

Does culture shape our understanding of others' thoughts and emotions? An investigation across 12 countries

Short title: Geographical variations of social cognition

Article accepted in Neuropsychology (March 2022)

Quesque François¹, Coutrot Antoine², Cox Sharon³, de Souza Leonardo Cruz⁴, Baez Sandra⁵, Cardona Juan Felipe⁶, Mulet-Perreault Hannah⁷, Flanagan Emma^{8,9}, Neely-Prado Alejandra¹⁰, Clarens Maria Florencia¹¹, Cassimiro Luciana¹², Musa Gada¹³, Kemp Jennifer¹⁴, Botzung Anne¹⁴, Philippi Nathalie¹⁴, Cosseddu Maura¹⁵, Trujillo Catalina⁶, Grisales Johan Sebastián⁶, Fittipaldi Sol^{16,17}, Magrath Guimet Nahuel¹¹, Calandri Ismael Luis¹¹, Crivelli Lucia¹¹, Sedeno Lucas¹⁷, Garcia Adolfo M^{16,17,18,19}, Moreno Fermin²⁰, Indakoetxea Begoña²⁰, Benussi Alberto²¹, Brandão Moura Millena Vieira⁴, Santamaria-Garcia Hernando²², Matallana Diana²², Prianishnikova Galina²³, Morozova Anna²³, Iakovleva Olga²³, Veryugina Nadezda²³, Levin Oleg²³, Zhao Lina²⁴, Liang Junhua²⁴, Duning Thomas²⁵, Lebouvier Thibaud¹, Pasquier Florence¹, Huepe David¹⁰, Barandiaran Myriam²⁰, Johnen Andreas²⁵, Lyashenko Elena²³, Allegri Ricardo F¹¹, Borroni Barbara²¹, Blanc Frederic¹⁴, Wang Fen²⁴, Yassuda Monica Sanches¹², Lillo Patricia¹³, Teixeira Antônio Lúcio⁴, Caramelli Paulo⁴, Hudon Carol⁷, Slachevsky Andrea^{26,27,28,29}, Ibáñez Agustin^{10,16,17,19,30†}, Hornberger Michael^{8,9†}, Bertoux Maxime^{1,9*}.

¹ Univ. Lille, Inserm, CHU Lille, U1172 - LiNCog - Lille Neuroscience & Cognition, LiCEND, F-59000 Lille, France

² Université de Nantes, CNRS, LS2N, F-44000 Nantes, France

³ Department of Behavioural Science and Health, Institute of Epidemiology and Healthcare, University College London, London, UK

⁴ University of Belo Horizonte, Brazil.

⁵ Universidad de los Andes, Bogotá, Colombia

⁶ Instituto de Psicología, Universidad del Valle, Cali, Colombia

⁷ Université Laval and CERVO Brain Research Centre, Québec, Canada

⁸ Norwich Medical School, University of East Anglia, UK

⁹ Department of Clinical Neurosciences, University of Cambridge, UK

¹⁰ Center for Social and Cognitive Neuroscience, School of Psychology, Adolfo Ibáñez University, Santiago, Chile

¹¹ FLENI Fondation, Department of Neurology, Buenos Aires, Argentina

- ¹² School of Arts, Sciences and Humanities, University of São Paulo, Department of Neurology, São Paulo, Brazil
- ¹³ Universidad de Chile, Santiago, Chile
- ¹⁴ University of Strasbourg, Strasbourg, France
- ¹⁵ Neurology Unit, Spedali Civili Hospital, Brescia, Italy
- ¹⁶ Universidad de San Andrés, Buenos Aires, Argentina
- ¹⁷ National Scientific and Technical Research Council (CONICET), Argentina
- ¹⁸ Departamento de Lingüística y Literatura, Facultad de Humanidades, Universidad de Santiago de Chile, Santiago, Chile.
- ¹⁹ Global Brain Health Institute (GBHI), University of California-San Francisco (UCSF), San Francisco, California, United States.
- ²⁰ Department of Neurology, Unit of Cognitive Disorders, Hospital Universitario Donostia, San Sebastian, Spain
- ²¹ Centre for Neurodegenerative Disorders, Department of Clinical and Experimental Sciences, University of Brescia, Brescia, Italy
- ²² School of Medicine, Neuroscience Doctorate. Aging Institute, Physiology and Psychiatry Department. Pontificia Universidad Javeriana, Bogotá, Colombia
- ²³ Central Clinic No 1 of the Ministry of Internal Affairs, Moskva, Russia
- ²⁴ Innovation center for neurological disorders, Department of Neurology, Xuan Wu Hospital, Capital Medical University, 45 Changchun Street, Beijing
- ²⁵ Clinic of Neurology with Institute for Translational Neurology, University Hospital Münster, Münster, Germany
- ²⁶ Geroscience Center for Brain Health and Metabolism (GERO), Faculty of Medicine, University of Chile, Santiago, Chile.
- ²⁷ Neuropsychology and Clinical Neuroscience Laboratory (LANNEC), Physiopathology Department - ICBM, Neuroscience and East Neuroscience Departments, Faculty of Medicine, University of Chile, Santiago, Chile.
- ²⁸ Memory and Neuropsychiatric Clinic (CMYN) Neurology Department, Hospital del Salvador and Faculty of Medicine, University of Chile, Santiago, Chile.
- ²⁹ Servicio de Neurología, Departamento de Medicina, Clínica Alemana-Universidad del Desarrollo, Santiago, Chile.
- ³⁰ Universidad Autónoma del Caribe, Barranquilla, Colombia.

