Investigating Object Orientation Effects Across 18 Languages

Sau-Chin Chen1, Erin Buchanan2, Zoltan Kekecs3,4, Jeremy K. Miller5, Anna Szabelska6, Balazs Aczel3, Pablo Bernabeu7, Patrick Forscher8,9, Attila Szuts3, Zahir Vally10, Ali H. Al-Hoorie11, Mai Helmy12,13, Caio Santos Alves da Silva14, Luana Oliveira da Silva14, Yago Luksevicius de Moraes14, Rafael Ming C. S. Hsu14, Anthonieta Looman Mafra14, Jaroslava V. Valentova14, Marco Antonio Correa Varella14, Barnaby Dixon15, Kim Peters15, Nik Steffens15, Omid Ghaesmi16, Andrew Roberts16, Robert M. Ross16, Ian D. Stephen16,17, Marina Milyavskaya18, Kelly Wang18, Kaitlyn M. Werner18, Dawn L. Holford19, Miroslav Sirota19, Thomas Rhys Evans20, Dermot Lynott7, Bethany M. Lane21, Danny Riis21, Glenn P. Williams22, Chrystalle B. Y. Tan23, Alicia Foo24, Steve M. J. Janssen24, Nwadiogo Chisom Arinze25, Izuchukwu Lawrence Gabriel Ndukaihe25, David Moreau26, Brianna Jurosic27, Brynna Leach27, Savannah Lewis27, Peter R. Mallik27, Kathleen Schmidt28, William J. Chopik29, Leigh Ann Vaughn30, Manyu Li31, Carmel A. Levitan32, Daniel Storage33, Carlota Batres34, Janina Enachescu35, Jerome Olsen35, Martin Voracek35, Claus Lamm36, Ekaterina Pronizius36, Tilli Ripp37, Jan Philipp Röer37, Roxane Schnepper37, Marietta Papadatou-Pastou38, Aviv Mokady39, Niv Reggev39, Priyanka Chandel40, Pratibha Kujur40, Babita Pande40, Arti Parganiha40, Noorshama Parveen40, Sraddha Pradhan40, Margaret Messiah Singh40, Max Korbmacher41, Jonas R. Kunst42, Christian K. Tamnes42, Frederike S. Woelfert42, Kristoffer Klevjer43, Sarah E. Martiny43, Gerit Pfuhl43, Sylwia Adamus44, Krystian Barzykowski44, Katarzyna Filip44, Patrícia Arriaga45, Vasilije Gvozdenović46, Vanja Kovic46, Tao-tao Gan47, Chuan-Peng Hu48, Qing-Lan Liu47, Zhong Chen49, Fei Gao49, Lisa Li49, Jozef Bavolár50, Monika Hricová50, Pavol Kacmár50, Matúš Adamkovic51,52, Peter Babincák51, Gabriel Baník51,52, Ivan Ropovik52,53, Danilo Zambrano Ricaurte54, Sara Álvarez Solas55, Harry Manley56, Panita Suavansri56, Chun-Chia Kung57, Belemir Çoktok58, Asil Ali Özdogru58, Çaglar Solak59, Sinem Söylemez59, Sami Çoksan60, John Protzko61, Ilker Dalgar62, Vinka Mlakic63, Elisabeth Oberzaucher64, Stefan Stieger63, Selina Volsa63, Janis Zickfeld65, & Christopher R. Chartier27

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](mailto: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 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

