Investigating Object Orientation Effects Across 18 Languages

Sau-Chin Chen[1], Erin Buchanan[2], Zoltan Kekecs[3,4], Jeremy K. Miller[5], Anna Szabelska[6], Balazs Aczel[3], Pablo Bernabeu[7], Patrick Forscher[8,9], Attila Szuts[3], Zahir Vally[10], Ali H. Al-Hoorie[11], Mai Helmy[12,13], Caio Santos Alves da Silva[14], Luana Oliveira da Silva[14], Yago Luksevicius de Moraes[14], Rafael Ming C. S. Hsu[14], Anthonieta Looman Mafra[14], Jaroslava V. Valentova[14], Marco Antonio Correa Varella[14], Barnaby Dixon[15], Kim Peters[15], Nik Steffens[15], Omid Ghaesmi[16], Andrew Roberts[16], Robert M. Ross[16], Ian D. Stephen[16,17], Marina Milyavskaya[18], Kelly Wang[18], Kaitlyn M. Werner[18], Dawn L. Holford[19], Miroslav Sirota[19], Thomas Rhys Evans[20], Dermot Lynott[7], Bethany M. Lane[21], Danny Riis[21], Glenn P. Williams[22], Chrystalle B. Y. Tan[23], Alicia Foo[24], Steve M. J. Janssen[24], Nwadiogo Chisom Arinze[25], Izuchukwu Lawrence Gabriel Ndukaihe[25], David Moreau[26], Brianna Jurosic[27], Brynna Leach[27], Savannah Lewis[27], Peter R. Mallik[27], Kathleen Schmidt[28], William J. Chopik[29], Leigh Ann Vaughn[30], Manyu Li[31], Carmel A. Levitan[32], Daniel Storage[33], Carlota Batres[34], Janina Enachescu[35], Jerome Olsen[35], Martin Voracek[35], Claus Lamm[36], Ekaterina Pronizius[36], Tilli Ripp[37], Jan Philipp Röer[37], Roxane Schnepper[37], Marietta Papadatou-Pastou[38], Aviv Mokady[39], Niv Reggev[39], Priyanka Chandel[40], Pratibha Kujur[40], Babita Pande[40], Arti Parganiha[40], Noorshama Parveen[40], Sraddha Pradhan[40], Margaret Messiah Singh[40], Max Korbmacher[41], Jonas R. Kunst[42], Christian K. Tamnes[42], Frederike S. Woelfert[42], Kristoffer Klevjer[43], Sarah E. Martiny[43], Gerit Pfuhl[43], Sylwia Adamus[44], Krystian Barzykowski[44], Katarzyna Filip[44], Patrícia Arriaga[45], Vasilije

Gvozdenović[46], Vanja Kovic[46], Tao-tao Gan[47], Chuan-Peng Hu[48], Qing-Lan Liu[47], Zhong Chen[49], Fei Gao[49], Lisa Li[49], Jozef Bavolár[50], Monika Hricová[50], Pavol Kacmár[50], Matúš Adamkovic[51,52], Peter Babincák[51], Gabriel Baník[51,52], Ivan Ropovik[52,53], Danilo Zambrano Ricaurte[54], Sara Álvarez Solas[55], Harry Manley[56], Panita Suavansri[56], Chun-Chia Kung[57], Belemir Çoktok[58], Asil Ali Özdogru[58], Çaglar Solak[59], Sinem Söylemez[59], Sami Çoksan[60], John Protzko[61], Ilker Dalgar[62], Vinka Mlakic[63], Elisabeth Oberzaucher[64], Stefan Stieger[63], Selina Volsa[63], Janis Zickfeld[65], & Christopher R. Chartier[27]

[1] Department of Human Development and Psychology, Tzu-Chi University, Hualien, Taiwan

[2] Harrisburg University of Science and Technology, Harrisburg, PA, USA

[3] Institute of Psychology, ELTE, Eotvos Lorand University, Budapest, Hungary

[4] Department of Psychology, Lund University, Lund, Sweden

[5] Department of Psychology, Willamette University,Salem OR, USA

[6] Institute of Cognition and Culture, Queen's University Belfast, UK

[7] Department of Psychology, Lancaster University, Lancaster, United Kingdom

[8] LIP/PC2S, Université Grenoble Alpes, Grenoble, France

[9] Busara Center for Behavioral Economics, Nairobi, Kenya

[10] Department of Clinical Psychology, United Arab Emirates University, Al Ain, UAE

[11] Royal Commission for Jubail and Yanbu, Jubail, Saudi Arabia

[12] Psychology Department, College of Education, Sultan Qaboos University, Muscat, Oman

[13] Psychology Department, Faculty of Arts, Menoufia University, Shebin El-Kom, Egypt

[14] Department of Experimental Psychology, Institute of Psychology, University of Sao Paulo, Sao Paulo, Brazil

[15] School of Psychology, University of Queensland, Brisbane, Australia

[16] Department of Psychology, Macquarie University, Sydney, Australia

[17] Department of Psychology, Nottingham Trent University, Nottingham, UK

[18] Department of Psychology, Carleton University, Ottawa, Canada

[19] Department of Psychology, University of Essex, Colchester, UK

[20] School of Social, Psychological and Behavioural Sciences, Coventry University, Coventry, UK

[21] Division of Psychology, School of Social and Health Sciences, Abertay University, Dundee, UK

[22] School of Psychology, Faculty of Health Sciences and Wellbeing, University of Sunderland, Sunderland, UK.

