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 their local institutions. 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 response time. The time difference between conditions in each task are the measurement of orientation effects and mental rotation scores. We did not select languages systematically, but instead based on our collaboration recruitment with the Psychological Science Accelerator (PSA, Moshontz et al., 2018).

(1) In the sentence-picture verification task, we expected response time to be shorter for matching compared to mismatching orientations within each language. In the picture-picture verification task, we expected shorter response time for identical orientation compared to different orientations. We did not have any specific hypotheses about the relative size of the object orientation match advantage in different languages.

(2) 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 null imagery score across languages and no association with mental simulation effects (see Chen et al., 2020).

Participants

The preregistered power analysis indicated n = 156 to 620 participants for 80% power for a directional one-sample t-test for a d = 0.20 and 0.10, respectively. A separate mixed-model simulation suggested that n = 400 participants with 100 items (i.e., 24 planned items nested within at least five languages) would produce 90% power to detect the same effect as Zwaan and Preacher (2012). The laboratories were allowed to follow a secondary plan: a team collected at least their preregistered minimum sample size (suggested 100 to 160 participants, most implemented 50), and then determine whether or not to continue data collection via Bayesian sequential analysis (stopping data collection if BF10 = 10 or - 10)[1].

In collaboration with the PSA, we collected data in 18 languages from 47 laboratories. Each laboratory chose a maximal sample size and an incremental n for sequential analysis before their data collection. Because the preregistered power analysis did not match the final analysis plan, we additionally completed a sensitivity analysis to ensure sample size was adequate to detect small effects, and the results indicated that each effect could be detected at a 2.23 millisecond range for the object orientation effect.

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 additional 4209 participants (2778 women; M = 23.75 years old) from 20 laboratories who completed the study. Web-based participants heard auditory instructions at the beginning of the study and had to correctly answer at least 2 of 3 comprehension check questions about the instructions. All participating laboratories had either ethical approval or institutional evaluation before data collection. All data and analysis scripts are available on the source files (https://osf.io/p7avr/). Appendix 1 summarizes the average

[1] Some laboratories requested withdrawal before they collected the requested minimum 50 participants for the unexpected affairs. The first team (GBR_006) stopped the data collection at 25th participant.

characteristics by language and laboratory.

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 & 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 where the same object was displayed, the pictures on each side were 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 other 4 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 Our preregistered analysis plan[2] employed the fixed-effects meta-analysis model that estimated the match advantage across laboratories and languages. The meta-analysis summarized the median reaction times by match condition then determine the effect size by laboratory. For the languages for which at least two teams collected data, we computed the meta-analytical effect size for these language data. 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, laboratory, and language as random intercepts (Baayen et al., 2008). 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). Because of the data from the Internet after COVID outbreaked, we at first evaluated the mixed-effects model with the fixed effects of match condition and data source and the four random intercepts. This analysis showed no difference between data sources: $b = 18.10$, $SE = 18.25$, $t(20.08) = 0.99$, $p = 0.33$. Therefore, the following mixed-effects models did not separate on-site and the web-based data. Language-specific mixed-effect models were conducted if the meta-analysis showed

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

the positive result.

Mental rotation 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 ANOVA. Because the later data collection was on the Internet, we used mixed models instead of ANOVA to evaluate the difference of data sources. The other planned analysis was the linear regression analysis in use of imagery scores as the predictor of match advantage. We evaluated the necessity of this analysis according to the outcomes of mixed-effect models.

Decision criterion. *p*-values were interpreted using the preregistered alpha level of .05. *p*-values for each effect were calculated using the Satterthwaite approximation for degrees of freedom (Luke, 2017).

Results

Within the data collected on-site, 2,006 participants finished the sentence-picture verification task and met the preregistered inclusion criterion ; 1,999 participants finished the picture-picture verification task. Within the data sets collected online, 1,390 participants finished both the tasks and met the preregistered inclusion criterion. One participant was removed because they did not reach our accuracy criterion.

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

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

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, $BF_{10} = -10$. Fourteen 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) two non-English laboratories (TUR_007, TWN_002) recruited English-speaking participants for the institutional policies. Lab-specific results were reported on a public website as each laboratory completed data collection (details available in Appendix 2).

Inter-lab analysis of final data

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 intercept (i.e., participants with the longest average response times). Only 49.62 % of participants' data could pass this criterion. 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. Two participants' data were excluded becauseif they did not fall between the acceptable minimum (160 ms) and maximum response time range (participant's median response time plus 2 median absolute deviation).

(Insert Table 1 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 merged data collection source. Ten teams were excluded because the data collection was incomplete[4]. The overall meta-analysis did not find a significant match advantage. Among the languages that had at least two laboratories, we conducted the meta-analysis for the languages with more than two teams (English, German, Norway, Simplified Chinese, Traditional Chinese, Slovak, and Turkey). In addition to the significant overall match advantage, German and Portuguese showed significant meta-analytic effects across laboratories (see Figure 1).