# 1 Introduction

The simulation of object properties is a major topic in conceptual processing research (Ostarek & Huettig, 2019; Scorolli, 2014). One well-known method for assessing mental simulation during sentence reading is the sentence-picture verification task. The task requires participants to read a probe sentence displayed on the screen. On the following screen, the participants see a picture of an object and must verify whether the object was mentioned in the probe sentence. For example, the probe sentence, ’Tom hammered the nail into the wall’ might be followed by a picture of a nail. Of central interest in the sentence-picture verification task is the match advantage, which occurs when people are faster to verify pictured objects whose properties match those of objects mentioned in the probe sentences. For example, in the sentence involving nails hammered into walls, an object orientation match advantage occurs when people are faster to verify horizontal nails than vertical ones. First discovered by Stanfield and Zwaan (2001), object orientation match advantages suggest that people mentally simulate objects during semantic processing (e.g., Barsalou, 1999, 2009). Researchers have found match advantages for shape (Zwaan et al., 2002) and color (Zwaan & Pecher, 2012; but see Connell, 2007). Consistent effects have appeared in English (Zwaan & Madden, 2005; Zwaan & Pecher, 2012), Chinese (Li & Shang, 2017), Dutch (De Koning et al., 2017; Engelen et al., 2011; Pecher et al., 2009; Rommers et al., 2013), German (Koster et al., 2018), Croatian (Šetić & Domijan, 2017), and Japanese (Sato et al., 2013). Object orientation, on the other hand, has produced mixed results across languages (Chen et al., 2020; De Koning et al., 2017; Koster et al., 2018; Zwaan & Madden, 2005; Zwaan & Pecher, 2012). To scrutinize the discrepancies across languages, we tested the match advantage of object orientation across 18 languages by means of a multi-lab collaboration. Among object properties, shape and color are described as intrinsic, which mean that these properties are relatively independent from the observer’s state. That is, regardless of the observers’ location or state, the object keeps its intrinsic properties constant. In contrast, an extrinsic property such as object orientation is dependent on the observer’s state. Studies on visual simulation have suggested that match advantages for intrinsic properties are more consistent than those for extrinsic properties (De Koning et al., 2017; Koster et al., 2018). On the other hand, studies on motor simulation (Beilock et al., 2008; Glenberg & Kaschak, 2002) showed the sentences that imply extrinsic properties (e.g., Joe sends the card to you) facilitated the responses to a congruent target (e.g., a “card” presented in a larger size). A later study suggests that orientation match advantages are stronger for large objects (Chen et al., 2020). Taken together, these results suggest that both intrinsic and extrinsic properties produce match advantages, though the advantage may be smaller for extrinsic properties.

## 1.1 Cross-linguistic and Experimental Factors

Several factors might contribute to cross-linguistic differences in the match advantage of orientation, including cultural, lexical and experimental ones. First, Ghandhari et al. (2020) found that cultural differences between Persian and Italian participants influenced the action sentence compatibility effect (for further differences studied in psycholinguistics, see Norcliffe et al., 2015). Second, languages differ in how they encode motion and placement events in sentences. Last, the potential role of mental rotation as a confound has been considered. We expand on the lexical and experimental factors below.

**Lexical Factors.** The probe sentences used in object orientation studies usually contain several motion events (e.g., “The ant walked towards the pot of honey and tried to climb in.”). The languages we probed in this study encode motion events in different ways, and grammatical differences between them could explain the different match advantage results. According to Verkerk (2014), Germanic languages (e.g., Dutch, English, German) generally encode the manner of motion in the verb (e.g., ‘The ant dashed’), while conveying the path information through satellite adjuncts (e.g., ‘towards the pot of honey’). In contrast, other languages, such as the Romance family (e.g., Portuguese, Spanish) more often encode path in the verb (e.g., ‘crossing,’ ‘exiting’). Crucially, the past research on the match advantage of object orientation is exclusively based on Germanic languages, and yet, there were differences across those languages, with English being the only one that consistently yielded the match advantage. As a minor difference across Germanic languages in this regard, Verkerk notes that path-only constructions (e.g., ‘The ant went to the feast’) are more common in English than in other Germanic languages.

Another topic to be considered is the lexical encoding of placement in each language, as the stimuli contains several placement events (e.g., ‘Sara situated the expensive plate on its holder on the shelf.’). Chen et al. (2020) and Koster et al. (2018) noted that some Germanic languages, such as German and Dutch, often make the orientation of objects more explicit than English. Whereas in English one could use the verb “put” in both “She put the book on the table” and “She put the bottle on the table,” in both Dutch and German, one could instead say “She laid the book on the table,” and “She stood the bottle on the table.” In these literal translations from German and Dutch, the verb “lay” encodes a horizontal orientation, whereas the verb “stand” encodes a vertical orientation. This distinction extends to verbs indicating existence. As Newman (2002) exemplified, an English speaker would be likely to say “There’s a lamp in the corner,” whereas a Dutch speaker would be more likely to say “There ‘stands’ a lamp in the corner.” Nonetheless, we cannot conclude that these cross-linguistic differences are affecting the match advantage across languages because there is no previous theory allowing predictions, and because placement events are not sufficiently prevalent in the stimuli.

