable worldwide could be legitimately questioned. Indeed, while the influence of factors such as gender or education on cognition have been considered before, how culture impacts cognitive functioning or its measurement has been traditionally ignored in neuropsychology and thus remains largely unknown (Barrett, 2020). Cognitive tests are indeed most often developed in western, industrialized, and democratic countries and their norms based on well-educated and rich people (referred to as "WEIRD" people, Henrich et al., 2010a, 2010b). Yet, it is becoming increasingly clear that cognitive measures no longer have acceptable validity when used with individuals from populations that do not fit these specific cultural characteristics. Comparisons made between the performance on common tests obtained by different ethnic groups within a single English-speaking country have shown that participants who learned English as a first language tended to have better performances in verbal (digit span, naming and fluency tests) and nonverbal (visuoperceptual tests) tests as compared to those who learned English as a second language (Boone et al., 2007; Statucka & Cohn, 2019). The differences observed may not be entirely driven by language expertise as even when translated and administered in the participant's native language, cognitive tests can remain more difficult for nonnative English speaker participants (Goodman et al., 2021). Important variations between participants from different countries were also revealed for verbal and nonverbal tests when using North American neuropsychological tests (Daugherty et al., 2017). Altogether, this shows that cultural variation, despite being an imprecise concept, could explain some of the observed neuropsychological differences.

The cross-cultural validity of neuropsychological tests and the overrepresentation of White educated people in test development and norms is far from being a trivial issue, as WEIRD participants represent only 12% of humanity (Arnett, 2016). This narrow sampling of the world's population constitutes a serious threat on the theoretical level, as a conclusion drawn from WEIRD samples in neuroscience studies should not be applied *in extenso* to non-WEIRD populations (Matsumoto & Juang, 2016). As cognitive evaluation also informs psychological and medical practices, this bias could have devastating consequences for educational or health decisions, including misdiagnosis and potentially inadequate treatment prescription. As an example, one study reported that the differences in performance between participants of different

ALZ UK-20-639295 and Alzheimer's Association GBHI ALZ UK-22-865742, ANID/FONDECYT REGULAR (1210176, 1210195, 1220995), Programa Interdisciplinario de Investigación Experimental en Comunicación y Cognición (PIIECC), Facultad de Humanidades, USACH, NIH NIA R01 AG057234, the Réseau québécois de recherche sur le vieillissement, the National Natural Science Foundation of China (No. 81100797, 81530036), Beijing high-level health talents training project (No. 2015-3-068), Russian Foundation for Basic Research (No. 20-013-00600), CNPq, Brazil (Ref. 402853/2012-1), Projet Hospitalier de Recherche Clinique (PHRC) inter-régional (IDRCB 2012-A00992-41; HUS No. 5330), Alzheimer's Research U.K. (ARUK_PPG2014B-3; ARUK-DT2016-1), Alzheimer's Society U.K. (AS-SF-14-003), Deutsche Telekom, São Paulo Research Foundation (FAPESP—2016/07967-2), Multi-Partner Consortium to Expand Dementia Research in Latin America (ReDLat)—funded by the National Institutes of Aging of the National Institutes of Health under award number R01AG057234, an Alzheimer's Association grant (SG-20-725707-ReDLat), the Rainwater Foundation, and the Global Brain Health Institute. The authors also thank the Laboratory of Excellence Development of Innovative Strategies for a Transdisciplinary Approach to Alzheimer's Disease (LabEx DISTALZ) as well as the Région Hauts-de-France (Dispositif STaRS), and the Thérèse et René Planiol Foundation. The content is solely the responsibility of the authors and does not represent the official views of the funding agencies.

Data and scripts used for the analyses of these data are freely available online at https://osf.io/dksv7/?view_only=9386572b0206439ca98d1a03657685ab

François Quesque played lead role in writing of original draft, supporting role in conceptualization, project administration and visualization and equal role in data curation, formal analysis and writing of review and editing. Antoine Coutrot played lead role in formal analysis and visualization and equal role in conceptualization, methodology, writing of original draft and writing of review and editing. Sharon Cox played supporting role in writing of review and editing and equal role in conceptualization, data curation, funding acquisition and investigation. Leonardo Cruz de Souza played supporting role in project administration and equal role in conceptualization, data curation, funding acquisition, investigation and writing of review and editing. Sandra Baez played equal role in data curation, funding acquisition, investigation and writing of review and editing. Juan Felipe Cardona played equal role in data curation, funding acquisition, investigation and writing of review and editing. Hannah Mulet-Perreault played equal role in data curation, formal analysis, investigation and writing of review and editing. Emma Flanagan played supporting role in project administration and equal role in data curation, investigation and writing of review and editing. Alejandra Neely-Prado played supporting role in writing of review and editing and equal role in data curation, funding acquisition and investigation. Maria Florencia Clarens played supporting role in project administration and writing of review and editing and equal role in data curation, investigation and referent investigator for fleni. Luciana Cassimiro played supporting role in writing of review and editing and equal role in data curation and investigation. Gada Musa played equal role in data curation, funding acquisition, investigation and writing of review and editing. Jennifer Kemp played equal role in data curation, funding acquisition and investigation. Anne Botzung played supporting role in investigation and project administration and equal role in data curation and funding acquisition. Nathalie Philippi played supporting role in funding acquisition and project administration and equal role in data curation. Maura Cosseddu played supporting role in conceptualization and writing of review and editing and equal role in data curation, funding acquisition and investigation. Catalina Trujillo-Llano played supporting role in writing of review and editing and equal role in data curation and investigation. Johan Sebastián Grisales-Cardenas played supporting role in writing of review and editing and equal role in data curation and investigation. Sol Fittipaldi played supporting role in project administration and writing of review and editing and equal role in data curation, funding acquisition, investigation and methodology. Nahuel Magrath Guimet played supporting role in project administration and writing of review and editing and equal role in data curation and investigation. Ismael Luis Calandri played supporting role in writing of review and editing and equal role in data curation, funding acquisition and investigation. Lucia Crivelli played supporting role in writing of review and editing and equal role in data curation, funding acquisition and investigation. Lucas Sedeno played equal role in data curation, funding acquisition, investigation and writing of review and editing. Adolfo M. Garcia played supporting role in project administration and equal role in data curation, funding acquisition, investigation and writing of review and editing. Fermin Moreno played supporting role in writing of review and editing and equal role in data curation, funding

continued

nationalities in verbal and nonverbal testing of cognition could lead to diagnostic errors depending on the origin of participants (Daugherty et al., 2017). Another striking example is the erroneous classification of almost half of healthy Black Americans as cognitively impaired when using one of the commonly used memory tests alongside its original standardized scores, predominantly based on data from White individuals (Norman et al., 2000). Therefore, such a "universalist" approach should be abandoned.

The study of cultural variation has a long tradition in sociology (e.g., Bendix, 1963), economy (e.g., Wright, 1970), and psychology (e.g., Brislin et al., 1973). Empirical data collection has been carried out in a variety of countries in order to compare—in a systematic manner—how attitudes and beliefs are influenced by local specificities. Contrasting with these trends, within the field of neuropsychology, cultural influences on test performance have only been investigated recently, mainly by comparing populations originating from a handful of countries (e.g., Daugherty et al., 2017; Statucka & Cohn 2019). In addition, for each country involved, participants are generally recruited from a single site, which might generate important confounds. Indeed, Talhelm et al. (2014) have shown that there could be major cultural differences within large countries such as China, and that the East versus West contrast, classically performed to assess cultural variation, could be even more simplistic when relying on single site studies. Moreover, in psychology, samples have been generally composed of a few young and well-educated people, which constitute an additional limitation to the generalization of findings.

Since differences have been observed across countries in memory and spatial navigation tasks (Coutrot et al., 2018; Hayden et al., 2014), one could expect even higher variations in tests assessing social cognition, which involve culture-dependent concepts or rules. However, in a historical context marked by antisocialism opinions, research in social cognition has been historically neglected due to the caricature of "social" as involuntary, irrational and regimented, established by early social psychologists, as well as the apparent threat to liberalism that was posed by socially engaged forms of cognition (see Greenwood, 2004). Considering this historical neglect, we believe that there remains a general underestimation of the importance of social cognition and its determinants, in both research and clinical settings. As deficits in social cognition lead to varied interpersonal difficulties that have been recognized as more incapacitating than traditionally assessed cognitive deficits (Henry et al., 2016), deeply impacting both the patient's (Santamaría-García et al., 2020) and their relatives' quality of life (Spitzer et al., 2019), there is an urgent

acquisition and investigation. Begoña Indakoetxea played supporting role in writing of review and editing and equal role in data curation, funding acquisition and investigation. Alberto Benussi played supporting role in writing of review and editing and equal role in data curation, funding acquisition and investigation. Millena Vieira Brandão Moura played supporting role in writing of review and editing and equal role in data curation and investigation. Hernando Santamaria-Garcia played supporting role in writing of review and editing and equal role in data curation, funding acquisition and investigation. Diana Matallana played supporting role in writing of review and editing and equal role in data curation and investigation. Galina Pryanishnikova played supporting role in writing of review and editing and equal role in data curation and investigation. Anna Morozova played supporting role in writing of review and editing and equal role in data curation and investigation. Olga Iakovleva played supporting role in writing of review and editing and equal role in data curation and investigation. Nadezda Veryugina played supporting role in writing of review and editing and equal role in data curation and investigation. Oleg Levin played supporting role in writing of review and editing and equal role in data curation, funding acquisition and investigation. Lina Zhao played supporting role in writing of review and editing and equal role in data curation and investigation. Junhua Liang played supporting role in writing of review and editing and equal role in data curation and investigation. Thomas Duning played supporting role in writing of review and editing and equal role in data curation, funding acquisition and investigation. Thibaud Lebouvier played supporting role in writing of review and editing and equal role in data curation and investigation. Florence Pasquier played supporting role in writing of review and editing and equal role in data curation and investigation. David Huepe played supporting role in writing of review and editing and equal role in data curation, funding acquisition and investigation. Myriam Barandiaran played supporting role in conceptualization, project administration and writing of review and editing and equal role in data curation, funding acquisition and investigation. Andreas Johnen played supporting role in conceptualization and writing of review and editing and equal role in data curation and investigation. Elena Lyashenko played supporting role in project administration and writing of review and editing and equal role in data curation, funding acquisition and investigation. Ricardo F. Allegri played supporting role in writing of review and editing and equal role in data curation, funding acquisition and investigation. Barbara Borroni played supporting role in project administration and writing of review and editing and equal role in data curation, funding acquisition and investigation. Frederic Blanc played supporting role in writing of review and editing and equal role in data curation, funding acquisition and investigation. Fen Wang played supporting role in project administration and writing of review and editing and equal role in data curation, funding acquisition and investigation. Mônica Sanches Yassuda played supporting role in conceptualization and writing of review and editing and equal role in data curation, funding acquisition and investigation. Patricia Lillo played supporting role in writing of review and editing and equal role in data curation, funding acquisition and investigation. Antônio Lúcio Teixeira played supporting role in writing of review and editing and equal role in data curation, funding acquisition and investigation. Paulo Caramelli played supporting role in writing of review and editing and equal role in data curation, funding acquisition and investigation. Carol Hudon played supporting role in project administration and writing of review and editing and equal role in data curation, funding acquisition and investigation. Andrea Slachevsky played equal role in conceptualization, data curation, funding acquisition, investigation and writing of review and editing. Agustin Ibáñez played supporting role in conceptualization, formal analysis and project administration and equal role in data curation, funding acquisition, investigation, supervision and writing of review and editing. Michael Hornberger played supporting role in project administration and equal role in conceptualization, data curation, funding acquisition, investigation and writing of review and editing. Maxime Bertoux played lead role in conceptualization, data curation, investigation, project administration, resources, supervision, writing of original draft and writing of review and editing and equal role in formal analysis, funding acquisition and methodology.

The data are available at https://osf.io/dksv7/?view_only=9386572b0206439ca98d1a03657685ab

The experiment materials are available at https://osf.io/dksv7/?view_only=9386572b0206439ca98d1a03657685ab

Correspondence concerning this article should be addressed to Maxime Bertoux, Lille Neuroscience and Cognition, Inserm, CHU Lille, Université de Lille, F-59000, Lille, France. Email: maxime.bertoux@inserm.fr

challenge to explore and quantify the possible cultural variations that could be at stake in the assessment of this domain, to improve the quality of neuropsychological evaluation and its relevance. This is especially true when evaluating the ability to infer other's mental states and to recognize emotions from faces. These are consensually considered as core abilities of social cognition (Cotter et al., 2018; Henry et al., 2016) that have been widely explored by validated tests (Ekman & Friesen, 1976; Stone et al., 1998) that are regarded worldwide as reliable measures (Russell et al., 2020). Modified and reduced versions of these tests have been used in a wide diversity of clinical contexts, from frontotemporal degeneration and depression (Bertoux et al., 2012) to rheumatoid arthritis (Gwinnutt et al., 2021).