[23] Department of Psychiatry and Psychological Health, Universiti Malaysia Sabah, Sabah, Malaysia

[24] School of Psychology, University of Nottingham Malaysia, Selangor, Malaysia

[25] Department of Psychology, Alex Ekwueme Federal University, Ndufu-Alike, Nigeria

[26] School of Psychology, University of Auckland, Auckland, NZ

[27] Department of Psychology, Ashland University, Ashland, OH, USA

[28] School of Psychological and Behavioral Sciences, Southern Illinois University, Carbondale, IL, USA

[29] Department of Psychology, Michigan State University, East Lansing, MI, USA

[30] Department of Psychology, Ithaca College, Ithaca, NY, USA

[31] Department of Psychology, University of Louisiana at Lafayette, Lafayette, LA, USA

[32] Department of Cognitive Science, Occidental College, Los Angeles, USA

[33] Department of Psychology, University of Denver, Denver, CO, USA

[34] Department of Psychology, Franklin and Marshall College, Lancaster, PA, USA

[35] Faculty of Psychology, University of Vienna, Wien, Austria

[36] Department of Cognition, Emotion, and Methods in Psychology, Faculty of Psychology, University of Vienna, Wien, Austria

[37] Department of Psychology and Psychotherapy, Witten/Herdecke University, Germany

[38] School of Education, National and Kapodistrian University of Athens, Athens, Greece

[39] Department of Psychology, Ben Gurion University, Beersheba, Israel

[40] School of Studies in Life Science, Pt. Ravishankar Shukla University, Raipur, India

[41] Department of Biological and Medical Psychology, University of Bergen, Bergen, Norway
[42] Department of Psychology, University of Oslo, OSLO, Norway
[43] Department of Psychology, UiT - The Arctic University of Norway, Tromsø, Norway
[44] Institute of Psychology, Jagiellonian University, Krakow, Poland
[45] Iscte-University Institute of Lisbon, CIS-IUL, Lisbon, Portugal
[46] Laboratory for Neurocognition and Applied Cognition, Faculty of Philosophy, University of Belgrade, Belgrade, Serbia
[47] Department of Psychology, Hubei University, Wuhan, China
[48] School of Psychology, Nanjing Normal University, Nanjing, China
[49] Faculty of Arts and Humanities, University of Macau, Macau, China
[50] Department of Psychology, Faculty of Arts, Pavol Jozef Šafarik University in Košice, Košice, Slovakia
[51] Institute of Psychology, University of Presov, Prešov, Slovakia
[52] Institute for Research and Development of Education, Faculty of Education, Charles university, Prague, Czechia
[53] Faculty of Education, University of Presov, Prešov, Slovakia
[54] Faculty of Psychology, Fundación Universitaria Konrad Lorenz, Bogotá, Colombia
[55] Ecosystem Engineer, Universidad Regional Amazónica Ikiam, Tena, Ecuador
[56] Faculty of Psychology, Chulalongkorn University, Bangkok, Thailand
[57] Department of Psychology, National Cheng Kung University, Tainan, Taiwan
[58] Department of Psychology, Üsküdar University, Istanbul, Turkey
[59] Department of Psychology, Manisa Celal Bayar University, Manisa,Turkey
[60] Department of Psychology, Middle East Technical University, Ankara, Turkey
[61] Department of Psychological Science, Central Connecticut State University, New Britain, CT, USA
[62] Department of Psychology, Ankara Medipol University, Ankara, Turkey.
[63] Department of Psychology and Psychodynamics, Karl Landsteiner University of Health

Sciences, Krems an der Donau, Austria

[64] Department of Evolutionary Anthropology, University of Vienna, Wien, Austria

[65] Department of Management, Aarhus University, Aarhus, Denmark

Author Note

Author contributions: Sau-Chin Chen contributed to the study concept, the design analysis protocol and wrote the initial report draft. Patrick Forscher, Pablo Bernabeu, Balazs Aczel and Attila Szuts improved the analysis protocol. Zoltan Kekecs, Jeremy K. Miller and Anna Szabelska managed the project administration which was established by Christopher R. Chartier. All the rest of authors contributed to the material prepation and data collection. All authors commented on previous versions of the manuscript, read and approved the final manuscript.

Funding statement. Below authors had the individual funds supporiting their participations. Glenn P. Williams was supported by the Leverhulme Trust Research Project Grant (RPG-2016-093). Krystian Barzykowski was supported by the National Science Centre, Poland (2019/35/B/HS6/00528). Zoltan Kekecs was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Science. Erin Buchanan was supported by the National Institute on Mental Health (1R03MH110812-01). Patrícia Arriaga was supported by the Portuguese National Foundation for Science and Technology (UID/PSI/03125/2019). Gabriel Baník was supported by Charles University Grant Agency (PRIMUS/20/HUM/009).