**Experimental factors.** Stanfield and Zwaan (2001) found that participants who sufficiently understood the probe sentences showed a match advantage of object orientation. Later studies on this topic have examined the association between the match advantage and cognitive abilities. Spatial cognition is one of the relevant areas, which may be measured with mental rotation tasks. Studies have suggested that mental rotation tasks offer valid reflections of previous spatial experience (Frick & Möhring, 2013) and of current spatial cognition (Chu & Kita, 2008; Pouw et al., 2014). De Koning et al. (2017) suggested that mental rotation, as an alternative process to mental simulation, could quickly erase the mismatched orientation, replacing it with the orientation that matches the one described in the sentence (Cohen & Kubovy, 1993; Yaxley & Zwaan, 2007). Chen et al. (2020) investigated the relationship between the match advantage and mental rotation across three languages: English, Dutch and Chinese. They introduced the picture-picture verification task to examine how individuals process the target pictures regardless of their native language. This picture-picture verification task was designed using the mental rotation paradigm (Cohen & Kubovy, 1993). In each trial of this task, two pictures appear on opposite sides of the screen. Participants have to verify whether the pictures represent identical or different objects. The verification times for pictures of identical objects presented in the same orientation (i.e., two identical pictures presented in horizontal or vertical orientation) were shorter than those presented in different orientations (one horizontal; one vertical). Chen et al.’s findings suggested that crosslinguistic differences in the match advantage of object orientation are not confounded by mental rotation strategies.

## 1.2 Purposes of this study

Several explanations have been proposed for the inconsistent findings on the match advantage of object orientation, including the procedural details of some studies, such as participants not being required to verify the probe sentences they had read (see Zwaan, 2014). Without such a verification, participants might have paid less attention to the meaning of the probe sentences, in which case they would have been less likely to form a mental representation of the objects (e.g., Zwaan & van Oostendorp, 1993). In this regard, it is relevant to acknowledge that variability originating from individual differences and other characteristics of experiments can substantially influence the results (Barsalou, 2019).Thus, this study followed the original methods from Stanfield and Zwaan (2001) and addressed two primary questions: (1) How much of the match advantage of object orientation can be obtained within different languages and (2) How do differences in mental rotation affect the match advantage across languages?

# 2 Method

## 2.1 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 match 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 times 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 expected 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).

## 2.2 Participants

In collaboration with the PSA (Moshontz et al., 2018), we collected data in 18 languages. Our *a priori* power analysis recommended a language would have at least one thousand participants based on the current design[[1]](#footnote-1). The English data approached this number because 17 laboratories recruited native English speakers. 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 1,403 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. Appendix 1 summarizes the average characteristics by language and laboratory.

## 2.3 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. Two pictures showing the same critical object appeared in each “yes” trial; two pictures showing two different objects from the filler items appeared in each “no” trial.

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).

## 2.4 Analysis plan

**Confirmatory Analysis** According to our preregistered analysis plan[[2]](#footnote-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]](#footnote-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 indicate its accountability.

**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.

**Decision criterion** *p-*values were interpreted using the preregistered alpha level of .05. Because in our preregistered plan each language was assumed a standalone grtoup*,* *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).

# 3 Results

Within the data collected on-site, 1,980 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-seven 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. All data and analyses are available on the source files (<https://osf.io/p7avr/>).

## 3.1 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 (), or the maximal sample size. Most laboratories either chose a fixed sample size without applying sequential analysis, or applied sequential analysis and reached their maximal sample size.

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).