Some recent studies have investigated social cognition variation across countries using online surveys. For example, when exploring differences regarding moral dilemmas across 130 countries, Awad et al. (2018) identified three distinct "moral clusters" of countries ("western," "eastern," "southern") with different attitudes toward the same problems. While this methodological approach provides the means to accrue large quantities of data, it precludes formal cognitive assessment (i.e., performance-based) that would provide reliable findings regarding the evaluation of cognition in clinical settings. As a consequence, appeals for international collaborations have been conveyed over recent years (Barrett, 2020; Bauer, 2020). Anticipating these recommendations, the International Network on Social Cognition Disorders (INSCD) was developed in 2014. Specifically, we investigated normative variations across 12 countries in widely used cognitive tests that assess the core abilities of social cognition: mental state inferences and facial emotion recognition (FER). We focused on these two abilities and their associated tests because of the prominence they have in social cognitive evaluation (Cotter et al., 2018; Eddy, 2019; Henry et al., 2016; Quesque & Rossetti, 2020). While having originated from two western industrialized countries (i.e., USA & England), these tests are now used across the world in the context of neuropsychological evaluations. Fostering the diagnostic utility of social cognition, recent clinical recommendations advocate for a generalized use of these classical tests (Ducharme et al., 2020), but some questions, such as whether outcomes can be compared unequivocally across countries, remain unanswered. Moreover, from a fundamental perspective, with the increasing development of neuroscientific international projects relying on shared clinical data (e.g., Human Brain Project https://www.humanbrainproject.eu/en/, Enhancing NeuroImaging Genetics through Meta-Analysis [ENIGMA-Network] http://enigma.ini.usc.edu/, The Genetic Frontotemporal Initiative (GENFI) https://www.genfi.org/, The Neuropsychiatric International Consortium on Frontotemporal Dementia (NIC-FTD) https://www.alzheimercentrum.nl/wetenschap/lopend-onderzoek/nic-ftd/, Research Dementia Latin America (RedLat) https://www.gbhi.org/projects/multi-partner-consortium-expand-dementia-research-latin-america-redlat), it is absolutely critical that we investigate whether performances on widely used social cognition tests are influenced by local characteristics. We hypothesized that important variations between countries would emerge in both measures, alongside a general, significant influence of age, gender, and education on social cognitive performance.

Method

Participants

The recruitment and assessment of 587 healthy control participants (339 women, age range 18–89 years, M = 58.04 years ± 16.06 SD, mean education = 10.5 years ± 5.46 SD, where SD is standard deviation) was performed through 18 centers from 12 countries within the INSCD, a worldwide clinical consortium. Country-specific group sizes, assessment language, gender ratio, and mean years of education are reported in Table 1. Centers received local ethics approval (see Supplemental Material), and all participants signed informed consent prior to their inclusion. Common inclusion criteria included the following: (a) no cognitive complaints; (b) no depressive complaints; (c) a normal cognitive screening test (i.e., either the mini–mental state examination in 10/18 centers, the Montreal Cognitive Assessment in 5/18 centers, or the Addenbrooke's Cognitive Examination III in 3/18 centers, using the local normative data, considering age, gender, and education; see Supplemental Material 1); (d) no current psychiatric disorder; (e) no past or current neurological disease; (f) native language matching that of the assessment. All participants were recruited as controls for ongoing research studies, either focused on the establishment of local normative data or on comparison with patients with different diagnoses. Additional details

Table 1

Countries Specific Group Sizes, Assessment Language, Gender Ratio, Mean Years of Education as Well as Mean Mentalizing (Modified Faux Pas Test) and Facial Emotion Recognition Scores and Standard Deviations (SD)

Country	Assessment language	*N*	Gender ratio (women/all)	Mean education (years)	mFP (*M*)	mFP (*SD*)	FER (*M*)	FER (*SD*)
Argentina	Spanish	49	0.73	15.24	13.66	1.53	13.12	1.29
Brazil	Portuguese	53	0.68	11.74	12.88	1.76	11.75	1.08
Canada	French	55	0.82	14.73	12.30	1.74	12.12	1.10
Chile	Spanish	56	0.63	13.55	12.45	1.92	11.74	1.21
China	Chinese	34	0.56	11.59	12.28	1.54	11.86	1.02
Colombia	Spanish	30	0.37	12.67	13.56	1.23	12.47	0.97
England	English	97	0.45	14.78	14.56	0.67	13.64	1.10
France	French	88	0.56	14.71	13.22	1.62	12.58	1.22
Germany	German	19	0.53	11.68	NA	NA	12.03	0.82
Italy	Italian	19	0.47	11.63	13.98	1.20	11.82	0.90
Russia	Russian	29	0.62	15.72	13.14	1.57	12.63	1.29
Spain	Spanish	58	0.47	17.60	11.76	1.90	12.14	1.25

Note. mPF = modified and reduced version of the faux pas test; FER = facial emotion recognition.

such as center-specific inclusion criteria (e.g., normal magnetic resonance imaging [MRI] according to the investigators' expertise, normal neuropsychological examination), sample sizes, gender distribution and age ranges are presented in the Supplemental Material. In order to control for the influence of sample bias, age, gender, and education were included in our statistical analyses addressing cross countries' variations.

Ethics approval was given for the center for Addictive Behaviors Research, London South Bank University from the School of Applied Sciences Ethics Committee, London South Bank University. For the Clínica Neurólogos de Occidente, Cali, Colombia, the Comité Institucional de Revisión de Ética Humana (CIREH), Universidad del Valle (Number: CIREH 015-017). For the FLENI Foundation, Buenos Aires, Argentina, the Comité Investigación Fleni. For the Instituto de Neurología Cognitiva, Centro de Estudios de la Memoria y la Conducta, Buenos Aires, Argentina, the Comité Institucional de Ética del Instituto de Neurología Cognitiva, Centro de Estudios de la Memoria y la Conducta, Buenos Aires, Argentina (project: "El papel de las modulaciones contextuales durante el lenguaje de acción y el procesamiento emocional en la neurodegeneración." Approval issued on August 8, 2017). For the Universidad Adolfo Ibáñez, Santiago de Chile, Chile, the Comité Ético de Investigación de la Universidad Adolfo Ibáñez, Santiago de Chile, Chile (project: "The role of contextual modulations during action language and emotional processing in neurodegeneration." Approval issued on June 15, 2017). For the Centre de recherche CERVO, Université Laval, Québec, the Comité d'éthique de la recherche sectoriel en neurosciences et santé mentale (Number: 2019-1541). For the Xuanwu Hospital of Capital Medical University, China, the Ethics Committee of Xuanwu Hospital of Capital Medical University. For the University Hospital Muenster, Germany, the Ethics committee of the University Hospital Muenster and medical association of Westfalen-Lippe (Number: 2012-365-f-S). For the Centre de Neuropsychologie, Fédération de Neurologie, Hôpital de la Pitié-Salpêtrière, the Comité Éthique de l'Hôpital Pitié-Salpêtrière, Paris (Reference: RBM05-15). For the Faculdade de Medicina, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte (MG), Brazil, the Comitê de Ética em Pesquisa da UFMG (Reference: COEP-UFMG—17850513.2.0000.5149). For the Geroscience Center for Brain Health and Metabolism (GERO), Faculty of Medicine, University of Chile, Avenida Salvador 486, Providencia, Santiago, Chile, the Comite de etica Servicio de Salud Metropolitano Oriente Comité de etica Servicio de Salud Metropolitano Sur, Comité de etica Facultad de Medicina Universidad de Chile. For the Department of Clinical and Experimental Sciences, University of Brescia, Brescia, Italy, the Brescia Hospital Ethics Committee (Reference: NP2224). For the Hospital Universitario Donostia, San Sebastian, Spain, the Donostia University Hospital Ethics Committee. For the Centre Mémoire de Ressources et de Recherche, Hôpitaux Universitaires de Strasbourg, the Comité de Protection des Personnes Est IV. For the Department of Clinical Neurosciences, Addenbrookes Hospital, University of Cambridge and the Norwich Medical School, University of East Anglia, Norwich, U.K., the National Research Ethics Service (NRES) Committee London—Queens Square (respectively reference "COGENT" 14/LO/2045 and "TRACC-UEA" 16/LO/1366). For the University of São Paulo, School of Arts, Sciences and Humanities, São Paulo, São Paulo, Brazil, the Ethics Committee of the Hospital das Clínicas de São Paulo (Reference: CAAE 04970612.1.0000.0068).

Materials

The mini social cognition & emotional assessment (mini-SEA; Bertoux et al., 2012) was administered to all participants. This short battery is composed of adaptations of two classical tests: a modified and reduced version of the faux pas test (mFP, Stone et al., 1998), evaluating the ability to decode social rules and to infer others' knowledge, intention and feelings and a reduced version of the picture of facial affect test (Ekman & Friesen, 1976), assessing FER. Both tests are among three of the most used tasks to examine social cognition among neuropsychiatric populations (Eddy, 2019). The mini-SEA has been validated in different languages (e.g., Clarens et al., 2021; Quesque et al., 2020) and clinical contexts (e.g., Bertoux et al., 2013; El Grabli et al., 2022; Sensenbrenner et al., 2020) since its first publication and has been used in numerous studies to assess social cognition over the past decade. Details regarding its validation have been published elsewhere (Quesque et al., 2020). The computation of a general mini-SEA score (/30) and mFP and FER subscores (/15) was performed.

The modified and reduced version of the faux pas test (mFP) is composed of 10 short stories depicting a short social scene in which one character either commits (in 5 stories) or does not commit (in 5 others) a social faux pas. See Supplemental Table 1, for details. Three drawings illustrate each story (see Supplemental Material 3, for an example). The test involves the ability to decode social rules and to infer others' knowledge, intentions, and feelings (mentalizing). The task of the participants is to read each story aloud and to detect the presence or absence of a faux pas (an embarrassing action given the context). For example, in a "faux pas story," somebody says she never liked the bowl her friend just broke, although this was a wedding gift from the latter to the former. In a story without faux pas, somebody offers his seat to an older lady in a city bus. In stories with faux pas, if a faux pas is correctly detected (1 point), the participants have to answer to five questions in order to assess their ability to understand who committed the faux pas, what it was, the knowledge and intention of the person who committed the faux pas and the feeling experienced by the person who was victim of it (5 points, one per question). For the other stories, the accurate detection of the absence of faux pas provides two points. Two control questions (providing 1 point each) assess the general understanding of the text for all stories. Participants were invited to read the stories as many times as necessary to answer the questions. Answers were recorded textually by the clinicians. The final score on 40 (5×6 points for the items with a faux pas + 5×2 points for the items without faux pas) was normalized and reported on a 15 points scale (i.e., subscore 1). Finally, it is important to note that as this task was developed in England, the original "English" stories were used as the reference point. Therefore, the original faux pas was always considered as faux pas in all countries. This decision reflects current clinical practices.

The reduced version of the picture of facial affect test involves FER. Participants are sequentially presented with 35 Black and White photos of White human faces and, for each face, have to choose one label among seven (happiness, surprise, neutral, sadness, anger, disgust, fear) that matches the emotion displayed. All faces were presented in a fixed random order. For all countries except China, where a specific version was used (see Gong et al., 2011 and Supplemental Material 4, for more details), the original items were used. As it was the case for the faux pas task, the originally defined correct responses (see Ekman & Friesen, 1976) were used as references for all participants involved, independent of their respective culture. The final score out of 35 was also normalized and reported on a 15 points scale (i.e., subscore 2). Again, this decision reflects current clinical practices.

Procedure

Contrasting with cross-cultural surveys relying on questionnaires, the present study involved a genuine cognitive assessment under neuropsychological testing conditions (i.e., face-to-face performance-based psychological evaluation by a health professional). Such a methodology implied smaller sample sizes than in classical surveys and prevented the use of classical statistical procedures (e.g., confirmatory factor analysis) to evaluate measurement equivalence across cultures (Davidov et al., 2018). Regarding this point, we were aware that it was crucial to obtain a thorough insight on social cognitive skills in each country involved and to minimize, as far as possible, any construct bias. Interestingly, the tests included in this study are already used in each country involved in the present study, which implies that local experts judged them as appropriate in their culture (congruently with Johnson's, 1998 recommendations). Within the following paragraphs, we underline the efforts taken in order to limit the occurrence of method bias.