(Insert Figure **??** about here)

Evaluating match advantages using linear mixed-effects models. All models presented in this section are reported in Appendix 3. At first, we confirmed the null-effect model with all random intercept factors, including participants, items, teams, and languages, had the best model fit, AIC = 882504.1, BIC = 882558.9.[5] After adding the fixed effect predictor of matching orientation, this model did not reveal a significant effect of match advantage: $b = 0.889$, $SE = 1.09$, t(64957.195) = 0.815, $p = 0.415$, although this model had a better fitness than the null-effect model, χ^2 (6,7) = 0.665, $p = 0.415$. Among the other models considered, the model with highest theoretical interest had a random slope of matching condition on language. This model also showed no significant effect of match advantage: $b = 1.338$, $SE = 1.143$, t(43.194) = 1.171, $p = 0.248$, which was as equal fitness as the null effect model, χ^2 (6,9) = 1.343, $p = 0.719$. The illustration of match advantages by language were summarized in Figure 1).

We conducted mixed-effect models on German data because this was the only language indicated a significant match advantage in the meta-analysis. The best fitted

[4] Some laboratories requested withdrawal before they collected the requested minimum 50 participants for the unexpected affairs. The first team (GBR_006) stopped the data collection at 25th participant.

[5] All the estimated parameters of random intercepts are summarized in Appendix 3 and 4.

null-effect model had the random intercept factors of participants and items but teams[6], AIC = 59949.09, BIC = 59975.01. In comparison with the null-effect model, we indicated the significant difference with the model with orientation match condition as the fixed effects and the random effects of participants and items, $\chi^2 = (4,5) = 2.902$, $p = 0.088$. This model revealed the significant match advantage across teams: $b = 5.503$, $SE = 3.23$, $t(4544.905) = 1.704$, $p = 0.089$(see the detailed report in Appendix 3).

(Insert Figure 1 about here)