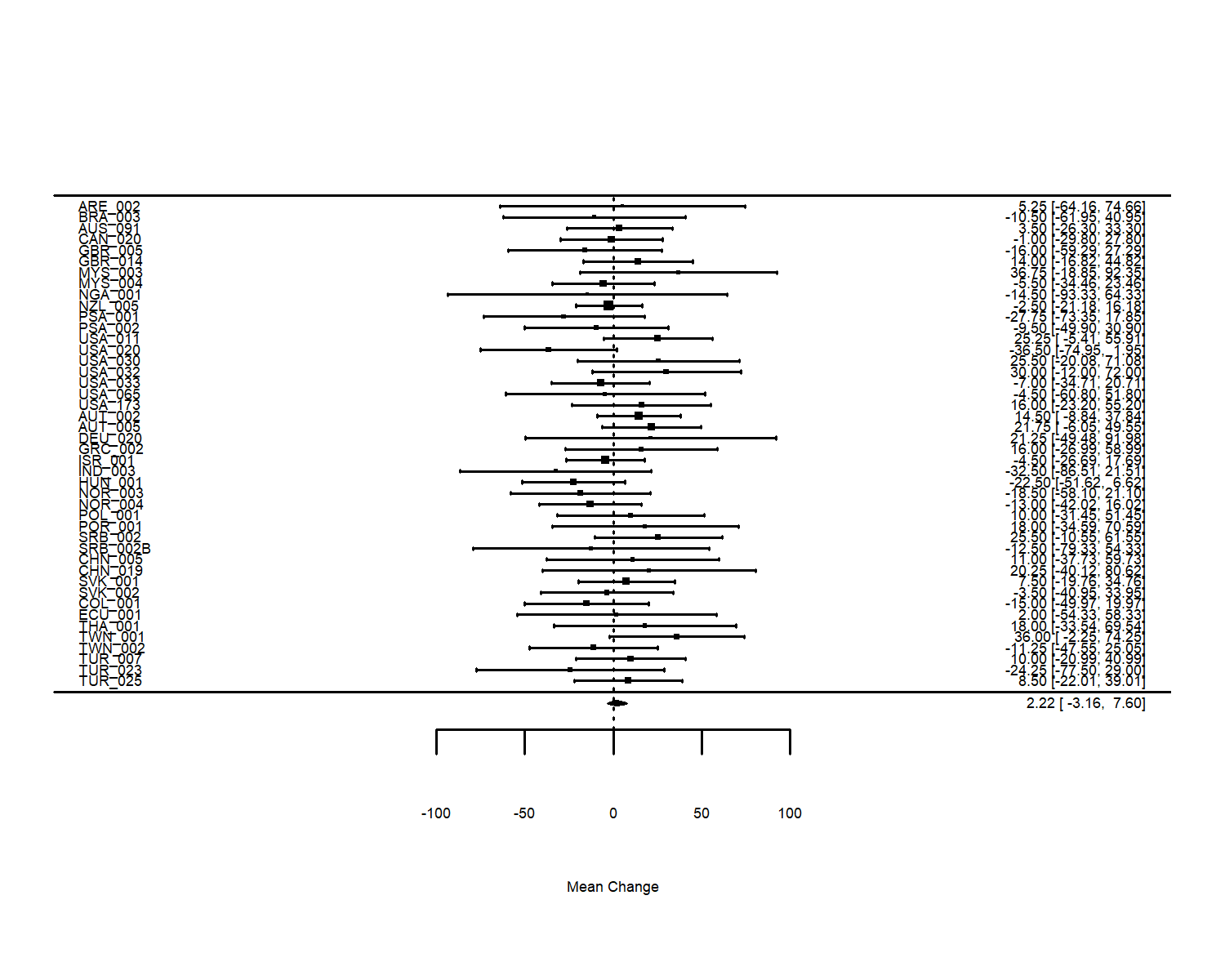
## 3.2 Confirmatory analysis: Inter-lab analysis of final data

**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). 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. Therefore, we used a criterion of 160 ms as the minimum response latency (CITE Hyman-Hicks). 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 ?? about here )

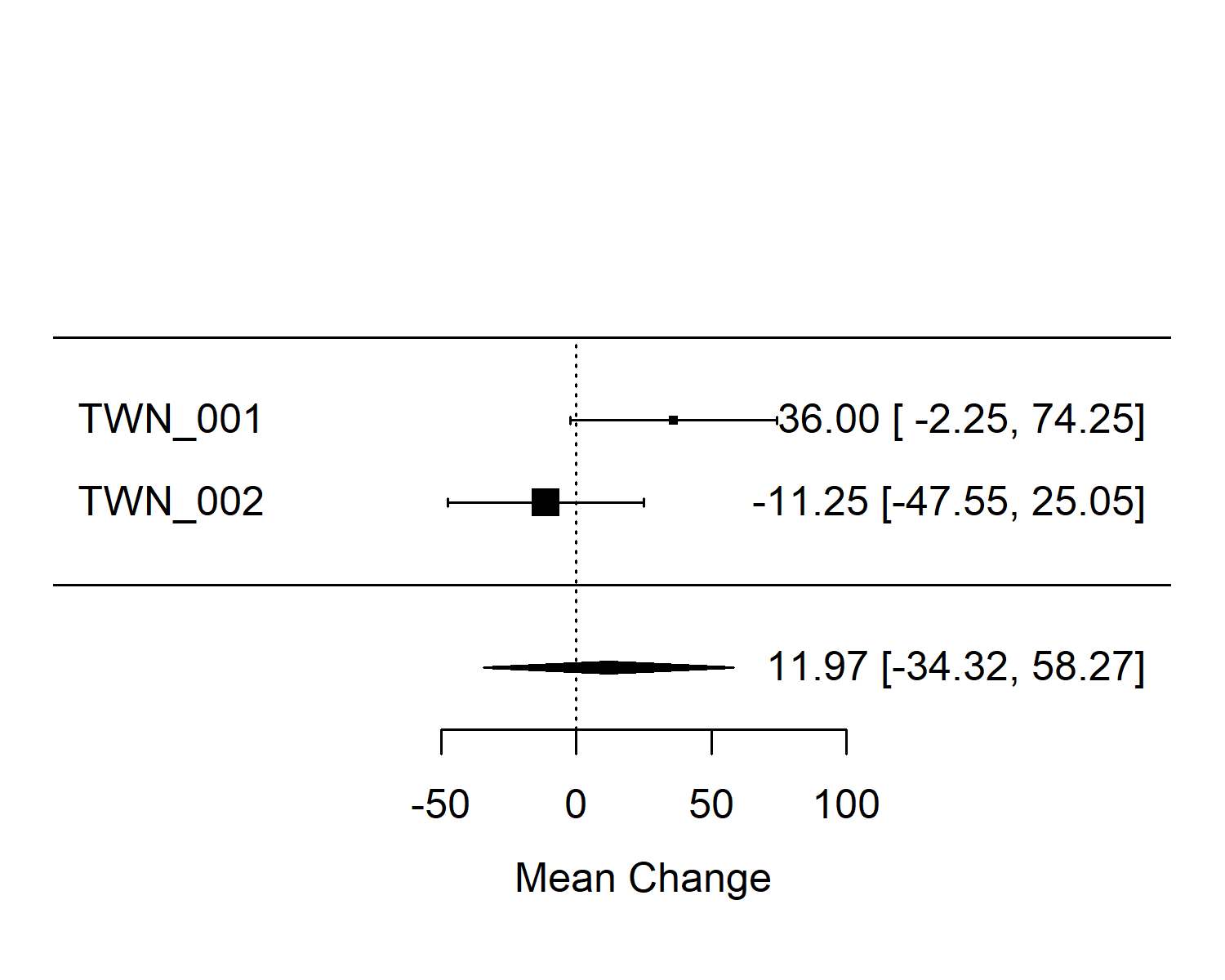
(Insert Table ?? about here )