Ethical approval statement. Authors who collected data on site and online had the ethical approval/agreement from the local institute. The latest status of ethical approval for all the participating authors is available at the public OSF folder (https://osf.io/e428p/ “IRB approvals” in Files).

Acknowledgement. We appreciated the major contributions from the contributors as below. Chris Chartier and Jeremy Miller managed and monitored progress. Erin Buchanan provided guidelines to improve the inter-lab progress website management and managed the JATOS server for online data collection. Arti Parganiha, Asil Özdoğru, Attila Szuts, Babita Pande, Danilo Zambrano Ricaurte, Gabriel Baník, Harry Manley, Jonas Kunst, Krystian Barzykowski, Marco Antonio Correa Varella, Marietta Papadatou Pastou, Niv Reggev, Patrícia Arriaga, Stefan Stieger, Vanja Ković and Zahir Vally managed the material translation from English to the other languages. Roles of each collaborator are available in the public table (https://osf.io/mz97h/). We thank the suggestions from the editor and two reviewers on our first and second proposals.

Correspondence concerning this article should be addressed to Sau-Chin Chen, No. 67, Jei-Ren St., Hualien City, Taiwan. E-mail: csc2009@mail.tcu.edu.tw

Abstract

Mental simulation theories of language comprehension propose that people automatically create mental representations of objects mentioned in sentences. Representation is often measured with the sentence-picture verification task, in which participants first read a sentence and, on a following screen, see a picture of an object. Participants then verify whether the latter object had been mentioned in the sentence. Crucially, two covert conditions exist: the sentence and the picture can either match or mismatch in terms of a perceptual property, including object orientation, shape, color and size. The key finding obtained in some studies is the match advantage, whereby responses were faster in the match condition; however, object orientation results are often inconsistent inconsistent findings across languages. This registered report describes our investigation of the match advantage of object orientation across 18 languages, which was undertaken by 33 laboratories and organized by the Psychological Science Accelerator. The preregistered analysis revealed that the match advantage was supported either overall or in any specific language.

Keywords: mental simulation, object orientation, mental rotation, language comprehension

Word count: 5,138 words in total; Introduction: 1,242 words

Investigating Object Orientation Effects Across 18 Languages

Method

Hypotheses and Design

The study design for the sentence-picture and picture-picture verification task was mixed using between-participant (language) and within-participant (match versus mismatch object orientation) independent variables. In the sentence-picture verification task, the match condition reflects a matching between the sentence and the picture, whereas in the picture-picture verification, it reflects a match in orientation between two pictures. The only dependent variable for both tasks was the response time. In the sentence-picture verification task, we expected response time to be shorter for matching compared to mismatching orientations within each language. We did not select languages systematically, but instead based on our collaboration recruitment with the Psychological Science Accelerator (PSA). We did not have any specific hypotheses about the relative size of the object orientation match advantage in different languages. In the picture-picture verification task, we expected shorter response time for identical orientation compared to different orientations. We computed an imagery score by subtracting the verification time for identical orientation from the verification time for different orientations. Based on the assumption that the mental rotation is a general cognitive aspect, we expect imagery scores to be the same on average across languages and can be used to predict a possible match advantage (see Chen et al., 2020).

Participants

In collaboration with the PSA (Moshontz et al., 2018), we collected data in 18 languages from 33 laboratories. Our a priori power analysis based on past English studies recommended one language would have at least one thousand participants based on the current design[1]. Among the 18 languages, only English data approached this number because 17 laboratories were in the countries English is the primary language. For the rest of 17 languages, the laboratories followed the secondary plan: they collected the latest number, 50 participants, then inspected the reached effect by the Bayesian sequential analysis.

Based on the preregistered plan, the available participants' accuracy had to reach 70%. Before the pandemic outbreak, 2,340 participants (1,104 women; M = 21.46 years old) from 33 laboratories joined and finished the study. After the study migrated online, there were 1403 participants (926 women; M = 23.75 years old) from 20 laboratories completed the study. Web-based participants at the beginning heard the auditory instruction and had to correctly answer at least 2 of 3 comprehension check questions about the instructions. All participating laboratories had ethical approval before data collection. All data and analyssis scripts are available on the source files (https://osf.io/p7avr/). Appendix 1 summarizes the average characteristics by language and laboratory.

(Erin suggested a section for power here)

General Procedure and Materials

In the beginning of the sentence-picture verification task, participants had to correctly answer all the practice trials. Each trial started with a left-justified and horizontally centered fixation point displayed for 1000 ms, immediately followed by the probe sentence. The sentence was presented until the participant pressed the space key, acknowledging that they understood the sentence. Then, the object picture was presented in the center of the screen until the participant responded otherwise it disappeared after 2 seconds. Participants were instructed to verify the object picture mentioned in the probe sentence as quickly and accurately as they could. Following the original study (Stanfield &

[1] See details of power analysis in the preregistered plan, p. 13 ~ 15. https://psyarxiv.com/t2pjv/

Zwaan, 2001), a memory check test was carried out after every three to eight trials to ensure that the participants had read each sentence carefully.