As the original tests were published in English, specific translations were used for each language prior to the study (the same translation was used in Argentina, Chile, Colombia and Spain, http://www.autismresearchcentre.com/tests). Then, specific instructions and scoring procedures were applied to the mini-SEA. Examples of scoring were shared across the network to ensure a rigorous standardization of the assessment.

All participants were recruited through local advertisement (but methods varied across the network, e.g., poster, advertisement in local newspaper, newsletter ...). In all centers, the tests were administered by a well-trained professional, that is, a neuropsychologist in all centers, except in Moskva and Belo Horizonte where it was done by a senior neurologist trained in neuropsychology. This minimizes method bias (see Collins, 1980; Johnson, 1998). Moreover, the assessment was done in the participants' and clinicians' native languages, in compliance with Puente et al. (2013) recommendations. The items of the tests were printed on A4 pages and presented in the same order. All participants were examined individually in a quiet consultation room, ensuring standardized physical conditions, and thus reducing the occurrence of administration bias.

To ensure standardization of data, a procedure for the inclusion of fully anonymized data in center-specific databases was shared across the whole network. Databases were then centralized in Cambridge (U.K.) then Lille (France), where two individual raters performed quality checks of the data (e.g., thresholds for cognitive screening tests, appropriateness of ranges for every score and codes for item responses, consistency among individual responses, subscores and general scores, MRI normality, years of education, etc. were checked). These checks involved queries and corrections made by the centers in order to match the standardization of the data and meet the highest quality standard. Data were then made available online.

Data Analysis

We fit a linear mixed model for each score with age, gender, and education as fixed effect and country as random effect: score ~ age + gender + education + (1|country). In these models, we controlled for age, gender, and education while allowing the intercept to vary by country. The parameters of the linear mixed models were estimated with the restricted maximum likelihood method, and the covariance matrix of the random effects was estimated with the Cholesky parameterization. The variance partitioning coefficients (VPC), also named intraclass coefficients, quantified the proportion of observed variation in the outcome that is attributable to the effect of clustering by country. It is the ratio of the between-cluster variance to the total variance. See Equation 1, in which σ_0^2 is the variance between the countries and σ^2 the residual variance.

$$\text{VPC} = \frac{\sigma_0^2}{\sigma_0^2 + \sigma^2}. \quad (1)$$

To compute the confidence intervals around the VPC, we created a bootstrapped distribution of the VPC (10,000 iterations), then obtained the relevant quantiles from that distribution. Hedge's g was employed as the effect size for education and gender effects (positive values corresponding to an advantage for women and more educated participants respectively). Scripts used for the data analysis were made available online as well.

Results

Effect of Age, Gender, and Education

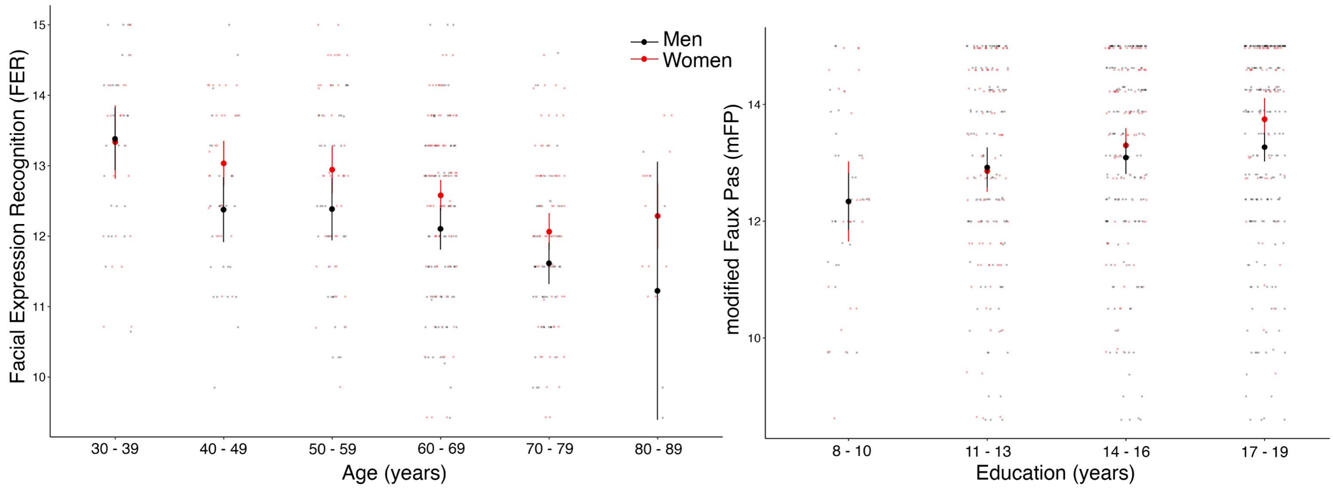
Mentalizing Score

Independent of the origin of participants, we found that age had a negative effect on the modified faux pas score, $t(537) = -2.01$, $p = .03$. We found a small effect of gender, $t(537) = -1.98$, $p = .048$, but the Hedge's g effect size is almost null: $g = 0.03$, 95% CI [−0.14, 0.20], positive values corresponding to an advantage for women. We found a positive effect of education on the modified faux pas score, $t(537) = 4.00$, $p < .001$. Figure 1b shows an increase of faux pas detection score with education, both for men and women. We computed the effect size of education with Hedge's g, comparing participants with less than the median education duration (14 years, $N = 271$) to participants with the median education duration or more ($N = 313$), $g = 0.35$, 95% CI [0.17, 0.51], with positive values corresponding to an advantage for more educated participants.

Facial Emotion Recognition Score

Age also had a negative effect on the FER score, $t(566) = -6.08$, $p < .001$, with a monotonic decline over the lifespan. Figure 1a shows the decline of FER scores with age, both for men and women. Gender also had an effect on the FER score, with women having higher scores than men, $t(566) = 4.95$, $p < .001$. We computed the effect size of gender with Hedge's g, $g = 0.23$; 95% CI [0.06, 0.40], positive values corresponding to an advantage for women. We did not find an effect of education, $t(566) = 0.91$; $p = .36$. Effect of age on the mFP and education on the FER are illustrated on the Supplemental Material 6.

Figure 1
Age, Education and Gender Effects



Note. (Left graph) Effect of education and gender on the modified faux pas score (mFP). mFP is averaged within 3-year windows for education. (Right graph) Effect of age and gender on facial emotion recognition score (FER). FER is averaged within 10-year windows. Error bars correspond to 95% confidence intervals, center of the error bars corresponds to the mean, and dots correspond to individual data points. See the online article for the color version of this figure.

Effect of Nationality

To estimate variations across countries, we compared the linear mixed model described above to a model only including age, gender, and education. The likelihood ratio test statistic is LRStat(1) = 118.53, $p < .001$ for mFP, LRStat(1) = 119.01; $p < .001$ for FER, indicating that the scores are better predicted when the country random effect was included in the model.

Variation of Mentalizing Scores Across Countries

The VPC indicates that 24.52%, 95% CI [9.10, 41.32], of the variance in the modified faux pas scores can be attributed to differences between nationalities after controlling for age, gender and education. Figure 2a represents the countries ranked according to their conditional modes. We computed Hedges' g between every pair of countries, see heatmap Figure 2b (see also the Supplemental Material 7, for the significance of each pairwise country differences). Unlike conditional modes, Hedge's g is not corrected for the potential effect of other variables, but they have the advantage of being straightforward and to allow a direct comparison between studies. Hedge's g between the first and last countries for the faux pas detection score (England vs. Spain) was $g = 2.17$, 95% CI [1.35, 2.97]. We computed the gender effect size for each country (Figure 2c). It went from $g = 1.42$, 95% CI [0.51, 2.33] in Russia to $g = -0.40$, 95% CI[−1.07, 0.28] in China, with positive values corresponding to higher scores for women. Education effect sizes for each country (comparison between low vs. high education, relying on median-split) went from $g = 1.50$; 95% CI [0.35, 2.62] in Russia to $g = -0.05$; 95% CI [−0.68, 0.57] in Argentina.

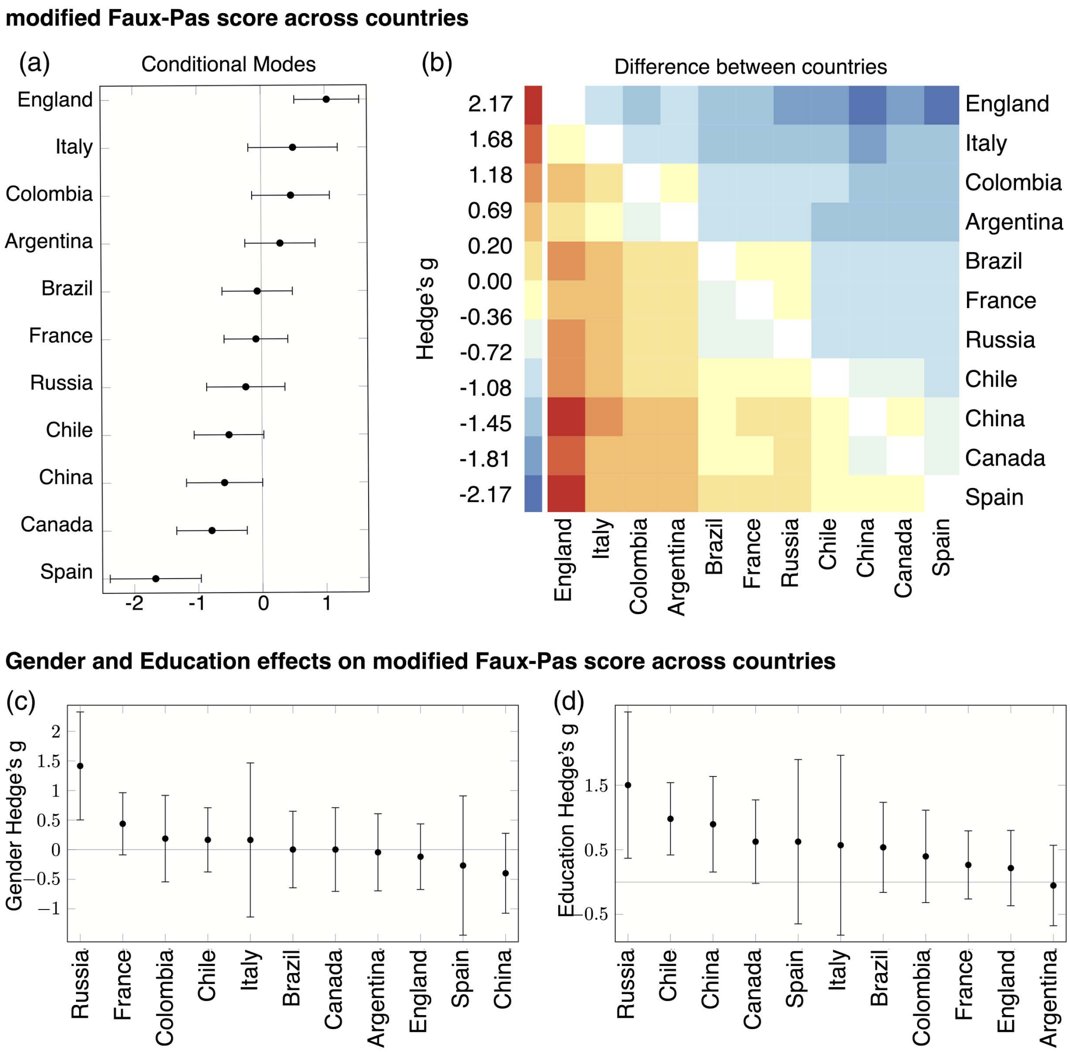
We separated the stories that actually contained a faux pas from the stories that did not, and computed the same linear mixed models as above on the scores resulting from these two groups of stimuli. We found that the country VPC was higher for the stories with a faux pas (VPC_FP = 23.8%) than for the stories without faux pas (NFP, no faux-pas; VPC_NFP = 11.6%), indicating that variations across countries are more important when there actually is a faux pas to identify.