**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 no 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 Traditional Chinese 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 orienation for all datasets

(Insert Figure 1 about here)



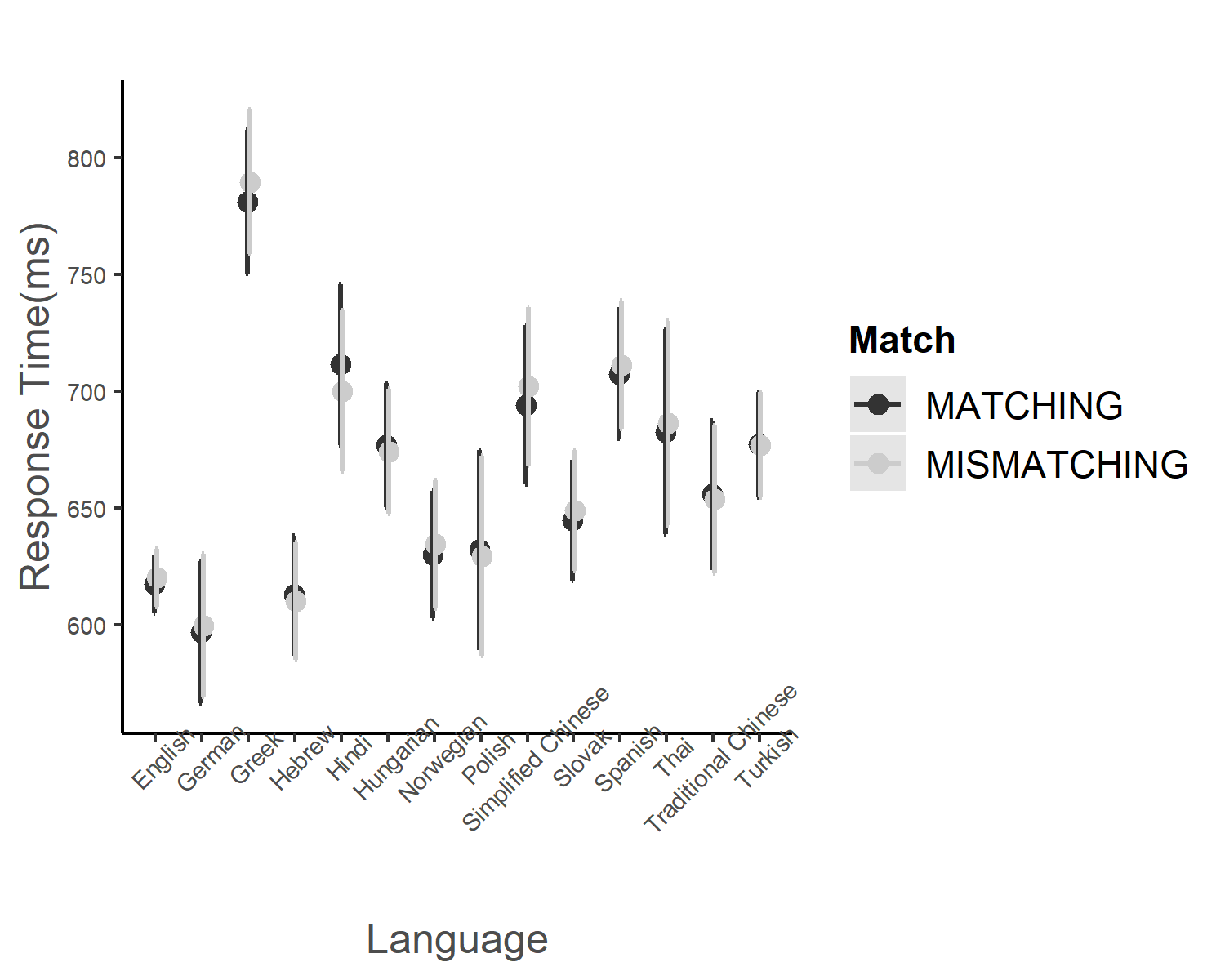
*Figure* *2.*  Meta-analysis on match advantage of object orienation for Traditional Chinese datasets.