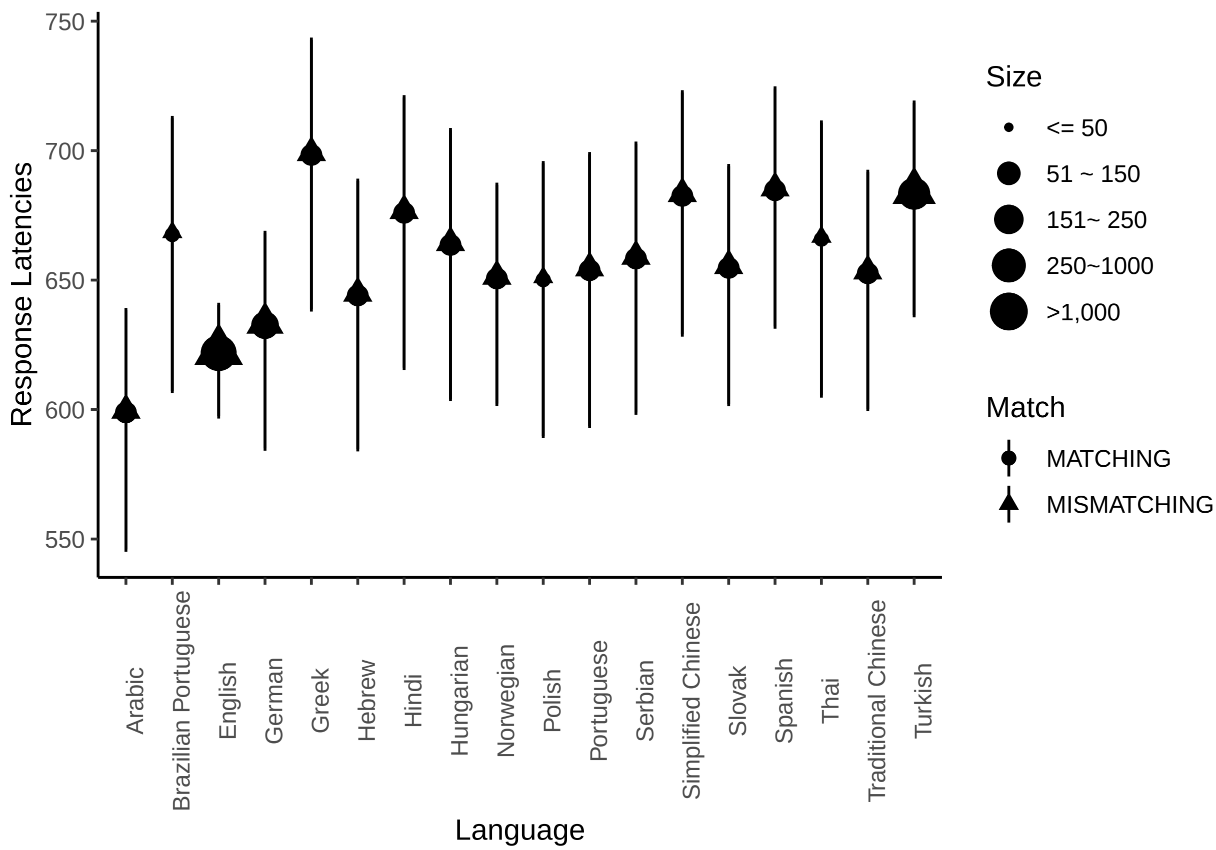

Figure 1. Average Response times and 95% CI in the sentence-picture verification task by match condition in each language

Analysis of mental rotation scores. This analysis treated the object orientation settings (index of mental rotation scores) and languages as the fixed effects and

[6] All the estimated parameters of random intercepts are summarized in Appendix 3 and 4.

participants, items, and teams as random effects.The null effect model with all the random intercept factors had the best fitness[7],AIC = 839191.3, BIC = 839237.2. When language and object orientation settings entered, the comparisons indicated the differences to the null effect model either for the model with the addition of the two fixed effects, χ^2 (5,23) = 3115.137, $p < .01$, or for the the interaction of the two fixed effects, χ^2 (5,40) = 3150.159, $p < .01$. Further comparison indicated the best fitness for the interaction model, χ^2 (23,40) = 35.022, $p < .01$. The interaction model indicated he significant mental rotation scores, b = 27.573, *SE* = 3.211, t(68313.425) = 8.588, $p < .01$. The response times illustrated in Figure 2 indicated that mental rotation scores varied among languages. The coefficients of all considered mixed-effects models are reported in Appendix 4.

was significant, —>

(Insert Figure 2 about here)

The last preregistered plan was to build a regression model to predict the match advantage in the sentence-picture task by the mental rotation score calculated from the picture-picture task. If mental rotation scores predicted match advantage, the regression model with languages and mental rotation scores should fit the data better than the regression model with languages only. However, the model comparison indicated the better fitted regression model had languages only, $F < 1$. As Table 2 illustrated, none of the language set of mental rotation scores sufficiently predict the match advantages.

```
## A data.frame with 6 labelled columns:
##
##                                term estimate        conf.int statistic   df p.value
## 1                         Intercept   -11.03   [-28.07, 6.01]     -1.27 3321    .204
## 2  LanguageBrazilian Portuguese      0.26 [-29.83, 30.36]      0.02 3321    .986
```

[7] All the estimated parameters of random intercepts are summarized in Appendix 3 and 4.