Variation of Facial Emotion Recognition Scores Across Countries

The VPC indicates that 20.76%, 95% CI [8.26, 35.69], of the variance in emotion recognition scores can be attributed to differences among nationalities after controlling for age, gender, and education. Figure 3a represents the countries ranked according to their conditional modes, that is, the difference between the global average-predicted response in score and the response predicted for a particular country. As above, we also computed Hedges' g between every pair of countries, see heatmap Figure 3b (see also the Supplemental Material 7). To mitigate the effect of age (which had the biggest effect on FER scores) in Hedge's g calculations, only participants above 50 years old were considered for this analysis ($N = 432$, i.e., 74% of the participants). The countries' ranking was the same for conditional modes as for Hedge's g. Hedge's g between the first and last countries in emotion recognition score (England vs. Italy), was $g = 1.89$; 95% CI [0.81, 2.95]. Since gender had a significant effect on the FER score, we computed the gender effect size for each country (Figure 3c). It went from $g = 1.17$, 95% CI [0.37, 1.97] in Colombia to $g = -0.22$, 95% CI [−1.04, 0.60] in Russia, with positive values corresponding to higher scores for women.

Variations across countries for each emotion are illustrated in Figure 4, which represents country-specific mean performances for each emotion in the FER test. We fitted a linear mixed model for each emotion with age, gender, and education as fixed effects and country as a random effect: score_emotion ∼ age + gender + education + (1|country). To compare the effect of countries across the different emotions, we computed a VPC for each emotion. The

Figure 2
Mental State Inference Across Countries



Note. (a) Country conditional modes from a linear mixed model controlling for age, gender, and education. Higher conditional modes correspond to better performances. (b) Heatmap of the difference between every pair of countries measured with Hedge's *g*. Positive values correspond to a better performance of column countries compared to row countries. (c) Gender effect size in each country measured with Hedge's *g*. Positive values correspond to women scoring higher than men. (d) Education effect size in each country, measured with Hedge's *g*. Positive values correspond to a positive effect of education. Error bars correspond to 95% confidence intervals. See the online article for the color version of this figure.

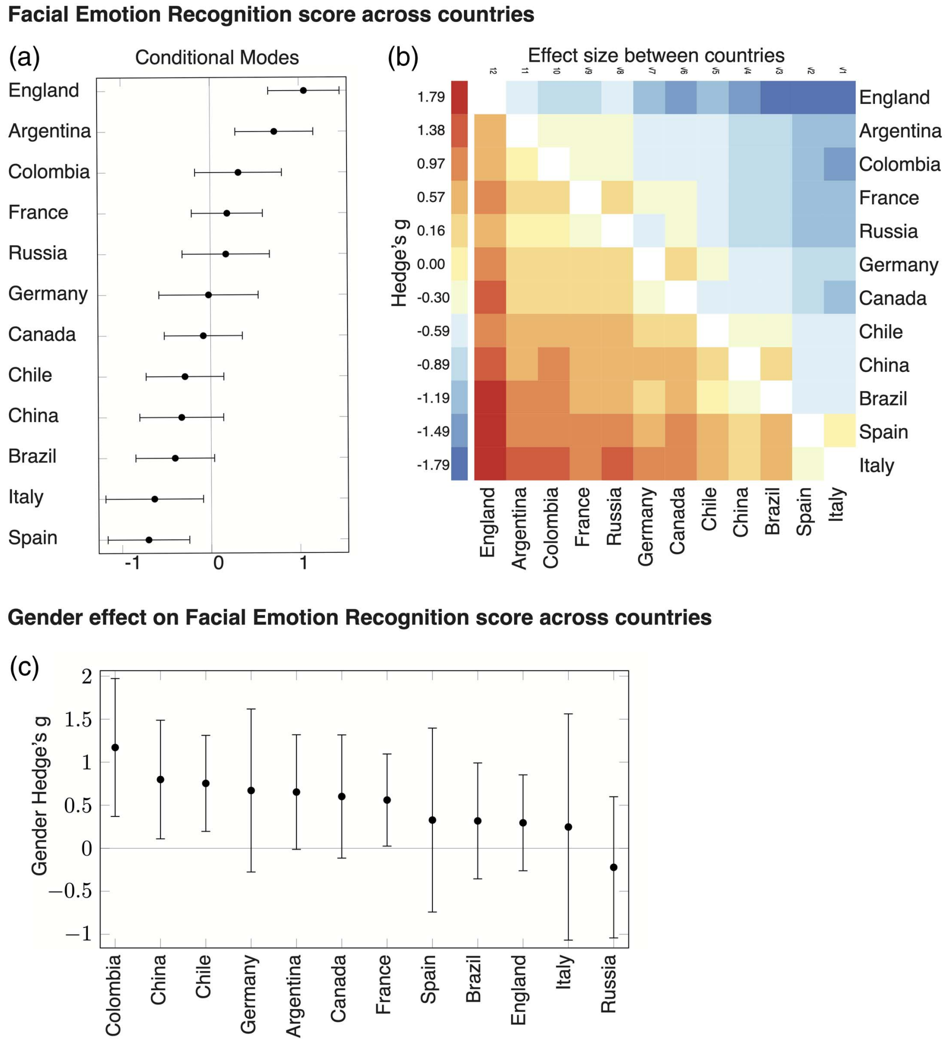
emotion with the largest country effect after controlling for age, gender, and education was fear, VPC = 21.42%, 95% CI [8.66, 37.43], followed by sadness, VPC = 15.31%, 95% CI [7.38, 28.59], anger, VPC = 12.73%, 95% CI [7.05, 25.01], and disgust, VPC = 9.09%, 95% CI [6.54, 20.03].

We also computed confusion tables for each country (Figure 5). The overall structure was similar across countries, but a number of differences can still be identified. For instance, fear items were often misclassified as surprise items, but the error rate widely varied across countries, from 25% in Germany to 50% in Canada. Italians misclassified sadness items as neutral 21% of the time, but it never happened in China.

Potential Confounding Factors

We first ruled out a potential effect of translation. Indeed, the original tests were developed in English, and their translation to other languages could have had an influence on performance. To test this, we added to the linear mixed model previously described "language" as a fixed effect: score ~ age + gender + education + language + (1|country). Language was a categorical variable corresponding to the language the test was translated into for each country. Interestingly, it seems that translations did not have a significant effect on FER performances, $t(559) = 1.06$, $p = .35$, or on faux pas detection performances, $t(531) = 0.74$, $p = .77$.

Figure 3
Facial Emotion Recognition Across Countries



Note. (a) Country conditional modes from a linear mixed model controlling for age, gender, and education. Higher conditional modes correspond to better performances. (b) Heatmap of the difference between every pair of countries measured with Hedge's *g*. Positive values correspond to a better performance of column countries compared to row countries. (c) Gender effect size in each country measured with Hedge's *g*. Positive values correspond to women scoring higher than men. Error bars correspond to 95% confidence intervals. See the online article for the color version of this figure.

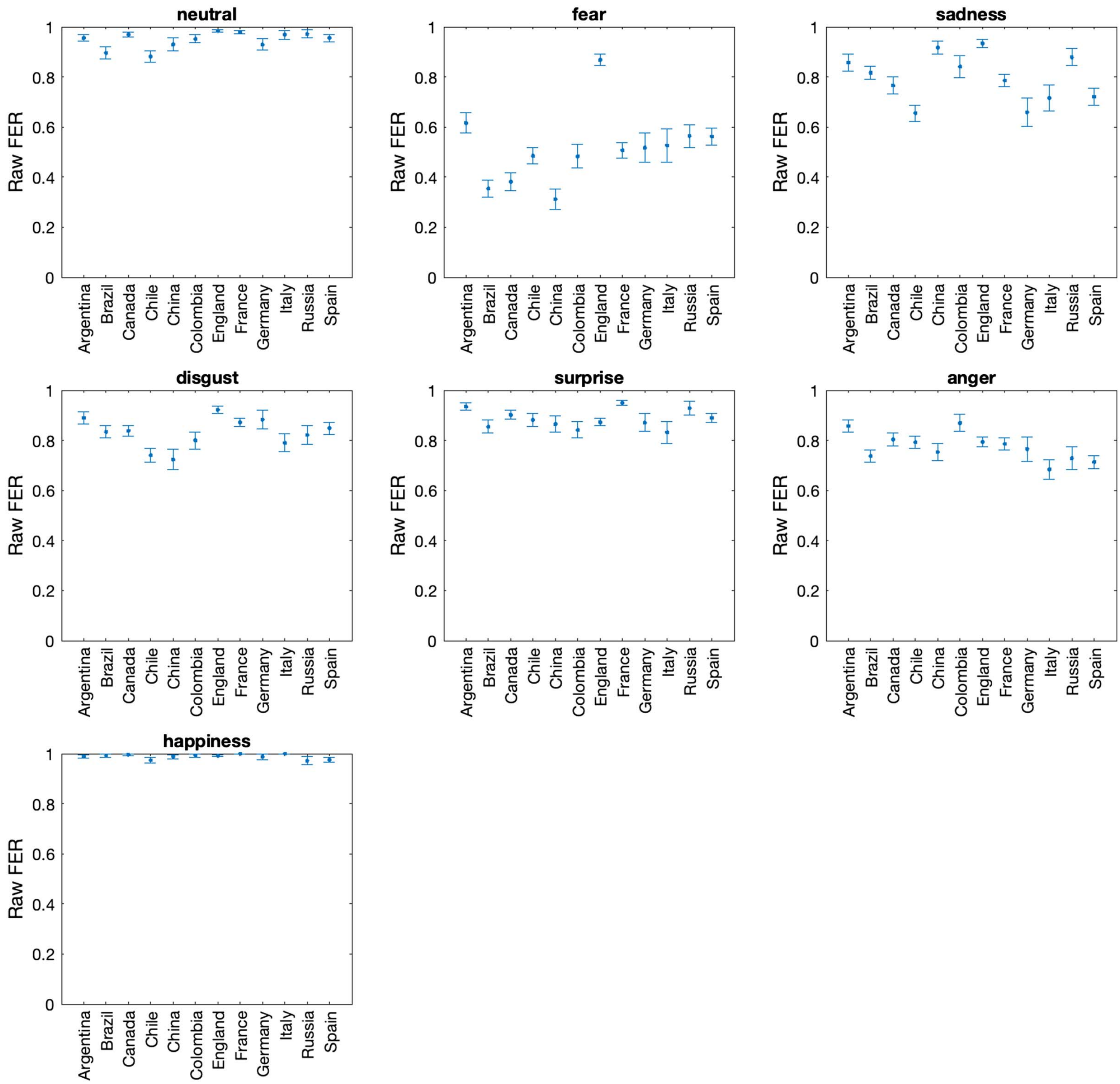
We also compared the variation in scores between countries to the variation in scores between the different centers within countries. We added to our initial linear mixed model a random effect for centers nested within countries: score ~ age + gender + education + (1|country) + (1|country:center). For FER, VPC_country = 11.70% and VPC_country:center = 10.33%. For faux pas detection, VPC_country = 16.29% and VPC_country:center = 9.00%. This shows that not only is there some variance between countries but also between centers within countries. Differences in the magnitude of these effects can however not be interpreted confidently in the present study.

Discussion

This study demonstrated that individual and cultural factors strongly impact measures of social cognition abilities. More specifically, it revealed a continuous decline across the adult lifespan in mentalizing and emotion recognition scores, as measured by

Figure 4

Country-Specific Mean Performance for Each Emotion at the Facial Emotion Recognition (FER) Test



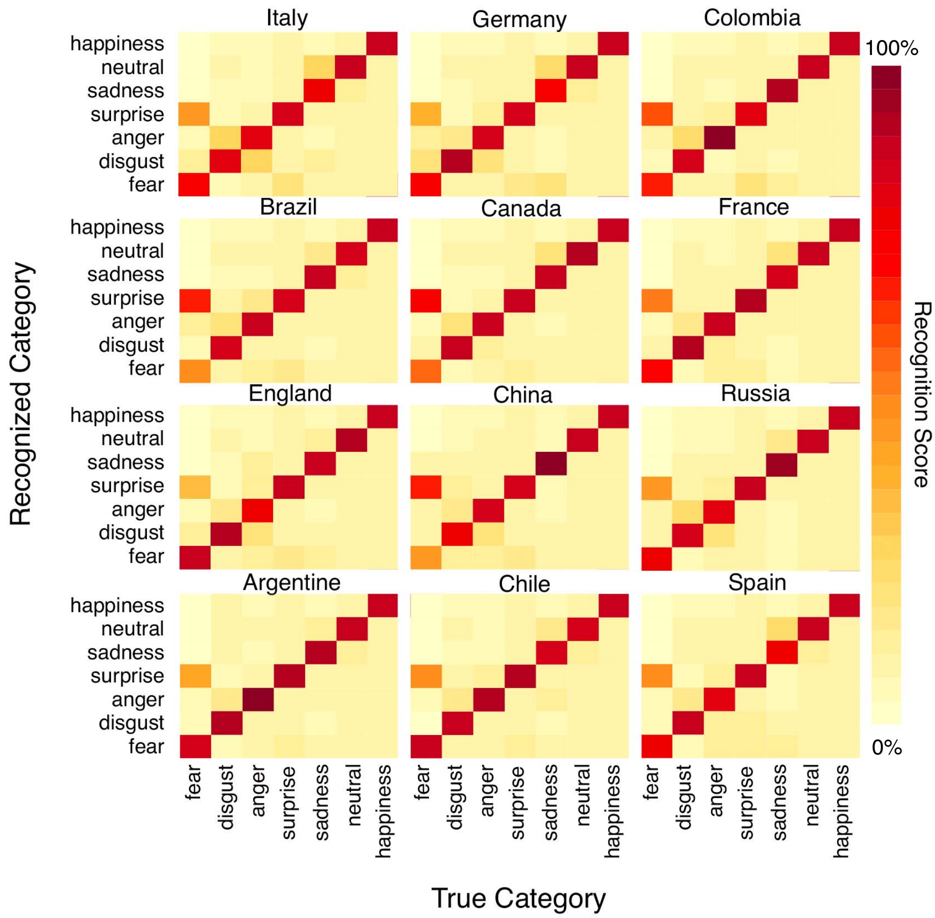
Note. See the online article for the color version of this figure.

classical neuropsychological tools. Education largely influenced the former, but not the latter. We also found that women outperformed men on both tasks, especially the FER. Finally, we observed important variations across countries on both tests' performances, revealing that classical measures of social cognition can drastically vary depending on the nationality of participants.