(Insert Figure 2 about here)

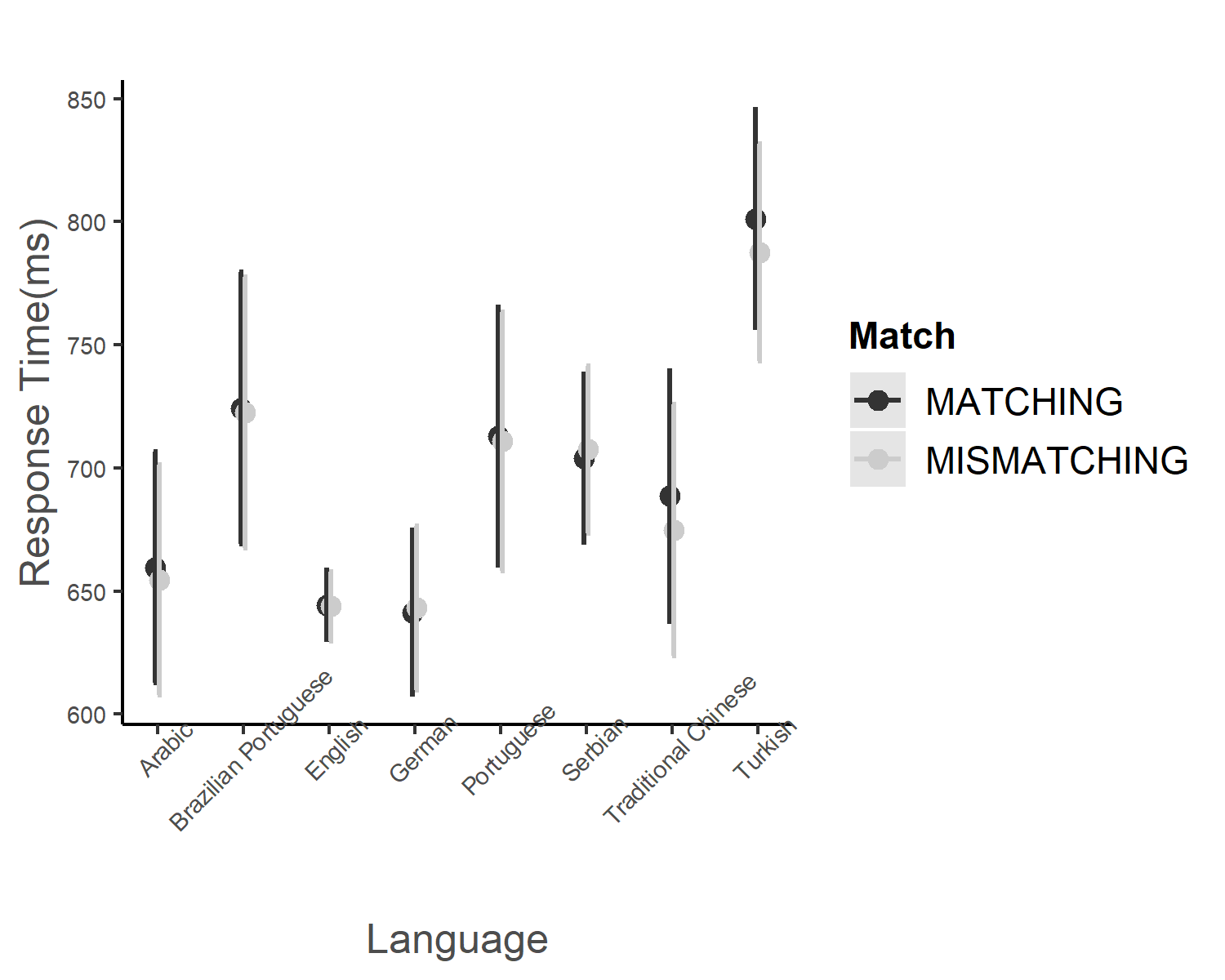
**Evaluating match advantage using linear mixed-effects models.** We excluded data from languages with less than 25 participants in each data source (Portuguese – on-site; Norwegian – web-based). Because the sources of data collection were varied, we evaluated whether one mixed-effects model sufficiently fit all the data. This analysis showed a significant difference between data sources: *b* = -9.724, *SE* = 21.468, *t*(6.996) = -0.453, *p* < .001. Thus, the on-site and the web-based data were analyzed separately.