The picture-picture verification task used the same object pictures. In each trial, two objects appeared on either side of the central fixation point until either the participant indicated that the pictures displayed the same object or two different objects or until 2 seconds elapsed. In the trials displayed the same object, the pictures on each side presented the same orientation (both were horizontal/vertical) or different orientations (one was horizontal; one was vertical).

The study was executed using OpenSesame software for millisecond timing (Mathôt et al., 2012). Before the Covid-19 pandemic broke out, 29 participating laboratories had completed data collection. The remaining laboratories had to stop in person data collection because of local lockdowns. The project team decided to move data collection online. To minimize the differences between on-site and web-based studies, we converted the original Python code to Javascript and collected the data using OpenSesame through a JATOS server (Lange et al., 2015). After the changes in the procedure were approved by the journal editor and reviewers, we proceeded with the online study from February to June 2021. For the remote version, a recorded set of verbal instructions was played at the beginning of the study. Participants had to confirm they were native speakers of the targeted language. All verbal briefings were packaged in the language-specific scripts. Appendix 2 describes the deployment of the scripts and the results of participants' fluency tests. Following the literature, we did not anticipate any theoretically important differences between the two data sources (see Anwyl-Irvine et al., 2020; Bridges et al., 2020; de Leeuw & Motz, 2016). The instructions and experimental scripts are available at the public OSF folder (https://osf.io/e428p/ "Materials" in Files).

Analysis plan

Confirmatory Analysis According to our preregistered analysis plan[2], this study used meta-analysis and mixed-effect models to estimate the match advantage across languages. The meta-analysis summarized the median reaction times by match condition to determine the global effect size. This approach was compatible with ANOVA used by the original study (Stanfield & Zwaan, 2001). The mixed-effect models used each individual response time as the dependent variable and analyzed the fixed effects of matching condition using participant, target item, and lab id as random intercepts (Baayen et al., 2008). This approach was used by recent studies (Chen et al., 2020; Koster et al., 2018). The statistical analyses were conducted by R packages including *metafor* for meta analysis (Viechtbauer, 2010), *lme4* (Bates et al., 2015) and *lmerTest* (Kuznetsova et al., 2017) for mixed-effects models, as well as multiple regression through *R* base package (Version 4.1.1; R Core Team, 2021).

Imagery scores were the dependent measure of the picture-picture verification responses. Response times were summarized by the difference between the identical and different orientation. According to our preregistered analysis plan [3], we first evaluated the equality of imagery scores across languages in use of the mixed-effects models. Our other linear regression analysis evaluated the imagery scores as the predictor of match advantage. In a best fit model having the imagery score as the predictor, the slope would represent the association of image score and match advantage..

Exploratory Analysis. We conducted mixed-effect models for languages that reached the recommended sample size from our power analyses and for languages that showed a significant match advantage.

[2] See the analysis plan in the preregistered plan, p. 19 ~ 20. https://psyarxiv.com/t2pjv/

[3] See the analysis plan in the preregistered plan, p. 21. https://psyarxiv.com/t2pjv/

Decision criterion p-values were interpreted using the preregistered alpha level of .05. Because in our preregistered plan each language was assumed a standalone group, p-values of the analysis by each language were not corrected (Armstrong, 2014). All the final mixed-effects models were selected by pursuing a maximal random-effects structure whilst allowing the model to converge (Bates et al., 2015). p-values for each effect were calculated using the Satterthwaite approximation for degrees of freedom (Luke, 2017).

Results

Within the data collected on-site, 1,979 participants finished the sentence-picture verification task and met the preregistered inclusion criterion (accuracy percentile > 70%); 2,007 participants finished the picture-picture verification task. Raw data files containing data for twenty-eight participants were lost due to human error. Within the data sets collected online, 1,337 participants finished the sentence-picture verification task and met the preregistered inclusion criterion; 1,402 participants finished the picture-picture verification task.

Confirmatory analysis: Intra-lab analysis during data collection

Before data collection, each lab decided whether they wanted to apply a sequential analysis (Schönbrodt et al., 2017) or whether they wanted to settle for a fixed sample size. The preregistered protocol for labs applying sequential analysis established that they could stop data collection upon reaching the preregistered criterion ($BF_{10} = 10 \; or \; -10$), or the maximal sample size. Each laboratory chose a fixed sample size and an incremental n for sequential analysis before their data collection.

Two laboratories (HUN 001, TWN 001) stopped data collection at the preregistered criterion. Some laboratories did not conduct the sequential analysis on all their data because of one of the following reasons: (1) their data collection was interrupted by the pandemic outbreak; (2) participants performed worse in the online study; (3) too many of their participants were non-native speakers. Lab-specific results were reported on a public website as each laboratory completed data collection (details available in Appendix 2).

Confirmatory analysis: Inter-lab analysis of final data

Table 1

Median reaction times and accuracy percentages (in parentheses) per match condition (Mismatching, Matching); Match advantage (difference in response times) by language in the on-site data.