The impact of age observed here is congruent with previous cross-sectional studies in which older adults have consistently shown lower accuracy compared with younger adults when having to detect social faux pas (Halberstadt et al., 2011; Wang & Su, 2006) but also when labeling facial expressions of emotion (Ruffman et al., 2008, 2009). In the present case, however, only five out of our 18 test centers recruited participants younger than 40 and the majority of our participants were aged 50 years or older. The specific linear decline (and its early onset) observed in Figure 1 should then be interpreted with caution, especially as it might differ from other recent findings suggesting a delayed decline of FER over the lifespan (e.g., McDonald et al., 2018; Ruffman et al., 2008).

Figure 5

Confusion Tables for Emotion Recognition in Each Country



Note. Hotter colors correspond to higher recognition scores. See the online article for the color version of this figure.

Interestingly, in our study, the effects of age were independent of the participants' nationality, suggesting a universal decline of social cognitive abilities with age. Concerning the effect of education, the fact that we observed no effect on FER performance partially contrasts with previous findings (e.g., de Souza et al., 2018). However, in the study by de Souza et al., education was only associated with FER in Brazilian participants and not French. This apparent difference could be the consequence of cultural differences on the influence of education on FER or FER testing. In contrast, the advantage for individuals with higher education on the mFP seems to be common across a wide variety of countries. To our knowledge, as developmental studies of social cognition classically focus on childhood and are conducted within a single country (e.g., Filippova & Astington, 2008), we are the first to report such a finding.

Our study indicates that a gender effect on emotion recognition and social reasoning can be found across a wide variety of countries, consistently over the lifespan. In the literature, two contradictory views currently coexist (Lausen & Schacht, 2018). The "gender difference hypothesis" suggests that women have better socioemotional decoding skills than men, even if this advantage is relatively small according to meta-analyses (Kirkland et al., 2013; Kret & De Gelder, 2012). Conversely, the "gender similarity hypothesis" suggests that most of the time, gender differences only reflect experimental artifacts and concludes that in many cases women and men are rather similar on most psychological dimensions (Baez et al., 2017; Hyde, 2014). Our study, based on a large sample and relying on traditional tasks of social cognition, indicates that a gender effect on emotion recognition and social reasoning is, indeed, found across a wide variety of countries (see also Merten, 2005, for similar findings), consistently over the lifespan. In addition, major variations of the magnitude of gender differences across countries were observed in our data. As previously underlined (Eisenberg et al., 1998), the meanings of emotion expression and appropriateness of behaviors are largely shaped by socialization since early childhood, through parental reactions and expectations that are gender-differentiated (Denham et al., 2007). In a binary vision of gender, women are expected to be better social decoders than men (Graham & Ickes, 1997), it is thus probable that this difference in results reflects some higher motivational aspects (Ickes et al., 2000) along with compliance to cultural gender roles (Baez et al., 2017). In future studies, the gender identity of the participants should be asked more explicitly in order to go beyond the binary opposition between women and men and deepen our understanding of the role of gender in social cognition performance. Also, the inclusion of a larger range of countries would provide the opportunity to investigate the determinants of such gender differences across countries (e.g., correlation with national "Gender Gap Index," see Coutrot et al., 2018).

The most striking findings of our study are the variations across countries observed after controlling for age, gender and education. In our population, the proportion of variance explained by the country of origin was indeed considerable, reaching 24.52% on the modified faux pas scores, and 20.76% on emotion recognition scores. This is almost 10 times higher than the effects reported in previous large-scale cross-cultural studies focusing on memory and attention (Hayden et al., 2014) or on spatial navigation abilities (Coutrot et al., 2018). This underlines that, more than other neurocognitive tasks, social stimuli and measures—often relying on cultural norms or utilizing words with multiple meanings—are highly dependent on local specificities. In our study, these variations cannot be explained by translation issues as both between-countries analyses controlling for the languages in which the tests are administered, and within-countries multisite comparisons, indicated that the languages in which the tests were administered did not have an effect on the tests' performance. This specific result is of particular significance as a major limitation in previous studies was based on the fact that it was not possible to isolate cultural variations from the impact of material translations in comparisons over two or three countries, although language and social cognition are extensively related (Fiedler, 2008). In the present study, between-country and between-center differences were not directly comparable and both were found to explain a significant proportion of variance. Moreover, the origin of between-center variations remains quite speculative at this stage and could reflect within-country cultural specificities (e.g., Beaupré & Hess, 2005) or be the mere consequence of practical aspects (e.g., clinicians' influence). Future studies with larger samples are needed to deepen understanding of the cultural differences reported here. The total number of variables (5×6 items for the faux pas items with faux pas + 5 items for the faux pas items without faux pas + 10 control questions + 35 items for the picture of facial affect test) recorded in our study exceeded the number of participants in most centers. Consequently, it was not possible to conduct finer-grained analyses that would reveal which items are differentially failed. Nor was it possible to identify the factorial structure of the supposed underlying psychological constructs across countries (and whether such structure is consistent across centers within a given country). This constraint is why we focused on the tests' total score. For the same reasons, it was not possible to directly test for specific item bias (Van de Vijver & Leung, 1997). However, we did check the typical patterns of responses to allow a more qualitative reading of cultural difference, which we briefly discuss in the following lines.

Regarding the ability to mentalize, participants from different countries have already been compared in past research taking a cross-cultural perspective. Wu and Keysar (2007) have shown that cultural differences induce different profiles of perspective-taking, either favoring other people's (Chinese participants) or self-related (U.S. participants) perspectives. In another study using the faux pas test, Malaysian participants were found to have lower performance than participants from the U.K. (Yong et al., 2021). Variations in the timing of false-belief development have also been observed across Canada, China (mainland), Hong-Kong, and USA, ranging up to 2 years' difference between countries (Liu et al., 2008). While variations across countries could be expected, the present study is, however, the first to demonstrate a cultural variation on the faux-pas test through a large-scale multicentric standardized and objective assessment. Interestingly, the best performances were obtained by participants from England, which is the country in which the test has been developed. As identifying a faux pas requires detecting that an implicit social rule has been broken, our results question the applicability of the test in other countries as it is easy to conceive that social rules fluctuate from one country to another. As pointed out by Van de Vijver (2019), norm-driven adaptations should be used to accommodate cultural differences in norms, values, and practices as specific cultural contexts may not apply in all countries. As social norms not only drive one's actions but also others' expectations of one's actions, these cultural differences also have the potential to modulate the inference of mental states per se. As an example from the task, mistaking a customer for a waiter in a restaurant unequivocally constitutes a faux pas for all participants (100%) from England, but not for Canadians (65.4%). In the same vein, 21.2% of Chinese participants considered that it was a faux pas to give up a seat to an older passenger while riding on a city bus, although 100% of English participants considered this was a normal behavior. Although this very last example was a control item (i.e., originating from a story with no faux pas), we found that variations across countries were stronger when there was actually a faux pas to identify. Altogether, the variations reported here should invite neuropsychological experts to develop concept-driven adaptations of the mFP test rather than literal translations (which is currently the case).

Regarding emotion recognition, cross-country comparisons revealed that some emotions were consensually labeled by all participants, while for others, cross-country variations existed. Specifically, happy expressions were unambiguously categorized as such in every country (see Nelson & Russell, 2013 for congruent results). By contrast, the recognition of negative items of the FER test appeared to be more culture-dependent and this was particularly the case for the Fear items. Many factors can account for these specificities, both at the practical and theoretical levels. First, as most of the items of the FER test are negative, it is not surprising to observe higher variations for fearful or disgusted faces than for happy faces, especially as a ceiling effect is often observed for the latter. For example, within-countries classification patterns revealed that the majority of Brazilian and Canadian participants confused "fear" and "surprise" faces, a bias that was almost absent among English and Argentinian participants. Second, fear has classically been found to be more difficult to identify compared to other basic emotions in a variety of cultural groups (Beaupré & Hess, 2005; McAndrew, 1986). Given the similarities in their visual configuration, the "attentional limitation hypothesis" has been formulated to explain the confusion between "fear" and "surprise" faces (Chamberland et al., 2017), but cross-cultural validation of this hypothesis is lacking. As such confusion could be observed in some countries and not in others, our findings would rather contradict the attentional limitation hypothesis, but our study was not designed to assess its validity. Third, from a cultural perspective, lower recognition accuracy for negative emotions has been associated with cultural rules against displaying and acknowledging these emotions in others, which is hypothetically prevalent in some Asian cultures (e.g., Biehl et al., 1997; Matsumoto, 1990, 1992; McAndrew, 1986; Russell et al., 1993).

There is a long tradition of research regarding the cultural variations of emotion recognition. In contrast with the widespread conception of the universal nature of a few basic emotion expressions (e.g., Darwin, 1872; Ekman & Friesen, 1971), the influence of cultural factors has been acknowledged—though minimized—from the beginning of this field of research (e.g., Ekman & Friesen, 1969; Klineberg, 1938) and cross-national differences have been reported many decades ago (e.g., Biehl et al., 1997; Matsumoto, 1992) revealing decoding strategies that vary across cultures (Matsumoto, 1990). Meta-analyses have shown that people are better at recognizing facial emotions from their own cultures rather from other cultures (Elfenbein & Ambady, 2002). Cross-country explorations including both a large range of countries and the use of standardized tools, such as the present study, have, however, rarely been done in the past. In most previous work the comparison was restricted to a few countries and opposed "western" (i.e., WEIRD) versus "Eastern" groups (Elfenbein & Ambady, 2002; Jack et al., 2009, 2012) which therefore limits the interpretations to a rough dichotomy rather than to the existence of underlying socio psychological variables (e.g., "display rules," Matsumoto & Hwang, 2019). These "western" versus "eastern" comparisons have, however, helped to point out several key cultural differences undermining the universality of the six basic emotions that was previously assumed (Ekman & Friesen, 1971). For example, a neat differentiation of emotion categories was only observed in WEIRD but not "Eastern" (Chinese) perceivers (Jack et al., 2012). A number of experiments also demonstrated that the emotion labels (e.g., anger, fear) given to participants in traditional FER paradigms provide a context for emotion perception and participate in constructing the perceptual representations of facial emotions presented (Gendron et al., 2012). In addition, the representational structure of emotion expressions in visual face-processing brain regions has been found to be predicted by intracultural conceptual similarity between emotions (Brooks et al., 2019). Linguistic and clinical investigations have respectively shown that emotion conceptual knowledge, which has different patterns of association in different language families (Jackson et al., 2019), is critically involved in the recognition of facial expression (Bertoux et al., 2020). Along with these studies, our current findings suggest that conceptual knowledge about emotions scaffolds FER. The variability retrieved here with and without WEIRD participants and the country-specific misclassifying biases support this constructionist perspective of emotion recognition (Barrett & Satpute, 2019). According to this perspective, local and culture-dependent conceptual knowledge not only provides labels to describe the perception of emotions but also shapes it (Lindquist et al., 2015). Using a test conceived by authors from the USA to assess FER through a forced choice methodology may have favored the cultural variations we observed in our study, as the test requires selection within discrete category (thus favoring performance of WEIRD participants) to describe stereotypic emotions which are further acted (and not expressed) by White American people only.