The final models examined the interaction between language and match advantage in each data source, as reported below. All other models are reported in Appendix 4. It must be acknowledged that the languages with larger sample sizes (see Tables ?? and ??) have more reliable results. Furthermore, most of the languages were underpowered, being far from the 1,200 participants suggested by an a priori power analysis.

In each data source, we compared the fit of the models with and without the random slope of matching condition. Both indicated that the models without the random slope had the best fit. The model from the on-site data revealed no significant effect of match advantage: *b* = 2.898, *SE* = 2.659, *t*(38311.12) = 1.09, *p* = .276. The model from the web-based data also failed to reveal a significant effect: *b* = -5.06, *SE* = 12.673, t( 20798.168 ) = -0.399, *p* = .69. The latter model had a negative coefficient, unlike the on-site data. Although neither effect was significant, the difference in direction resounds with the match advantages and disadvantages found in experiments using the property of color (cf. Connell, 2007; Zwaan & Pecher, 2012).

s

*Figure* *3.*  Response times and standard error in the sentence-picture verification task by match condition in each language (on-site data only).



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

Figure 3 illustrates the response times from the on-site data. Eight languages presented significant intercepts (see “Models including languages” section in Appendix 4).

(Insert Figure 3 about here)

Figure 4 illustrates the response times in the web-based data. One languages presented significant effects (see “Models included languages” section in Appendix 4).

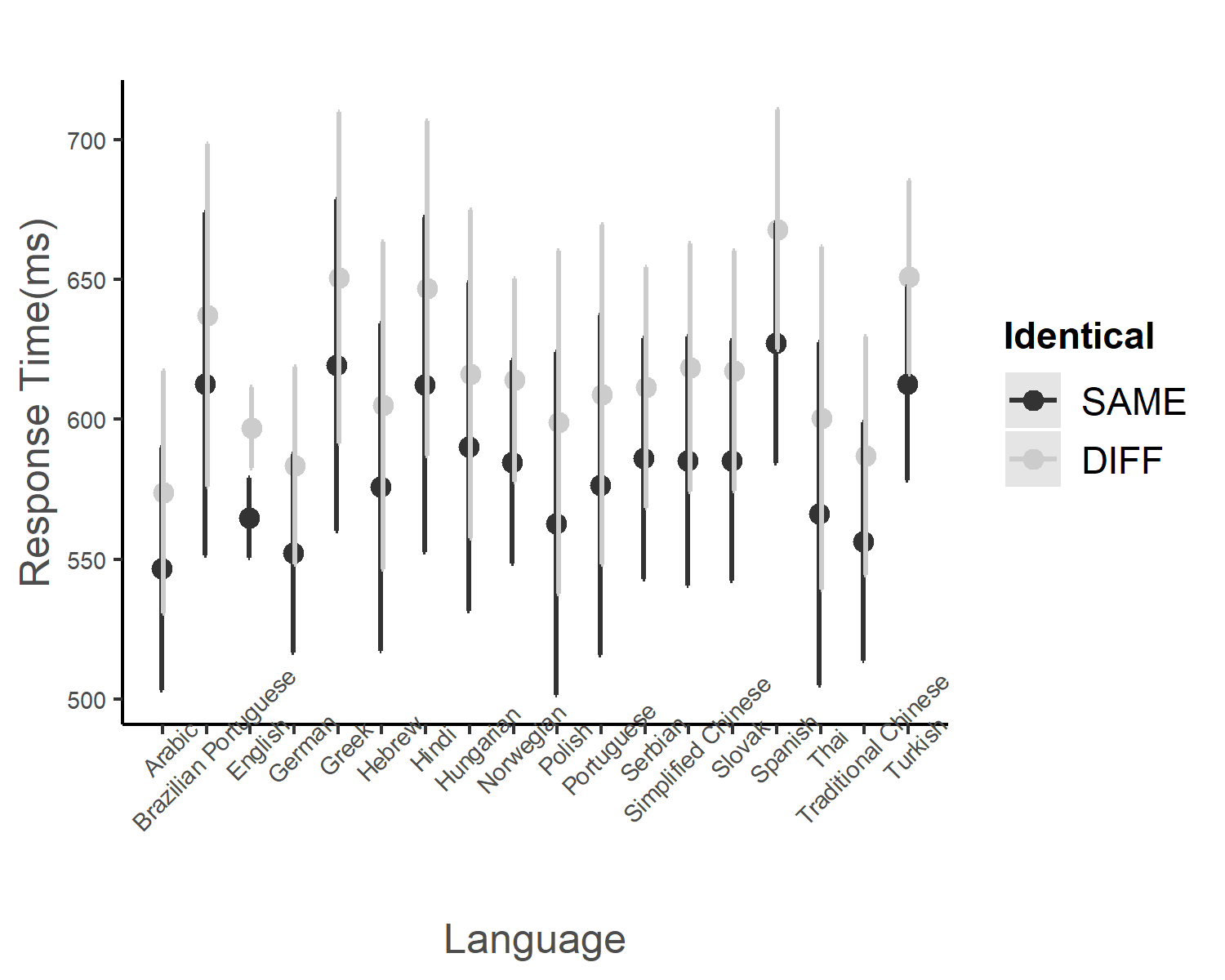
(Insert Figure 4 about here)