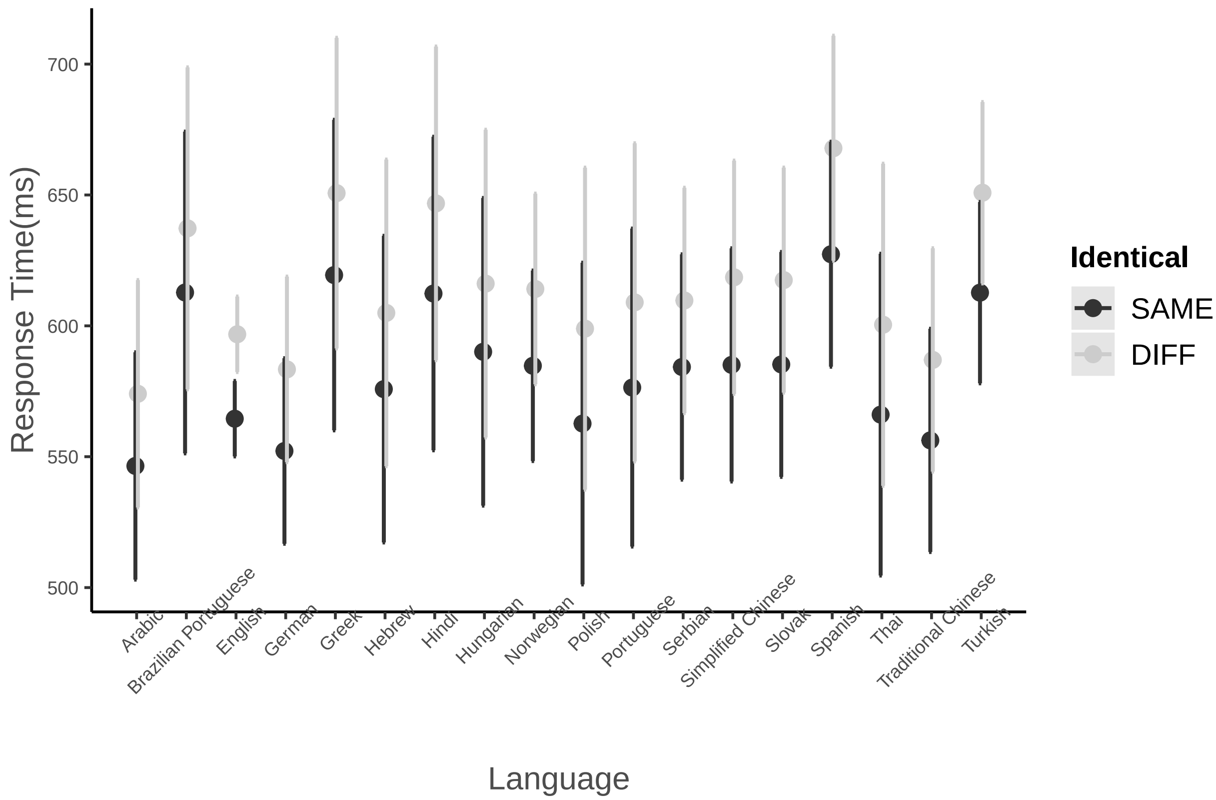

Figure 2. Response times and standard error in the picture-picture verification task by match condition in each language (both on-site and web-based data).

```
## 3            LanguageEnglish    7.39 [-10.30, 25.08]     0.82 3321    .413
## 4             LanguageGerman   17.15  [-3.40, 37.70]     1.64 3321    .102
## 5              LanguageGreek   15.27  [-9.31, 39.86]     1.22 3321    .223
## 6             LanguageHebrew    6.94 [-15.48, 29.36]     0.61 3321    .544
## 7              LanguageHindi   -2.35 [-28.52, 23.82]    -0.18 3321    .861
## 8          LanguageHungarian    1.09 [-21.91, 24.09]     0.09 3321    .926
## 9          LanguageNorwegian   20.69  [-1.82, 43.21]     1.80 3321    .072
## 10            LanguagePolish   10.33 [-19.76, 40.43]     0.67 3321    .501
## 11        LanguagePortuguese   25.77  [-2.57, 54.12]     1.78 3321    .075
## 12           LanguageSerbian    4.63 [-21.26, 30.52]     0.35 3321    .726
```

```
## 13   LanguageSimplified Chinese    17.92  [-8.06, 43.90]      1.35 3321    .176
## 14               LanguageSlovak    16.46  [-6.20, 39.11]      1.42 3321    .154
## 15              LanguageSpanish    12.81 [-10.27, 35.89]      1.09 3321    .276
## 16                 LanguageThai    14.37 [-15.72, 44.47]      0.94 3321    .349
## 17  LanguageTraditional Chinese    12.30  [-9.96, 34.56]      1.08 3321    .279
## 18              LanguageTurkish    16.63  [-3.58, 36.83]      1.61 3321    .107
##
## term     : Predictor
## estimate : $b$
## conf.int : 95\\% CI
## statistic: $t$
## df       : $\\mathit{df}$
## p.value  : $p$
```

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

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

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

Table 1

Descriptive statistics by language: Total sample size, Average accuracy percentage, Median response times and median absolute deviations (in parentheses) per match condition (Mismatching, Matching); Match advantage (difference in response times).

Language	N	Accuracy Percentages	Mismatching	Matching	Match Advantage
Arabic	106	77	539(219.05)	515(240.92)	24.50
Brazilian Portuguese	50	94	634(166.79)	622(126.76)	11.00
English	1363	93	567(127.50)	566(128.99)	1.00
German	233	96	582(106.01)	568(103.78)	14.50
Greek	98	89	754(232.77)	728(230.91)	25.00
Hebrew	146	96	570(103.41)	574(111.57)	-4.25
Hindi	79	88	630(200.15)	666(255.01)	-36.00
Hungarian	129	95	623(112.68)	646(129.73)	-22.50
Norwegian	144	96	590(131.58)	607(135.29)	-17.25
Polish	50	95	595(140.85)	585(117.87)	10.00
Portuguese	60	95	628(138.62)	583(137.14)	45.00
Serbian	129	94	604(150.48)	606(157.53)	-2.75
Simplified Chinese	81	90	658(177.17)	644(160.12)	14.00
Slovak	138	94	622(120.83)	608(114.90)	13.25
Spanish	127	91	663(154.19)	683(163.83)	-20.00
Thai	50	90	652(177.91)	650(129.36)	2.50
Traditional Chinese	150	93	617(139.74)	604(117.87)	13.00
Turkish	262	93	648(152.34)	632(131.58)	15.50