† These authors contributed equally

*** Corresponding author**

Maxime Bertoux: maxime.bertoux@inserm.fr

CMRR, Département de Neurologie, Hôpital Roger Salengro, rue Emile Laine, CHU de Lille, 59000

Lille, France.

Data and scripts used for the analyses of these data are freely available online at

https://osf.io/dksv7/?view_only=9386572b0206439ca98d1a03657685ab

Abstract

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.

Keywords: Social cognition, Emotion recognition, Mentalizing, Theory of mind, Cross-Cultural, Culture, bvFTD, autism, schizophrenia

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.

Introduction

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 “W.E.I.R.D.” people, Henrich, Heine & Norenzayan, 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 non-verbal (visuoperceptual tests) tests as compared to those who learned English as a second language (Boone, Victor, Wen, Razani, & Pontón, 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 non-native 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, Puente, Fasfous, Hidalgo-Ruzzante & Pérez-García, 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, 2008). 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 et al., 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 nationalities in verbal and non-verbal 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, Ewan, Miller & Heaton, 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, Lonner, & Thorndike's, 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. Cohn & Statucka, 2019; Daugherty et al., 2017). 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 vs 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 constitutes 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 anti-socialism 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 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 (Henry et al., 2016; Cotter et al., 2018) that have been widely explored by validated tests (Stone et al., 1998; Ekman & Friesen, 1976) 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, 2019). Anticipating these recommendations, the International Network on Social Cognition Disorders (INSCD) was developed at the University of Cambridge 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. 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/> , 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.

Methods

Participants

The recruitment and assessment of 587 healthy control participants (339 women, age range 18–89 years, mean=58.04 years \pm 16.06 s.d., mean education=10.5 years \pm 5.46 s.d., where s.d. 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: (1) no cognitive complaints; (2) no depressive complaints; (3) a normal cognitive screening test (i.e. either the Mini Mental State Examination in 10/18 centres, the Montreal Cognitive Assessment in 5/18 centres, or the Addenbrooke's Cognitive Examination III in 3/18 centres, using the local normative data, considering age, gender and education; see Supplementary Material 1); (4) no current psychiatric disorder; (5) no past or current neurological disease; (6) 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 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.

Please insert Table 1

Ethics approval was given for the centre for Addictive Behaviours 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, UK, the 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-SEA (mini Social cognition & Emotional Assessment, 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 facial emotion recognition (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. Quesque et al. 2020; Clarens et al., 2021) and clinical contexts (e.g. Bertoux et al, 2013; El Grabli et al., 2021; 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 Supplementary Table 1 for details.

Three drawings illustrate each story (see Supplementary 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 (one 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 (five points, one per question). For the other stories, the accurate detection of the absence of faux pas provides two points. Two control questions (providing one 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, Huang, Wang, et Luo, 2011 and Supplementary 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, Schmidt, Billiet & Meulemann, 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 minimise, 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, 1998's 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, i.e. 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 (UK) 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 & 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: $\text{score} \sim \text{age} + \text{gender} + \text{education} + (1 | \text{country})$. 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 were 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.