Language	N	Mismatching	Matching	Match Advantage
English	613	585(83.31)	582(84.60)	2.50
German	99	584(84.68)	570(87.54)	14.50
Greek	98	754(80.36)	728(83.33)	25.00
Hebrew	146	570(86.53)	574(86.82)	-4.25
Hindi	79	630(78.80)	666(79.54)	-36.00
Hungarian	129	623(87.60)	646(87.02)	-22.50
Norwegian	123	585(86.38)	608(87.53)	-23.00
Polish	50	595(86.00)	585(86.83)	10.00
Portuguese	5	579(86.67)	576(90.00)	3.00
Simplified Chinese	81	658(82.30)	644(81.17)	14.00
Slovak	138	622(86.23)	608(84.84)	13.25
Spanish	127	663(81.76)	683(81.76)	-20.00
Thai	50	652(83.17)	650(80.17)	2.50
Traditional Chinese	93	640(84.23)	630(85.22)	10.00
Turkish	183	638(85.56)	622(84.79)	16.00

Identification of outliers. Our preregistered plan included excluding outliers based on a linear mixed-model analysis for participants in the third quantile of the grand

Table 2

Median reaction times and accuracy percentages (in parentheses) per match condition (Mismatching, Matching); Match advantage (difference in response times) by language in the web-based data.

Language	N	Mismatching	Matching	Match Advantage
Arabic	106	539(66.59)	515(66.98)	24.50
Brazilian Portuguese	50	634(86.33)	622(84.83)	11.00
English	770	555(83.01)	550(84.20)	5.00
German	134	582(85.95)	558(86.38)	23.75
Norwegian	21	622(85.32)	589(86.90)	33.00
Portuguese	55	643(86.52)	596(84.70)	47.00
Serbian	128	604(84.96)	606(86.20)	-2.75
Traditional Chinese	57	600(84.65)	583(85.96)	17.00
Turkish	79	698(85.97)	657(83.86)	41.00

intercept (i.e., participants with the longest average response times). After examining the data from both online and in-person data collection, it became clear that both a minimum response latency and maximum response latency should be employed, as improbable times existed at both ends of the distribution (**kvalsethHickLawEquivalent2021?**; **proctorHickLawChoice2018?**). The maximum response latency was calculated as two times the mean absolute deviation plus the median calculated separately for each participant. Individual participants were removed if they did not reach our accuracy criterion, and individual data points were excluded if they did not fall within the acceptable response time range. All the below data analysis depended on the datasets excluding the outliers.

(Insert Table 1 about here)

(Insert Table 2 about here)

Meta-analysis of match advantages across laboratories. Because the preregistered analysis plan did not consider the data collected online, we conducted the overall meta-analyses for the complete dataset and separately by data collection source. Since data from small samples may contribute to a biased estimate, nine datasets with sample sizes smaller than 25 were excluded from the analyses. The overall meta-analysis found insignificant match advantage (Figure 1).Among the languages that had at least two datasets, we conducted the meta-analysis for English, German, Norway, Traditional Chinese, Slovak, and Turkey. Only German showed a significant meta-analytic effect across laboratories(see Figure 2). Results of the other languages are available in Appendix 3.

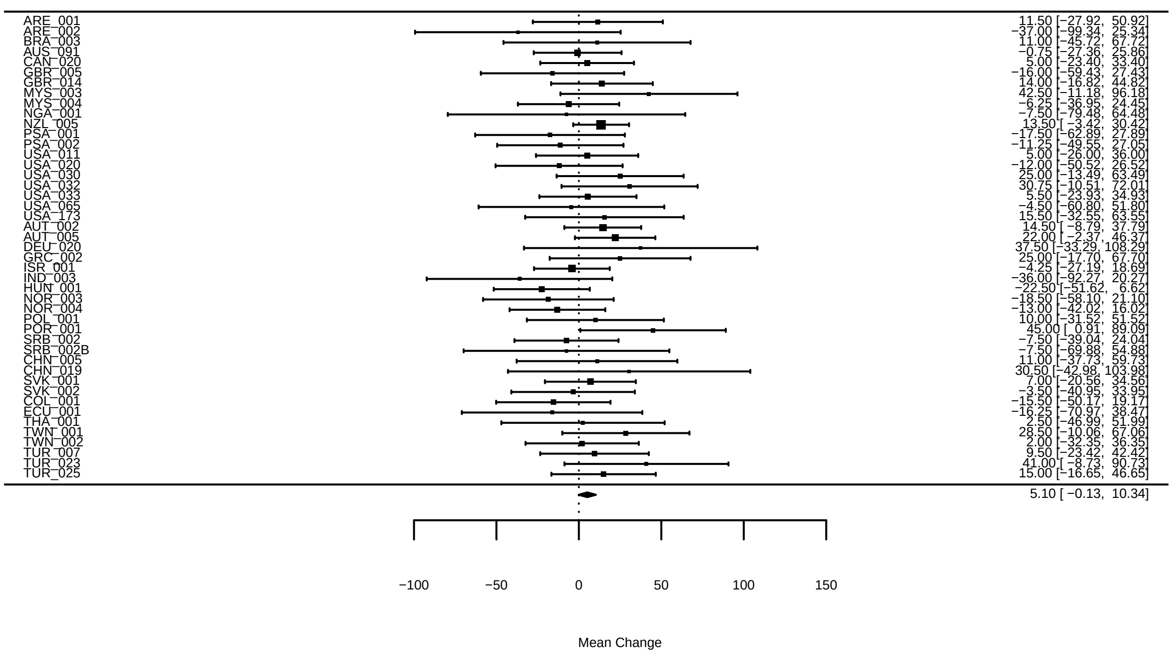

Figure 1. Meta-analysis on match advantage of object orientation for all datasets