Although we ranked countries according to their performance in the tests employed, it seems important to specify that we never considered or hypothesized the superiority of one country over another. Our findings do not show that English participants are better at understanding faux pas, but do show that they have a better performance in a task that totally fits their culture, as the aforementioned faux pas originated from English culture. While being commonly employed tests of social cognition abilities, both tests considered here originated from and were validated in specific local contexts, at specific periods, before being employed all over the world. Taken together, our findings suggest that either in supposedly "low-level" (e.g., emotion recognition) or "higher-level" (e.g., mentalizing and social norms decoding) mechanisms, social cognition measures are shaped by individual, as well as by contextual, factors. Beyond the effect of age, gender, and education, categorizing emotions as well as inferring others' intentions or beliefs to explain others' behaviors may be reasonably influenced by local concepts, norms and habits, and thus be differentially apprehended across countries. This raises questions regarding good practices in neuropsychology and human neurosciences in general as currently, a universalist approach involving the negligence of possible cultural variations is rather favored in order to generalize the results (Bauer, 2020). We believe that the question of the generalizability of findings is a pressing and consequential problem in neurosciences (Yarkoni, 2020). Generalizability is applied to many observed results in order to be more appealing and disseminated to a wider audience. What is observed in a given—often small—sample is generalized to the entire humanity. This led us to build most neuroscientific knowledge on observations based on WEIRD participants, commonly performed by WEIRD researchers (Barrett, 2020). However, by highlighting important variations across countries in classical measures of cognition, our study and others (e.g., Henrich et al., 2010b) underline a strong limit of this approach. Therefore, the field needs to engage in more epistemological and critical thinking in order to identify and overcome the societal or cultural biases and systemic constraints that impact the design of studies and dissemination of results. As larger-scale, multicultural studies are also needed to counterbalance this effect, we call states, foundations, and charities to better fund these costly initiatives and to favor international collaborations similar to the INSCD. In this vein, groups such as the European Consortium for Cross-Cultural Neuropsychology (ECCroN) could help prioritize the validation of cross-cultural tests or favor training of practitioners regarding cultural differences. Recent initiatives aiming to harmonize neuropsychological assessment across states or countries (e.g., Boccardi et al., 2022), and especially those centered on social cognition, such as the clinical use of SocIal coGNnition measures for the AssessmenT of neURocognitivE disorders (SIGNATURE Initiative), should also take into consideration these cultural variations. Within our network, future studies will investigate cross-cultural validity of the tests as well as cross-country comparisons of the brain regions involved in social cognition tasks.

Finally, we believe that our findings could also have important clinical impact. The detection and treatment of cognitive disorders, indeed, represents one of the biggest challenges in the fields of neurology, psychiatry, and developmental disorders (Alzheimer's Disease International, 2013; Mackin & Areán, 2009; McWhirter et al., 2020). With a rising awareness of these disorders or atypical functioning in the general population, higher survival rates of children and elders and the generalization of better diagnosis strategies, their prevalence is expected to increase in the next decades. The strong impact of social and cultural factors observed here in tasks that are among the most used measures of social cognition (Bertoux et al., 2012; Eddy, 2019; Henry et al., 2016; Quesque & Rossetti, 2020) lead us to argue that a mature neuroscience aiming to have relevance to clinical neuropsychology, psychiatry, neurology or geriatry can no longer ignore these factors. This is particularly true as our findings were observed in control participants without any cognitive decline or complaint. This indicates that if a test is validated in one country, it cannot be automatically used across the world for assessment or diagnosis purposes (Daugherty et al., 2017). In the absence of local validation, clinical studies should, thus, include a control group so that a proper local reference is available to assess the local test version's sensitivity or specificity. International studies, such as those favored by modern consortia (e.g., ENIGMA, Human Brain Project, etc.), should also perform cross-cultural comparison prior to any clinical groups contrast (controls vs. patients) to quantify any cultural effect, or should at least consider the countries of recruitment as a nuisance covariate during the analyses. Other good practices would involve the development of non-gender-biased instructions and items designed to limit the influence of motivational aspects and gender-expectations on social cognitive tasks. Similarly, setting universality assumptions aside requires innovation in measurement (Gendron, 2017). For cross-cultural use of a measurement tool, concept-driven adaptations rather than literal translation (which constitutes, by far, the default practice) is critical as this ensures that the tool relates to the social context actually encountered by people within their culture (Mehta et al., 2011). Going even further, before developing a new cognitive test, one should first probe whether the underlying principles of the task actually fit to the culture of the targeted population. For example, asking participants to categorize facial expressions into different categories of emotions might not be suitable among people that would not spontaneously use emotional terms (e.g., "he is angry") to describe such material but rather use physical descriptions (e.g., "he frowns," see Gendron et al., 2014a, 2014b). More generally, when adapting a task implying forced-choice response, researchers should not, a priori, define the possible responses but rather consider what types of attributions are spontaneously made by perceivers. Given the lead social sciences have taken in exploring and considering cultural diversity, we also call for richer and stronger interactions between neurosciences and the humanities, which is a necessary path toward higher rigor, relevance, and inclusivity.

In terms of study limitations, we acknowledge that the present project only represents a first, necessary step toward the development of more valid theories and tools to quantify social cognition abilities. Although our study took place in many different countries in Asia, Europe, North and South America as well as Russia, we did not include participants from other regions (e.g., Africa or Middle East). We hope centers from these regions and beyond are willing to join our network so that our next study of cultural variation will be more comprehensive. In addition, although we succeeded in including several centers from some countries, this was not always possible. In vast countries such as Canada, China, and Russia, only one center formed part of the study, although within-country variations could be expected. Future studies should take this issue into consideration as well. Given the potential influence of a huge number of sociodemographic and geographical indicators that are not considered in the present project, our operationalization of culture appears to be very basic, as we equated geographical location with culture. Of course, we acknowledge that culture is much more complex than the set of shared representations, concepts, and norms within a country, but we do believe that it was the best way to approach cultural variations in our current data set. Furthermore, this is a limit our study shared with most studies in neuropsychology. The same concern could be raised regarding individual factors, as we limited our investigations to investigating the influences of age, gender, and the level of formal education, without being able to control for these factors entirely. Importantly, given that most (75%) of our participants were 50 years or older, we cannot guarantee that our findings would be strictly similar in a younger population. One hypothesis is that such variations may be lower in younger participants as cognitive decline in older participants could explain the variations we observed. However, in our study, all participants had normal cognitive efficiency, no cognitive complaint and no psychiatric/neurologic antecedents. Most of our participants had additional screening prerequisites, such as a normal MRI (see Supplemental Materials 1 and 2). A third of our participants had comprehensive neurological and neuropsychological examinations showing no cognitive decline or any abnormality. We therefore think it is unlikely that cognitive decline could have played a role in the variations we observed, especially as in our study, age-related variations started to be observed from the 20's and were constant across the life span. An alternative hypothesis would be that we might expect larger variations in younger participants' performance given that the tests we used were not recent and may be culturally less-adapted to a younger cohort. Regarding the other individual factors, we acknowledge that we retained a binary approach of gender in our study (i.e., men/women) and think that future studies should cover other gender identities. In addition, while we had a satisfactory gender ratio in the total population (with 60% women and 40% men), it was unbalanced toward women in some centers. This makes some specific results difficult to interpret, as for example, the difference in the magnitude of the gender effect in Russia between the mFP and FER only relies on the contrast between 11 men and 18 women. Finally, many other factors, interacting with cultural ones, could impact cognitive performance and its measurement. In this line, a questionnaire such as the one included in RedLat, exploring many other psycho-sociological dimensions (e.g., past and present level of material ease, psychological or physical violence endured during life . . .) could be a useful example to follow. As an opportunity to develop such research trajectories, INSCD represents an important collaborative initiative and seeks to involve new partners. Collecting and sharing data from a larger sample of countries could ultimately lead to the opportunity to study the links between countries' cultural distances estimates (Muthukrishna et al., 2020), given that traditional cultural indicators (e.g., Hofstede's dimensions) have been questioned (see McSweeney, 2002; Oyserman et al., 2002). Another promising direction for future studies lies in large-scale cross-cultural comparisons of the brain substrates involved while performing classical social cognition tasks, as opposed to classical studies contrasting participants from two cultures (Adams et al., 2010; Kobayashi et al., 2006). So far, the only comparisons that have been done regarding mentalizing performance have yielded inconsistent results, either finding a high level of consistency between participants from different countries (e.g., Adams et al., 2010) or culture-dependent brain correlates (Kobayashi et al., 2006). Assessing such correlates in a larger data set could reveal the existence of culture-specific neural involvement in measures that might be linked to the variations of behavioral performances. More generally, we believe that the interest in large-scale and multicentric research programs should be generalized to all subfields of neuropsychology. We are aware that this would only constitute a first step toward better practices. A paradigm shift remains highly needed, and we hope that such collective initiatives will encourage the emergence of more representative theories and measurements of human cognition in all its diversity.

References

Adams, R. B., Jr., Rule, N. O., Franklin, R. G., Jr., Wang, E., Stevenson, M. T., Yoshikawa, S., Nomura, M., Sato, W., Kveraga, K., & Ambady, N. (2010). Cross-cultural reading the mind in the eyes: An fMRI investigation. *Journal of Cognitive Neuroscience*, *22*(1), 97–108. https://doi.org/10.1162/jocn.2009.21187

Alzheimer's Disease International. (2013). *Policy brief for heads of government: The global impact of dementia 2013–2050*.

Arnett, J. J. (2016). The neglected 95%: Why American psychology needs to become less American. *American Psychologist*, *63*(7), 602–614. https://doi.org/10.1037/0003-066X.63.7.602

Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J.-F., & Rahwan, I. (2018). The moral machine experiment. *Nature*, *563*(7729), 59–64. https://doi.org/10.1038/s41586-018-0637-6

Baez, S., Flichtentrei, D., Prats, M., Mastandueno, R., García, A. M., Cetkovich, M., & Ibáñez, A. (2017). Men, women … who cares? A population-based study on sex differences and gender roles in empathy and moral cognition. *PLOS ONE*, *12*(6), Article e0179336. https://doi.org/10.1371/journal.pone.0179336

Barrett, H. C. (2020). Towards a cognitive science of the human: Cross-cultural approaches and their urgency. *Trends in Cognitive Sciences*, *24*(8), 620–638. https://doi.org/10.1016/j.tics.2020.05.007

Barrett, L. F., & Satpute, A. B. (2019). Historical pitfalls and new directions in the neuroscience of emotion. *Neuroscience Letters*, *693*, 9–18. https://doi.org/10.1016/j.neulet.2017.07.045

Bauer, P. (2020). Expanding the reach of psychological science. *Psychological Science*, *31*(1), 3–5. https://doi.org/10.1177/0956797619898664

Beaupré, M. G., & Hess, U. (2005). Cross-cultural emotion recognition among Canadian ethnic groups. *Journal of Cross-Cultural Psychology*, *36*(3), 355–370. https://doi.org/10.1177/0022022104273656

Bendix, R. (1963). Concepts and generalizations in comparative sociological studies. *American Sociological Review*, *28*(4), 532–539. https://doi.org/10.2307/2090069

Bergeron, D., Flynn, K., Verret, L., Poulin, S., Bouchard, R. W., Bocti, C., Fülöp, T., Lacombe, G., Gauthier, S., Nasreddine, Z., & Laforce, R. J. (2017). Multicenter validation of an MMSE-MoCA conversion table. *Journal of the American Geriatrics Society*, *65*(5), 1067–1072. https://doi.org/10.1111/jgs.14779

Bertoux, M., Delavest, M., de Souza, L. C., Funkiewiez, A., Lépine, J. P., Fossati, P., Dubois, B., & Sarazin, M. (2012). Social cognition and emotional assessment differentiates frontotemporal dementia from depression. *Journal of Neurology, Neurosurgery, and Psychiatry*, *83*(4), 411–416. https://doi.org/10.1136/jnnp-2011-301849

Bertoux, M., Duclos, H., Caillaud, M., Segobin, S., Merck, C., de La Sayette, V., Belliard, S., Desgranges, B., Eustache, F., & Laisney, M. (2020). When affect overlaps with concept: Emotion recognition in semantic variant of primary progressive aphasia. *Brain: A Journal of Neurology*, *143*(12), 3850–3864. https://doi.org/10.1093/brain/awaa313

Bertoux, M., Funkiewiez, A., O'Callaghan, C., Dubois, B., & Hornberger, M. (2013). Sensitivity and specificity of ventromedial prefrontal cortex tests in behavioral variant frontotemporal dementia. *Alzheimer's & Dementia*, *9*(5, Suppl.), S84–S94. https://doi.org/10.1016/j.jalz.2012.09.010

Biehl, M., Matsumoto, D., Ekman, P., Hearn, V., Heider, K., Kudoh, T., & Ton, V. (1997). Matsumoto and Ekman's Japanese and Caucasian Facial Expressions of Emotion (JACFEE): Reliability data and cross-national differences. *Journal of Nonverbal Behavior*, *21*(1), 3–21. https://doi.org/10.1023/A:1024902500935

Boccardi, M., Monsch, A. U., Ferrari, C., Altomare, D., Berres, M., Bos, I., Buchmann, A., Cerami, C., Didic, M., Festari, C., Nicolosi, V., Sacco, L., Aerts, L., Albanese, E., Annoni, J. M., Ballhausen, N., Chicherio, C., Démonet, J. F., Descloux, V., … the Consortium for the Harmonization of Neuropsychological Assessment for Neurocognitive Disorders. (2022). Harmonizing neuropsychological assessment for mild neurocognitive disorders in Europe. *Alzheimer's & Dementia*, *18*(1), 29–42. https://doi.org/10.1002/alz.12365

Boone, K. B., Victor, T. L., Wen, J., Razani, J., & Pontón, M. (2007). The association between neuropsychological scores and ethnicity, language, and acculturation variables in a large patient population. *Archives of Clinical Neuropsychology*, *22*(3), 355–365. https://doi.org/10.1016/j.acn.2007.01.010

Brislin, R. W., Lonner, W. J., & Thorndike, R. (1973). *Cross-cultural research methods*. Wiley.