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

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 & 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 ($t(537)=-2.01$, $p=0.03$). We found a small effect of gender ($t(537)=-1.98$, $p=0.048$), but the Hedge's g effect size is almost null: $g=0.03$, $95\%IC=[-0.14\ 0.20]$, positive values corresponding to an advantage for women. We found a positive effect of education on the modified faux pas score ($t(537)=4.00$, $p<0.001$). 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 facial emotion recognition score ($t(566)=-6.08$, $p<0.001$), with a monotonic decline over the lifespan. Figure 1A shows the decline of facial emotion recognition scores with age, both for men and women. Gender also had an effect on the facial emotion recognition score, with women having higher scores than men ($t(566)=4.95$, $p<0.001$). 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 ($t(566)=0.91$; $p=0.36$). Effect of age on the mFP and education on the FER are illustrated on the Supplementary Material 6.

Please insert Figure 1

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 $LRStat(1) = 118.53$, $p<0.001$ for mFP, $LRStat(1) = 119.01$; $p<0.001$ 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 variance partition coefficient (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 Supplementary Material 7 for the significance of each pairwise country differences). Unlike conditional modes, Hedge's g are 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 (VPC_NFP = 11.6%), indicating that variations across countries are more important when there actually is a faux pas to identify.

Please insert Figure 2

Variation of facial emotion recognition scores across countries - The variance partition coefficient (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, i.e. 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 Supplementary 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 facial emotion recognition 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.

Please insert Figure 3

Variations across countries for each emotion are illustrated in Figure 4, which represents country-specific mean performances for each emotion in the facial emotion recognition test. We fitted a linear mixed model for each emotion with age, gender and education as fixed effects, and country as a random effect: $\text{score_emotion} \sim \text{age} + \text{gender} + \text{education} + (1 | \text{country})$. To compare the effect of countries across the different emotions, we computed a VPC for each emotion. The 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%, $95\%CI=[6.54\ 20.03]$).

Please insert Figure 4

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.

Please insert Figure 5

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: $\text{score} \sim \text{age} + \text{gender} + \text{education} + \text{language} + (1 | \text{country})$. 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 facial emotion recognition performances ($t(559) = 1.06, p = 0.35$) or on faux pas detection performances ($t(531) = 0.74, p = 0.77$).

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: $\text{score} \sim \text{age} + \text{gender} + \text{education} + (1 | \text{country}) + (1 | \text{country}:\text{center})$. For facial emotion recognition, $\text{VPC}_{\text{country}} = 11.70\%$ and $\text{VPC}_{\text{country}:\text{center}} = 10.33\%$. For faux pas detection, $\text{VPC}_{\text{country}} = 16.29\%$ and $\text{VPC}_{\text{country}:\text{center}} = 9.00\%$. 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 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, Ruffman, Murray, Taumoepeau, & Ryan, 2011; Wang & Su, 2006) but also when labeling facial expressions of emotion (Ruffman, Halberstadt, & Murray, 2009; Ruffman et al., 2008). In the present case, however, only 5 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, 2017; Ruffman et al, 2008). 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 co-exist (Lausen & Schacht, 2018). The “*gender difference hypothesis*” suggests that women have better socio-emotional decoding skills than men, even if this advantage is relatively small according to meta-analyses (Kirkland, Peterson, Baker, Miller & Pulos, 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, Cumberland & Spinrad, 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, Bassett & Wyatt, 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, Gesn & Graham, 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 ten 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 multi-site 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 2 or 3 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 & Keysar (2007) have shown that cultural differences induce different profiles of perspective-taking, either favoring other people's (Chinese participants) or self-related (US participants) perspectives. In another study using the faux pas test, Malaysian participants were found to have lower performance than participants from the UK (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 two 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 multi-centric 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 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 labelled by all participants, while for others, cross-country variations existed. Specifically, happy expressions were unambiguously categorized as such in every country (see Nelson & Russel, 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 than 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 current 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) vs “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” vs “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 intra-cultural 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 facial emotion recognition 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, 2019). We believe that the question of the generalizability of findings is a pressing and consequential problem in neurosciences (Yarkoni, 2019). 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, multi-cultural 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., 2021) 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 & Arean, 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 geriatrics 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, Roberson, van der Vyver & Barrett, 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 towards 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 socio-demographic 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 dataset. 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 Supplementary Material 1 & 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 variation 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, Glover & Temple, 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 dataset 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 multi-centric research programs should be generalized to all subfields of neuropsychology. We are aware that this would only constitute a first step towards 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.