(Insert Figure 1 about here)

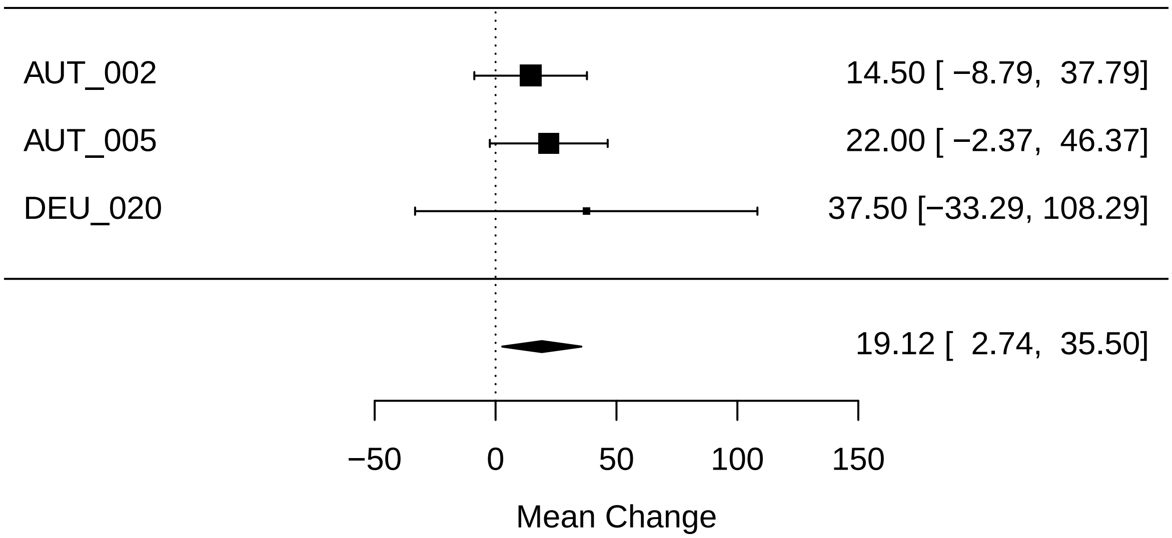

Figure 2. Meta-analysis on match advantage of object orientation for German datasets.

(Insert Figure 2 about here)

Evaluating match advantages using linear mixed-effects models. We at first evaluated whether one mixed-effects model sufficiently fitted all the data from on-site and web-based data. This analysis showed no difference between data sources: b = 16.761, *SE* = 15.232, t(27.679) = 1.1, p = 0.281. Therefore, the following analysis did not separate on-site and the web-based data. All other models are reported in Appendix 4.

We compared the fitness of the models with and without the random intercept of items. The result indicated that the models with the random intercept had the best fitness. The model revealed no significant effect of match advantage: b = -4.025, *SE* = 7.855, t(64265.353) = -0.512, p = 0.608 and no interaction of match advantage and any language, all *ps* > .05. (see “Models including languages” section in Appendix 4).

(Insert Figure 3 about here)