Brooks, J. A., Chikazoe, J., Sadato, N., & Freeman, J. B. (2019). The neural representation of facial-emotion categories reflects conceptual structure. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(32), 15861–15870. https://doi.org/10.1073/pnas.1816408116

Chamberland, J., Roy-Charland, A., Perron, M., & Dickinson, J. (2017). Distinction between fear and surprise: An interpretation-independent test of the perceptual-attentional limitation hypothesis. *Social Neuroscience*, *12*(6), 751–768. https://doi.org/10.1080/17470919.2016.1251964

Clarens, M. F., Crivelli, L., Martín, M. E., Fernández, R., Martínez Canyazo, C., Arruabarrena, M., Tabernero, M. E., Cervino, C., Varela, Y., Prestupa, R., Pereyra, L., Rossi, F., Sarasola, D. R., & Allegri, R. F. (2021). Baremos en español del Mini-SEA (Mini Social Cognition and Emotional Assessment) en adultos normales de Buenos Aires [Spanish Norms for Mini-SEA (Mini Social Cognition and Emotional Assessment) in adults in Buenos Aires]. *Vertex (Buenos Aires, Argentina)*, *XXXII*(152), 5–10. https://europepmc.org/article/med/34783788

Collins, M. (1980). Interviewer variability: A review of the problem. *Journal of the Market Research Society*, *22*(2), 77–95.

Cotter, J., Granger, K., Backx, R., Hobbs, M., Looi, C. Y., & Barnett, J. H. (2018). Social cognitive dysfunction as a clinical marker: A systematic review of meta-analyses across 30 clinical conditions. *Neuroscience and Biobehavioral Reviews*, *84*, 92–99. https://doi.org/10.1016/j.neubiorev.2017.11.014

Coutrot, A., Silva, R., Manley, E., de Cothi, W., Sami, S., Bohbot, V. D., Wiener, J. M., Hölscher, C., Dalton, R. C., Hornberger, M., & Spiers, H. J. (2018). Global determinants of navigation ability. *Current Biology*, *28*(17), 2861–2866.e4. https://doi.org/10.1016/j.cub.2018.06.009

Darwin, C. (1872). *The expression of the emotions in man and animals*. John Murray. https://doi.org/10.1037/10001-000

Daugherty, J. C., Puente, A. E., Fasfous, A. F., Hidalgo-Ruzzante, N., & Pérez-Garcia, M. (2017). Diagnostic mistakes of culturally diverse individuals when using North American neuropsychological tests. *Applied Neuropsychology: Adult*, *24*(1), 16–22. https://doi.org/10.1080/23279095.2015.1036992

Davidov, E., Schmidt, P., Billiet, J., & Meuleman, B. (Eds.). (2018). *Cross-cultural analysis: Methods and applications*. Routledge. https://doi.org/10.4324/9781315537078

de Souza, L. C., Bertoux, M., de Faria, Â. R. V., Corgosinho, L. T. S., Prado, A. C. A., Barbosa, I. G., Caramelli, P., Colosimo, E., & Teixeira, A. L. (2018). The effects of gender, age, schooling, and cultural background on the identification of facial emotions: A transcultural study. *International Psychogeriatrics*, *30*(12), 1861–1870. https://doi.org/10.1017/S1041610218000443

Denham, S. A., Bassett, H. H., & Wyatt, T. (2007). The socialization of emotional competence. In J. E. Grusec & P. D. Hastings (Eds.), *Handbook of socialization: Theory and research* (pp. 614–637). Guilford Press.

Ducharme, S., Dols, A., Laforce, R., Devenney, E., Kumfor, F., van den Stock, J., Dallaire-Théroux, C., Seelaar, H., Gossink, F., Vijverberg, E., Huey, E., Vandenbulcke, M., Masellis, M., Trieu, C., Onyike, C., Caramelli, P., de Souza, L. C., Santillo, A., Waldö, M. L., . . . Pijnenburg, Y. (2020). Recommendations to distinguish behavioural variant frontotemporal dementia from psychiatric disorders. *Brain: A Journal of Neurology*, *143*(6), 1632–1650. https://doi.org/10.1093/brain/awaa018

Eddy, C. M. (2019). What do you have in mind? Measures to assess mental state reasoning in neuropsychiatric populations. *Frontiers in Psychiatry*, *10*, Article 425. https://doi.org/10.3389/fpsyt.2019.00425

Eisenberg, N., Cumberland, A., & Spinrad, T. L. (1998). Parental socialization of emotion. *Psychological Inquiry*, *9*(4), 241–273. https://doi.org/10.1207/s15327965pli0904_1

Ekman, P., & Friesen, W. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica*, *1*(1), 49–98. https://doi.org/10.1515/semi.1969.1.1.49

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, *17*(2), 124–129. https://doi.org/10.1037/h0030377

Ekman, P., & Friesen, W. V. (1976). *Pictures of facial affect*. Consulting Psychologists Press.

El Grabli, F., Quesque, F., Borg, C., Witthöft, M., Michael, G. A., Lucas, C., Pasquier, F., Lebouvier, T., & Bertoux, M. (2022). Interoception and social cognition in chronic low back pain: A common inference disturbance? An exploratory study. *Pain Management*, *12*(4), 471–485. https://doi.org/10.2217/pmt-2021-0090

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, *128*(2), 203–235. https://doi.org/10.1037/0033-2909.128.2.203

Fiedler, K. (2008). Language: A toolbox for sharing and influencing social reality. *Perspectives on Psychological Science*, *3*(1), 38–47. https://doi.org/10.1111/j.1745-6916.2008.00060.x

Filippova, E., & Astington, J. W. (2008). Further development in social reasoning revealed in discourse irony understanding. *Child Development*, *79*(1), 126–138. https://doi.org/10.1111/j.1467-8624.2007.01115.x

Gendron, M. (2017). Revisiting diversity: Cultural variation reveals the constructed nature of emotion perception. *Current Opinion in Psychology*, *17*, 145–150. https://doi.org/10.1016/j.copsyc.2017.07.014

Gendron, M., Lindquist, K. A., Barsalou, L., & Barrett, L. F. (2012). Emotion words shape emotion percepts. *Emotion*, *12*(2), 314–325. https://doi.org/10.1037/a0026007

Gendron, M., Roberson, D., van der Vyver, J. M., & Barrett, L. F. (2014a). Cultural relativity in perceiving emotion from vocalizations. *Psychological Science*, *25*(4), 911–920. https://doi.org/10.1177/0956797613517239

Gendron, M., Roberson, D., van der Vyver, J. M., & Barrett, L. F. (2014b). Perceptions of emotion from facial expressions are not culturally universal: Evidence from a remote culture. *Emotion*, *14*(2), 251–262. https://doi.org/10.1037/a0036052

Gong, X., Huang, Y. X., Wang, Y., Luo, Y. J. (2011). Revision of the Chinese facial affective picture system. *Chinese Mental Health*, *25*(1), 40–46.

Goodman, Z. T., Llabre, M. M., González, H. M., Lamar, M., Gallo, L. C., Tarraf, W., Perreira, K. M., López-Cevallos, D. F., Vásquez, P. M., Medina, L. D., Perera, M. J., Zeng, D., & Bainter, S. A. (2021). Testing measurement equivalence of neurocognitive assessments across language in the Hispanic Community Health Study/Study of Latinos. *Neuropsychology*, *35*(4), 423–433. https://doi.org/10.1037/neu0000725

Graham, T., & Ickes, W. (1997). When women's intuition isn't greater than men's. In W. Ickes (Ed.), *Empathic accuracy* (pp. 117–143). Guilford Press.

Greenwood, J. D. (2004). What happened to the "social" in social psychology? *Journal for the Theory of Social Behaviour*, *34*(1), 19–34. https://doi.org/10.1111/j.1468-5914.2004.00232.x

Gwinnutt, J. M., Toyoda, T., Jeffs, S., Flanagan, E., Chipping, J. R., Dainty, J. R., Mioshi, E., Hornberger, M., & MacGregor, A. (2021). Reduced cognitive ability in people with rheumatoid arthritis compared with age-matched healthy controls. *Rheumatology Advances in Practice*, *5*(2), Article rkab044. https://doi.org/10.1093/rap/rkab044

Halberstadt, J., Ruffman, T., Murray, J., Taumoepeau, M., & Ryan, M. (2011). Emotion perception explains age-related differences in the perception of social gaffes. *Psychology and Aging*, *26*(1), 133–136. https://doi.org/10.1037/a0021366

Hayden, K. M., Makeeva, O. A., Newby, L. K., Plassman, B. L., Markova, V. V., Dunham, A., Romero, H. R., Melikyan, Z. A., Germain, C. M., Welsh-Bohmer, K. A., Roses, A. D., & the TOMSK-DUKE Study Group Investigators. (2014). A comparison of neuropsychological performance between US and Russia: Preparing for a global clinical trial. *Alzheimer's & Dementia*, *10*(6), 760–768.e1. https://doi.org/10.1016/j.jalz.2014.02.008

Henrich, J., Heine, S. J., & Norenzayan, A. (2010a). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Henrich, J., Heine, S. J., & Norenzayan, A. (2010b). Most people are not WEIRD. *Nature*, *466*(7302), 29. https://doi.org/10.1038/466029a

Henry, J. D., von Hippel, W., Molenberghs, P., Lee, T., & Sachdev, P. S. (2016). Clinical assessment of social cognitive function in neurological disorders. *Nature Reviews. Neurology*, *12*(1), 28–39. https://doi.org/10.1038/nrneurol.2015.229

Hyde, J. S. (2014). Gender similarities and differences. *Annual Review of Psychology*, *65*(1), 373–398. https://doi.org/10.1146/annurev-psych-010213-115057

Ickes, W., Gesn, P. R., & Graham, T. (2000). Gender differences in empathic accuracy: Differential ability or differential motivation? *Personal Relationships*, *7*(1), 95–109. https://doi.org/10.1111/j.1475-6811.2000.tb00006.x

Jack, R. E., Blais, C., Scheepers, C., Schyns, P. G., & Caldara, R. (2009). Cultural confusions show that facial expressions are not universal. *Current Biology*, *19*(18), 1543–1548. https://doi.org/10.1016/j.cub.2009.07.051

Jack, R. E., Garrod, O. G., Yu, H., Caldara, R., & Schyns, P. G. (2012). Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(19), 7241–7244. https://doi.org/10.1073/pnas.1200155109

Jackson, J. C., Watts, J., Henry, T. R., List, J. M., Forkel, R., Mucha, P. J., Greenhill, S. J., Gray, R. D., & Lindquist, K. A. (2019). Emotion semantics show both cultural variation and universal structure. *Science*, *366*(6472), 1517–1522. https://doi.org/10.1126/science.aaw8160

Johnson, T. (1998). Approaches to equivalence in cross-cultural and cross-national survey research. In J. A. Harkness (Ed.), *Zuma-Nachrichten Spezial Volume 3. Cross-cultural survey equivalence* (pp. 1–40). Zuma.