Acknowledgement

We would like to thank Prof. Stephanie D. Preston and Prof. Skye McDonald as well as a third anonymous reviewer for their constructive comments during the revision of this paper. We also deeply appreciated the help of Prof. McDonald for English revision. For their administrative or clinical support related to the acquisition of data, we would like to thank María Goldfeder, Agostina Galiani (INECO Foundation, Argentina), Ma Eugenia Martin, Cecilia Cervino, Ma Eugenia Taberner, 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

(University of Cambridge, UK), Stephen Jeffs (University of East Anglia, UK), 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), CONICET, FONCyT-PICT 2017-1818, FONCyT-PICT 2017 1820, ANID/FONDAP/15150012, ANID /FONDECYT/1140423 &1100975, GBHI ALZ UK-20-639295, NIH NIA R01 AG057234, and the MULTI-PARTNER CONSORTIUM TO EXPAND DEMENTIA RESEARCH IN LATIN AMERICA (ReDLat; National Institutes of Aging of the National Institutes of Health under award number R01AG057234, Alzheimer's Association grant SG-20-725707-ReDLat, the Rainwater Foundation, and the Global Brain Health Institute), 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. 18-313-20006), CNPq, Brazil (Ref. 402853/2012-1), Projet Hospitalier de Recherche Clinique (PHRC) inter-régional (IDRCB 2012-A00992-41; HUS N°5330), Alzheimer's Research UK (ARUK_PPG2014B-3; ARUK-DT2016-1), Alzheimer's Society UK (AS-SF-14-003), Deutsche Telekom, São Paulo Research Foundation (FAPESP - 2016/07967-2), Laboratory of Excellence Development of Innovative Strategies for a Transdisciplinary Approach to Alzheimer's Disease (LabEx DISTALZ) as well as 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.

References

- Adams Jr, R. B., Rule, N. O., Franklin Jr, R. G., Wang, E., Stevenson, M. T., Yoshikawa, S., ... & Ambady, N. (2010). Cross-cultural reading the mind in the eyes: an fMRI investigation. *Journal of cognitive neuroscience*, 22(1), 97-108.
- Adolphs, R. (2009). The social brain: neural basis of social knowledge. *Annual review of psychology*, 60, 693-716.
- 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*, 71(1), 1-10.
- Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., ... & Rahwan, I. (2018). The moral machine experiment. *Nature*, 563(7729), 59-64.
- 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), e0179336.
- 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.
- Barrett, L. F., & Satpute, A. B. (2019). Historical pitfalls and new directions in the neuroscience of emotion. *Neuroscience letters*, 693, 9-18.
- Bendix, R. (1963). Concepts and generalizations in comparative sociological studies. *American Sociological Review*, 28, 532-539.
- Beaupré, M. G., & Hess, U. (2005). Cross-cultural emotion recognition among Canadian ethnic groups. *Journal of cross-cultural psychology*, 36(3), 355-370.
- Bertoux, M., Delavest, M., de Souza, L. C., Funkiewiez, A., Lépine, J. P., Fossati, P., ... & Sarazin, M. (2012). Social cognition and emotional assessment differentiates frontotemporal dementia from depression. *Journal of Neurology, Neurosurgery & Psychiatry*, 83(4), 411-416.

- 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), S84-S94.
- Bertoux, M., Duclos, H., Caillaud, M., Segobin, S., Merck, C., de La Sayette, V., ... & Laisney, M. (2020). When affect overlaps with concept: emotion recognition in semantic variant of primary progressive aphasia. *Brain*, 143(12), 3850-3864.
- Boccardi, M., Monsch, A. U., Ferrari, C., Altomare, D., Berres, M., Bos, I., ... Consortium for the Harmonization of Neuropsychological Assessment for Neurocognitive Disorders (2021). Harmonizing neuropsychological assessment for mild neurocognitive disorders in Europe. *Alzheimer's & dementia: the journal of the Alzheimer's Association*, 10.1002/alz.12365.
- 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.
- 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.
- Brislin, R. W., Lonner, W. J., & Thorndike, R. (1973). *Cross-cultural research methods*. New York: 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*, 116(32), 15861-15870.
- 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.
- Darwin, C. (1872). *The expression of the emotions in man and animals*. London: John Murray, Albermarle Street.

Collins, M. (1980). Interviewer variability: a review of the problem. *Journal of the Market Research Society*.

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 & Biobehavioral Reviews*, 84, 92-99.

Coutrot, A., Silva, R., Manley, E., de Cothi, W., Sami, S., Bohbot, V. D., ... & Spiers, H. J. (2018). Global determinants of navigation ability. *Current Biology*, 28(17), 2861-2866.