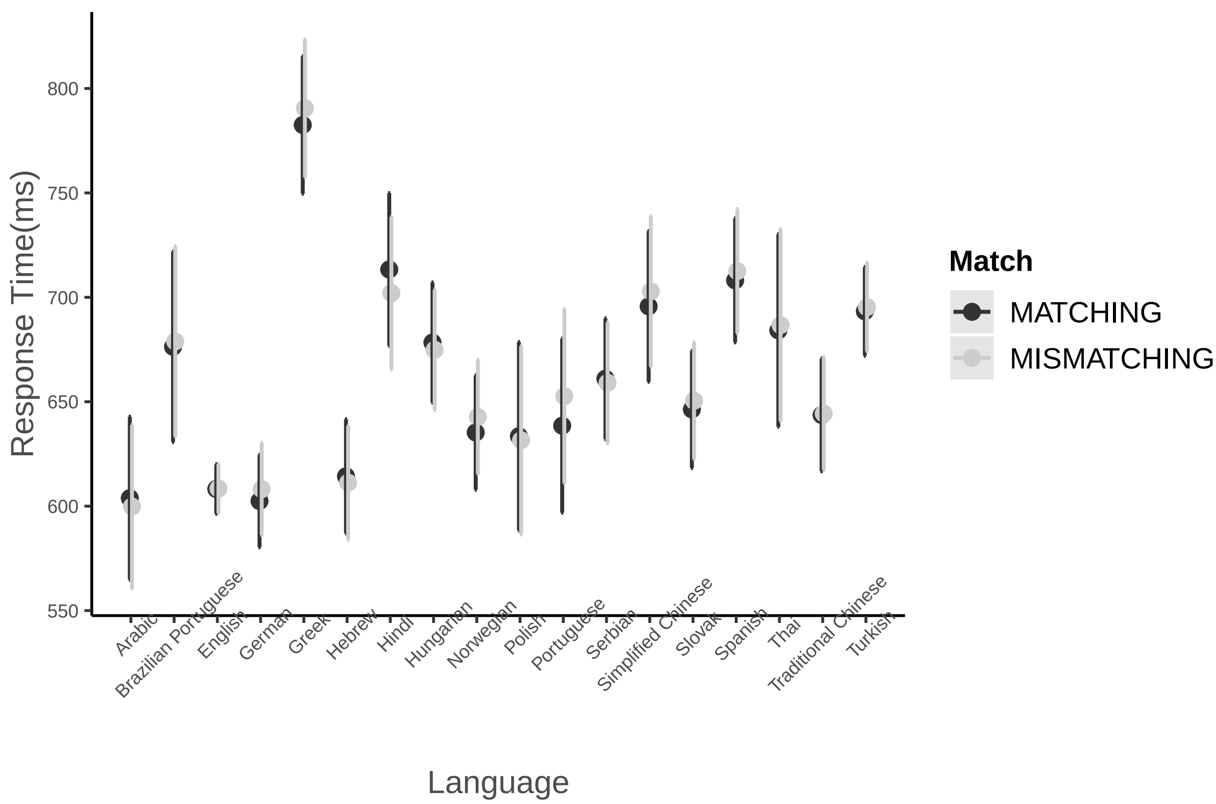

Figure 3. Response times and standard error in the sentence-picture verification task by match condition in each language.

Analysis of imagery scores. Prior to data collection, we assumed the imagery scores of every language group would be nearly equal. The best-fitting model included random intercepts for participants, targets and laboratories but no slopes for orientation. The fixed effect of orientation match was significant, $b = 27.366$, $SE = 2.275$, t(138198.589) = 12.032, $p < .001$. The response times illustrated in Figure 4 indicated that the imagery scores measured for each language were consistently positive supporting our hypothesis. The coefficients of all evaluated mixed-effects models are reported in Appendix 5.

(Insert Figure 4 about here)

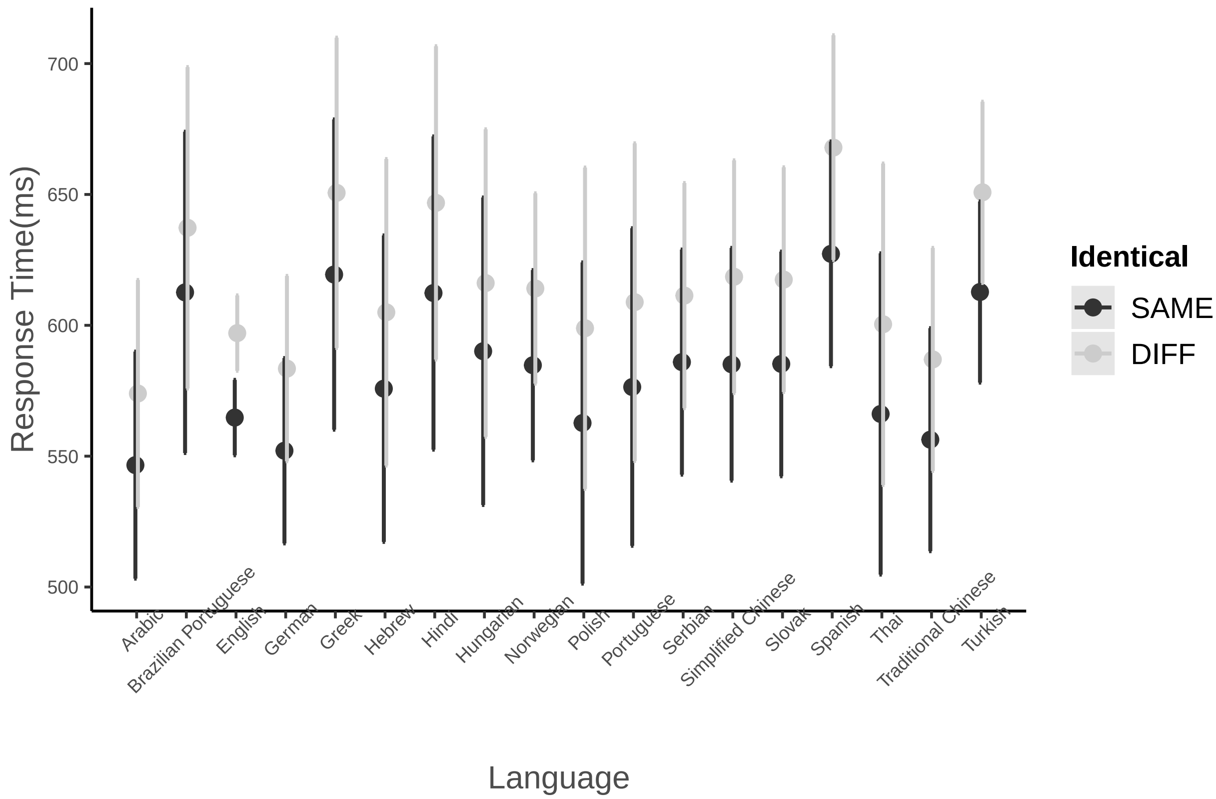

Figure 4. Response times and standard error in the picture-picture verification task by match condition in each language (both on-site and web-based data).