Kirkland, R. A., Peterson, E., Baker, C. A., Miller, S., & Pulos, S. (2013). Meta-analysis reveals adult female superiority in "Reading the Mind in the Eyes Test". *North American Journal of Psychology*, *15*(1), 121–146.

Klineberg, O. (1938). Emotional expression in Chinese literature. *Journal of Abnormal and Social Psychology*, *33*(4), 517–520. https://doi.org/10.1037/h0057105

Kobayashi, C., Glover, G. H., & Temple, E. (2006). Cultural and linguistic influence on neural bases of "Theory of Mind": An fMRI study with Japanese bilinguals. *Brain and Language*, *98*(2), 210–220. https://doi.org/10.1016/j.bandl.2006.04.013

Kret, M. E., & De Gelder, B. (2012). A review on sex differences in processing emotional signals. *Neuropsychologia*, *50*(7), 1211–1221. https://doi.org/10.1016/j.neuropsychologia.2011.12.022

Lausen, A., & Schacht, A. (2018). Gender differences in the recognition of vocal emotions. *Frontiers in Psychology*, *9*, Article 882. https://doi.org/10.3389/fpsyg.2018.00882

Lindquist, K. A., Satpute, A. B., & Gendron, M. (2015). Does language do more than communicate emotion? *Current Directions in Psychological Science*, *24*(2), 99–108. https://doi.org/10.1177/0963721414553440

Liu, D., Wellman, H. M., Tardif, T., & Sabbagh, M. A. (2008). Theory of mind development in Chinese children: A meta-analysis of false-belief understanding across cultures and languages. *Developmental Psychology*, *44*(2), 523–531. https://doi.org/10.1037/0012-1649.44.2.523

Mackin, R. S., & Areán, P. A. (2009). Incidence and documentation of cognitive impairment among older adults with severe mental illness in a community mental health setting. *The American Journal of Geriatric Psychiatry*, *17*(1), 75–82. https://doi.org/10.1097/JGP.0b013e31818cd3e5

Matías-Guiu, J. A., Pytel, V., Cortés-Martínez, A., Valles-Salgado, M., Rognoni, T., Moreno-Ramos, T., & Matías-Guiu, J. (2018). Conversion between Addenbrooke's Cognitive Examination III and Mini-Mental State Examination. *International Psychogeriatrics*, *30*(8), 1227–1233. https://doi.org/10.1017/S104161021700268X

Matsumoto, D. (1990). Cultural similarities and differences in display rules. *Motivation and Emotion*, *14*(3), 195–214. https://doi.org/10.1007/BF00995569

Matsumoto, D. (1992). American-Japanese cultural differences in the recognition of universal facial expressions. *Journal of Cross-Cultural Psychology*, *23*(1), 72–84. https://doi.org/10.1177/0022022192231005

Matsumoto, D., & Hwang, H. C. (2019). Culture and emotion: Integrating biological universality with cultural specificity. In D. Matsumoto & H. C. Hwang (Eds.), *The handbook of culture and psychology* (pp. 361–398). Oxford University Press.

Matsumoto, D., & Juang, L. (2016). *Culture and psychology*. Nelson Education.

McAndrew, F. T. (1986). A cross-cultural study of recognition thresholds for facial expressions of emotion. *Journal of Cross-Cultural Psychology*, *17*(2), 211–224. https://doi.org/10.1177/0022002186017002005

McDonald, S., Honan, C., Allen, S. K., El-Helou, R., Kelly, M., Kumfor, F., Piguet, O., Hazelton, J. L., Padgett, C., & Keage, H. A. D. (2018). Normal adult and adolescent performance on TASIT-S, a short version of The Assessment of Social Inference Test. *The Clinical Neuropsychologist*, *32*(4), 700–719. https://doi.org/10.1080/13854046.2017.1400106

McSweeney, B. (2002). Hofstede's model of national cultural differences and their consequences: A triumph of faith—A failure of analysis. *Human Relations*, *55*(1), 89–118. https://doi.org/10.1177/0018726702551004

McWhirter, L., Ritchie, C., Stone, J., & Carson, A. (2020). Functional cognitive disorders: A systematic review. *The Lancet Psychiatry*, *7*(2), 191–207. https://doi.org/10.1016/S2215-0366(19)30405-5

Mehta, U. M., Thirthalli, J., Naveen Kumar, C., Mahadevaiah, M., Rao, K., Subbakrishna, D. K., Gangadhar, B. N., & Keshavan, M. S. (2011). Validation of Social Cognition Rating Tools in Indian Setting (SOCRATIS): A new test-battery to assess social cognition. *Asian Journal of Psychiatry*, *4*(3), 203–209. https://doi.org/10.1016/j.ajp.2011.05.014

Merten, J. (2005). Culture, gender and the recognition of the basic emotions. *Psychologia*, *48*(4), 306–316. https://doi.org/10.2117/psysoc.2005.306

Muthukrishna, M., Bell, A. V., Henrich, J., Curtin, C. M., Gedranovich, A., McInerney, J., & Thue, B. (2020). Beyond Western, Educated, Industrial, Rich, and Democratic (WEIRD) psychology: Measuring and mapping scales of cultural and psychological distance. *Psychological Science*, *31*(6), 678–701. https://doi.org/10.1177/0956797620916782

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, *53*(4), 695–699. https://doi.org/10.1111/j.1532-5415.2005.53221.x

Nelson, N. L., & Russell, J. A. (2013). Universality revisited. *Emotion Review*, *5*(1), 8–15. https://doi.org/10.1177/1754073912457227

Norman, M. A., Evans, J. D., Miller, W. S., & Heaton, R. K. (2000). Demographically corrected norms for the California verbal learning test. *Journal of Clinical and Experimental Neuropsychology*, *22*(1), 80–94. https://doi.org/10.1076/1380-3395(200002)22:1;1-8;FT080

Oyserman, D., Coon, H. M., & Kemmelmeier, M. (2002). Rethinking individualism and collectivism: Evaluation of theoretical assumptions and meta-analyses. *Psychological Bulletin*, *128*(1), 3–72. https://doi.org/10.1037/0033-2909.128.1.3

Potts, C., Richardson, J., Bond, R. B., Price, R. K., Mulvenna, M. D., Zvolsky, P., Harvey, M., Hughes, C. F., & Duffy, F. (2021). Reliability of Addenbrooke's Cognitive Examination III in differentiating between dementia, mild cognitive impairment and older adults who have not reported cognitive problems. *European Journal of Ageing*, *22*, 1–13. https://doi.org/10.1007/s10433-021-00652-4

Puente, A. E., Pérez-García, M., Vilar-Lopez, R., Hidalgo-Ruzzante, N., & Fasfous, F. (2013). Neuropsychological assessment of culturally and educationally dissimilar individuals. In F. Paniagua & A. M. Yamada (Eds.), *Handbook of multicultural mental health: Assessment and treatment of diverse population* (2nd ed., pp. 225–242). Elsevier. https://doi.org/10.1016/B978-0-12-394420-7.00012-6

Quesque, F., Michalon, S., Blanc, F., & Bertoux, M. (2020). Validation de la mini-SEA dans une population française variée: Données de référence pour la pratique clinique. *Revue de Neuropsychologie*, *12*(4), 367–375. https://doi.org/10.1684/nrp.2020.0609

Quesque, F., & Rossetti, Y. (2020). What do theory-of-mind tasks actually measure? Theory and practice. *Perspectives on Psychological Science*, *15*(2), 384–396. https://doi.org/10.1177/1745691619896607

Ruffman, T., Halberstadt, J., & Murray, J. (2009). Recognition of facial, auditory, and bodily emotions in older adults. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *64*(6), 696–703. https://doi.org/10.1093/geronb/gbp072

Ruffman, T., Henry, J. D., Livingstone, V., & Phillips, L. H. (2008). A meta-analytic review of emotion recognition and aging: Implications for neuropsychological models of aging. *Neuroscience and Biobehavioral Reviews*, *32*(4), 863–881. https://doi.org/10.1016/j.neubiorev.2008.01.001

Russell, J. A., Suzuki, N., & Ishida, N. (1993). Canadian, Greek, and Japanese freely produced emotion labels for facial expressions. *Motivation and Emotion*, *17*(4), 337–351. https://doi.org/10.1007/BF00992324

Russell, L. L., Greaves, C. V., Bocchetta, M., Nicholas, J., Convery, R. S., Moore, K., Cash, D. M., van Swieten, J., Jiskoot, L., Moreno, F., Sanchez-Valle, R., Borroni, B., Laforce, R., Jr., Masellis, M., Tartaglia, M. C., Graff, C., Rotondo, E., Galimberti, D., Rowe, J. B., . . . the Genetic FTD Initiative, GENFI. (2020). Social cognition impairment in genetic frontotemporal dementia within the GENFI cohort. *Cortex*, *133*, 384–398. https://doi.org/10.1016/j.cortex.2020.08.023

Santamaría-García, H., Baez, S., Gómez, C., Rodriguez-Villagra, O., Portela, M., Reyes, P., Klahr, J., Matallana, D., & Ibanez, A. (2020). The role of social cognition skills and social determinants of health in predicting symptoms of mental illness. *Translational Psychiatry*, *10*, Article 165. https://doi.org/10.1038/s41398-020-0852-4

Sensenbrenner, B., Rouaud, O., Graule-Petot, A., Guillemin, S., Piver, A., Giroud, M., Béjot, Y., & Jacquin-Piques, A. (2020). High prevalence of social cognition disorders and mild cognitive impairment long term after stroke. *Alzheimer Disease and Associated Disorders*, *34*(1), 72–78. https://doi.org/10.1097/WAD.0000000000000355

Spitzer, N., Shafir, T., Lerman, Y., & Werner, P. (2019). The relationship between caregiver burden and emotion recognition deficits in persons with MCI and early AD: The mediating role of caregivers' subjective evaluations. *Alzheimer Disease and Associated Disorders*, *33*(3), 266–271. https://doi.org/10.1097/WAD.0000000000000323

Statucka, M., & Cohn, M. (2019). Origins matter: Culture impacts cognitive testing in Parkinson's disease. *Frontiers in Human Neuroscience*, *13*, Article 269. https://doi.org/10.3389/fnhum.2019.00269

Stone, V. E., Baron-Cohen, S., & Knight, R. T. (1998). Frontal lobe contributions to theory of mind. *Journal of Cognitive Neuroscience*, *10*(5), 640–656. https://doi.org/10.1162/089892998562942

Talhelm, T., Zhang, X., Oishi, S., Shimin, C., Duan, D., Lan, X., & Kitayama, S. (2014). Large-scale psychological differences within China explained by rice versus wheat agriculture. *Science*, *344*(6184), 603–608. https://doi.org/10.1126/science.1246850

Tang-Wai, D. F., Knopman, D. S., Geda, Y. E., Edland, S. D., Smith, G. E., Ivnik, R. J., Tangalos, E. G., Boeve, B. F., & Petersen, R. C. (2003). Comparison of the short test of mental status and the mini-mental state examination in mild cognitive impairment. *Archives of Neurology*, *60*(12), 1777–1781. https://doi.org/10.1001/archneur.60.12.1777

Van de Vijver, F. J. R. (2019). Cross-cultural research. In J. Edlund & A. Nichols (Eds.), *Advanced research methods and statistics for the behavioral and social sciences*. Cambridge University Press.

Van de Vijver, F. J. R., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Sage Publications.

Wang, Y., & Luo, Y. (2005). Standardization and assessment of college students' facial expression of emotion. *Chinese Journal of Clinical Psychology*, *13*(4), 396–398.

Wang, Y., & Su, Y. (2006). Theory of mind in old adults: The performance on Happé's stories and faux pas stories. *Psychologia*, *49*(4), 228–237. https://doi.org/10.2117/psysoc.2006.228

Wright, R. W. (1970). Trends in international business research. *Journal of International Business Studies*, *1*(1), 109–123. https://doi.org/10.1057/palgrave.jibs.8490724

Wu, S., & Keysar, B. (2007). The effect of culture on perspective taking. *Psychological Science*, *18*(7), 600–606. https://doi.org/10.1111/j.1467-9280.2007.01946.x

Yarkoni, T. (2020). The generalizability crisis. *Behavioral and Brain Sciences*, *45*, Article e1. https://doi.org/10.1017/S0140525X20001685

Yong, M. H., Lawrie, L., Schaefer, A., & Phillips, L. H. (2021). The effects of adult aging and culture on theory of mind. *The Journals of Gerontology: Series B*, *77*(2), 332–340. https://doi.org/10.1093/geronb/gbab093

Received October 11, 2021
Revision received April 14, 2022
Accepted May 16, 2022 ■