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

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

de Souza, L. C., Bertoux, M., de Faria, R. V., Corgosinho, L. T. S., Prado, A. C. d. 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.

Denham, S. A., Bassett, H. H., & Wyatt, T. (2007). The socialization of emotional competence. *Handbook of socialization: Theory and research*, 614-637.

Ducharme, S., Dols, A., Laforce, R., Devenney, E., Kumfor, F., van den Stock, J., ... & Pijnenburg, Y. (2020). Recommendations to distinguish behavioural variant frontotemporal dementia from psychiatric disorders. *Brain*, 143(6), 1632-1650.

Eddy, C. M. (2019). What do you have in mind? Measures to assess mental state reasoning in neuropsychiatric populations. *Frontiers in psychiatry*, 10, 425.

Eisenberg, N., Cumberland, A. & Spinrad, T.L. (1998). Parental socialization of emotion. *Psychological inquiry*, 9(4), 241-273.

Ekman P, & Friesen, W.V. (1976). Pictures of facial affect. Consulting psychologists. Press Palo Alto, California.

Ekman, P. & Friesen, W. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica*, 1(1), 49–98.

Ekman, P., & Friesen, W.V. (1971). Constants across cultures in the face and emotion. *Journal of personality and social psychology*, 17, 124.

El Grabli, F., Quesque, F., Borg, C., Witthöft, M., Michael, G. A., Lucas, C., ... & Bertoux, M. (2021). Interoception and social cognition in chronic low back pain: a common inference disturbance? An exploratory study. *Pain Management*, (0).

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: a meta-analysis. *Psychological bulletin*, 128(2), 203.

Fiedler, K. (2008). Language: A toolbox for sharing and influencing social reality. *Perspectives on psychological science*, 3(1), 38-47.

Filippova, E., & Astington, J. W. (2008). Further development in social reasoning revealed in discourse irony understanding. *Child development*, 79(1), 126-138.

Gendron, M., Lindquist, K. A., Barsalou, L. & Barrett, L.F. (2012). Emotion words shape emotion percepts. *Emotion*, 12(2), 314-325.

Gendron, M. (2017). Revisiting diversity: Cultural variation reveals the constructed nature of emotion perception. *Current opinion in psychology*, 17, 145-150.

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.

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

Goodman, Z. T., Llabre, M. M., González, H. M., Lamar, M., Gallo, L. C., Tarraf, W., ... & 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.

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

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), 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, 133–136.

Hayden, K.M., Makeeva, O.A., Newby, L.K., Plassman, B.L., Markova, V.V., Dunham, A. & Roses, A.D. (2014). A comparison of neuropsychological performance between US and Russia: Preparing for a global clinical trial. *Alzheimer's & Dementia*, 10(6), 760-768.

Henrich, J., Heine, S.J. & Norenzayan, A. (2010a). The weirdest people in the world? *Behavioral and brain sciences*, 33, 61-83.

Henrich, J., Heine, S.J. & Norenzayan, A. (2010b). Most people are not WEIRD. *Nature*, 466, 29-29.

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.

Hyde, J.S. (2014). Gender similarities and differences. *Annual review of psychology*, 65, 373-398.

Ickes, W., Gesn, P.R. & Graham, T. (2000). Gender differences in empathic accuracy: Differential ability or differential motivation? *Personal Relationships*, 7(1), 95-109.

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.

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*, 109(19), 7241-7244.

Jackson, J.C., Watts, J., Henry, T.R., List, J.M., Forkel, R., Mucha, P.J., ..., Lindquist, K.A. (2019). Emotion semantics show both cultural variation and universal structure. *Science*, 366(6472), 1517-1522.

Johnson, T. (1998). Approaches to equivalence in cross-cultural and cross-national survey research. In Zuma-Nachrichten Spezial Volume 3. *Cross-Cultural Survey Equivalence*, ed. JA Harkness, pp. 1-40. Mannheim: 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.

Klineberg, O. (1938). Emotional expression in Chinese literature. *The Journal of Abnormal and Social Psychology*, 33(4), 517-520.

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.

Kret, M.E. & de Gelder, B. (2012). A review on sex differences in processing emotional signals. *Neuropsychologia*, 50(7), 1211-1221.

Lausen, A. & Schacht, A. (2018). Gender differences in the recognition of vocal emotions. *Frontiers in psychology*, 9, 882.