The above analyses suggested that data sources did not influence the imagery scores but did influence the match advantage. Therefore, we compared the model of languages and imagery scores and the model with languages only. Both models included match advantage as the dependent variable. If imagery scores predicted match advantage, the model with languages and imagery scores should fit the data better than the model with languages only. Because the random slopes for items were zero (see Appendix 5), the data for building the regression models were the aggregated data by participants. (**Require advice for precise writing**)

In the linear regression analysis, we selected the best fit model from the model for only one predictor, language, and the model with two predictors, language and imagery scores. Because the analysis of match advantage revealed a difference between on-site versus web-based data, we conducted separate regression analyses for the two data sources. In the analysis of the on-site data, the model with language with language only fit the data as good as the model with language and imagery scores, F (18,3272) = 1.047, p =0.402.

Exploratory analysis: Language-specific match advantages

Based on our exploratory plan described earlier, we selected the English datasets (N = 1,345) and the German datasets (N = 233). For both languages, we are interested in whether the data sources would show differences in thematch advantage. Another topic of interest is if the match advantage changed with English dialects, namely American English and British English.

Using the data from 1,345 English speaking participants, we ran a mixed-effects model usingmatch condition and English dialects (American vs. British) as fixed effects. The fittest model indicated the difference between the dialects, b = -50.191, SE = 23.247, t(18.428) = -2.159, p = 0.044, but failed to reveal the match advantage, p = 0.821. Also, this exploratory analysis indicated null interaction of orientation match condition and English dialects: p = 0.465 (see the detailed report in Appendix 4).

We conducted another exploratory mixed-effect model on German data because his was the only language to show a significant result in the preregistered meta-analysis. The best fit model had orientation match condition and data sources as the fixed effects. The match advantage of orientation was far from the significant level: b = 5.503, SE = 3.23, t(4544.905) = 1.704, p = 0.0893 (see the detailed report in Appendix 4). This result suggested that German study could have a robust estimation in the circumstance multiple teams conducted the study in terms of one the same protocol. Combined with the previous results of German(Koster et al., 2018), future research on this language could explore any potential linguistic aspects that might result in the match advantage of object orientation. Although this study is unable to which aspects contributed to the results, the advantage for the future language-specific studies would have a precise sample size justification on the participants and stimulus items.

References

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Armstrong, R. A. (2014). When to use the Bonferroni correction. *Ophthalmic and Physiological Optics*, *34*(5), 502–508. https://doi.org/10.1111/opo.12131

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using Lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing mega-study: Comparing a range of experiment generators, both lab-based and online. *PeerJ*, *8*, e9414. https://doi.org/10.7717/peerj.9414

Chen, S.-C., de Koning, B. B., & Zwaan, R. A. (2020). Does object size matter with regard to the mental simulation of object orientation? *Experimental Psychology*, *67*(1), 56–72. https://doi.org/10.1027/1618-3169/a000468

de Leeuw, J. R., & Motz, B. A. (2016). Psychophysics in a Web browser? Comparing response times collected with JavaScript and Psychophysics Toolbox in a visual search task. *Behavior Research Methods*, *48*(1), 1–12. https://doi.org/10.3758/s13428-015-0567-2

Koster, D., Cadierno, T., & Chiarandini, M. (2018). Mental simulation of object orientation and size: A conceptual replication with second language learners. *Journal of the European Second Language Association*, *2*(1). https://doi.org/10.22599/jesla.39

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). **lmerTest**

Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13

Lange, K., Kühn, S., & Filevich, E. (2015). "Just Another Tool for Online Studies" (JATOS): An easy solution for setup and management of web servers supporting online studies. *PLOS ONE*, *10*(6), e0130834. https://doi.org/10.1371/journal.pone.0130834

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, *49*(4), 1494–1502. https://doi.org/10.3758/s13428-016-0809-y

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

Moshontz, H., Campbell, L., Ebersole, C. R., IJzerman, H., Urry, H. L., Forscher, P. S., Grahe, J. E., McCarthy, R. J., Musser, E. D., Antfolk, J., Castille, C. M., Evans, T. R., Fiedler, S., Flake, J. K., Forero, D. A., Janssen, S. M. J., Keene, J. R., Protzko, J., Aczel, B., … Chartier, C. R. (2018). The Psychological Science Accelerator: Advancing psychology through a distributed collaborative network. *Advances in Methods and Practices in Psychological Science*, *1*(4), 501–515. https://doi.org/10.1177/2515245918797607

Schönbrodt, F. D., Wagenmakers, E.-J., Zehetleitner, M., & Perugini, M. (2017). Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences. *Psychological Methods*, *22*(2), 322–339. https://doi.org/10.1037/met0000061

Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science*, *12*(2), 153–156. https://doi.org/10.1111/1467-9280.00326

Viechtbauer, W. (2010). Conducting Meta-Analyses in R with the metafor Package.

Journal of Statistical Software, *36*(1), 1–48.
https://doi.org/10.18637/jss.v036.i03