- 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.
- Liu, D., Wellman, H.M., Tardiff, 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.
- 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.
- Matsumoto, D. & Juang, L. (2016). *Culture and psychology*. Nelson Education.
- Matsumoto, D. (1990). Cultural similarities and differences in display rules. *Motivation and emotion*, 14(3), 195-214.
- Matsumoto, D. (1992). American-Japanese cultural differences in the recognition of universal facial expressions. *Journal of Cross-Cultural Psychology*, 23, 72-84
- 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.
- 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.
- McDonald S, Honan C, Allen SK, El-Helou R, Kelly M, Kumfor F, et al. Normal adult and adolescent performance on TASIT-S, a short version of The Assessment of Social Inference Test. *Clin Neuropsychol*. 2018;32(4):700-19.
- McSweeney, B. (2002). Hofstede's model of national cultural differences and their consequences: A triumph of faith – A failure of analysis. *Human Relations*, 55, 89–118.
- McWhirter, L., Ritchie, C., Stone, J. & Carson, A. (2020). Functional cognitive disorders: a systematic review. *The Lancet Psychiatry*, 7(2), 191-207.

- 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, 203-209.
- Merten, J. (2005). Culture, gender and the recognition of the basic emotions. *Psychologia*, 48(4), 306-316.
- 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*, 0956797620916782.
- Nelson, N.L. & Russell, J.A. (2013). Universality revisited. *Emotion Review*, 5(1), 8-15.
- 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, 80-94.
- Oyserman, D., Coon, H.M. & Kemmelmeier, M. (2002). Rethinking individualism and collectivism: Evaluation of theoretical assumptions and meta-analyses. *Psychological Bulletin*, 128, 3–72.
- 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). New York, NY: Elsevier.
- 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).
- Quesque, F. & Rossetti, Y. (2020). What do theory-of-mind tasks actually measure? Theory and practice. *Perspectives on Psychological Science*, 15(2), 384-396.

- Ruffman, T., Halberstadt, J. & Murray, J. (2009). Recognition of facial, auditory, and bodily emotions in older adults. *Journal of Gerontology: Psychological Sciences*, 64B, 696-703.
- 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 & Biobehavioral Reviews*, 32(4), 863–881.
- Russell, J. A., Suzuki, N., & Ishida, N. (1993). Canadian, Greek, and Japanese freely produced emotion labels for facial expressions. *Motivation and Emotion*, 17, 337-351.
- Russell, L.L., Greaves, C.V., Bocchetta, M., Nicholas, J., Convery, R.S., Moore, K., ... & Rohrer, J.D.; Genetic FTD Initiative, GENFI (2020). Social cognition impairment in genetic frontotemporal dementia within the GENFI cohort. *Cortex Dec*; 133:384-398.
- Santamaría-García, H., Baez, S., Gómez, C., Rodríguez-Villagra, O., Portela, M., Reyes, P., ... & Ibanez, A. (2020). The role of social cognition skills (SCS) and social determinants of health (SDH) in predicting symptoms of mental illness. *Translational Psychiatry* <https://doi.org/10.1038/s41398-020-0852-4>
- Sensenbrenner, B., Rouaud, O., Graule-Petot, A., Guillemin, S., Piver, A., Giroud, M., ... & Jacquin-Piques, A. (2020). High prevalence of social cognition disorders and mild cognitive impairment long term after stroke. *Alzheimer Disease & Associated Disorders*, 34(1), 72-78.
- 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. *Alzheimer Disease & Associated Disorders*, 33(3), 266-271.
- Statucka, M., & Cohn, M. (2019). Origins matter: Culture impacts cognitive testing in Parkinson's disease. *Frontiers in human neuroscience*, 13, 269.
- Stone, V.E., Baron-Cohen, S. & Knight, R.T. (1998). Frontal lobe contributions to theory of mind. *Journal of cognitive neuroscience*, 10, 640-656.

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, 603-608.

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

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, United Kingdom: Cambridge University Press.

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.

Wright, R. W. (1970). Trends in international business research. *Journal of International Business Studies*, 1, 109-123.

Wu, S. & Keysar, B. (2007). The effect of culture on perspective taking. *Psychological science*, 18, 600-606

Yarkoni, T. (2020). The Generalizability Crisis. *Behavioural and Brain Science*, 21, 1-37.

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*.

Data and scripts used for the analyses of these data are freely available online at

https://osf.io/dksv7/?view_only=9386572b0206439ca98d1a03657685ab