**Anecdotal evidence on the match advantage.** In the on-site data, only Greek presented a match advantage, *b* = 5.721, *SE* = 7.204, t( 38319.619 ) = 0.794, *p* = 0.427. It should be noted, however, that these results are not robust due to the underpowered sample sizes (see Discussion).

The median response times in Greek (*M* = 787.12, *SD* = 272.75) and Serbian (*M* = 705.37, *SD* = 287.72) was longer than the average across languages (M = 659.15, SD = 238.40). This might not be coincidental, as according to Yap et al. (2014), longer response times have been associated with larger effects in psycholinguistics (Schilling et al., 1998; Seidenberg, 1985; Tainturier, 1992).

**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( 138198.585 ) = 12.032, *p* < .001. The response times illustrated in Figure 5 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 5 about here)



*Figure* *5.*  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 evaluated the fit of the model with 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 in the analyses of the match advantage were zero (see Appendix 5), the data for building the regression models were the aggregated data by participants.

In the mixed-model linear regression analysis, we decided the best fit model from the model with only predictor, language, and the model with two predictors, languages and imagery scores. Because the analysis of match advantage revealed a difference between data sources, we conducted the mixed-model analysis by the data source respectively. In the analysis of the on-site data, the model with language and imagery scores fit better than the model with language only, *F* = 1.41, *p* = 0.14. In contrast, in the analysis of the web-based data, the model with language and imagery scores had a better fit than the model with language only, *F* (8,1266) = 0.896, *p* = 0.14. In the latter case, the effect of imagery scores was nonsignificant, , 95% CI , , , . Appendix 5 summarized the coefficients of the models included in these analyses.

## 3.3 Exploratory analysis: Language-specific match advantages

Based on our exploratory plan described earlier, we selected the English datasets (*N* = 1,346) and the Traditional Chinese datasets (*N* = 149). For both languages, we are interested in whether the data sources would show differences in the match advantage. Another topic of interest is if the match advantage changed with English dialects, namely American English and British English.

Using the data from 1,346 English speaking participants, we ran a mixed-effects model using match condition, English dialects (American vs. British) and data sources (on-site vs. web-based) as fixed effects. The best fitting model indicated that only data source (on-site vs. web-based) was significant, *b* = -64.173, *SE* = 34.369, t( 16.939 ) = -1.867, *p* < .001. Although the match advantage of orientation was nonsignificant, this exploratory analysis indicated the interaction of orientation match condition and English dialects: *b* = 0.501, *SE* = 6.823, t( 23295.317 ) = 0.073, *p* = 0.941 (see the detailed report in Appendix 4).

We conducted another exploratory mixed-effect model on Traditional Chinese data because this 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. This model indicated that data source was significant, *b* = -30.431, *SE* = 27.565, t( 166.143 ) = -1.104, *p* < .001. The match advantage of orientation was nearly significant: *b* = -11.47, *SE* = 10.955, t( 2667.781 ) = -1.047, *p* = 0.2953, but the interaction of match advantage and data sources was nonsignificant: *b* = 9.302, *SE* = 13.496, t( 2666.773 ) = 0.689, *p* = 0.491 (see the detailed report in Appendix 4). This result suggested that Traditional Chinese 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 Traditional Chinese (Chen et al., 2020), future research on this language could explore any potential linguistic aspects that might result in the match advantage of object orientation and other properties. Although this study is unable to provide further advice, the advantage for the future Traditional Chinese studies would be a precise sample size justification on the participants and stimulus items.

# 4 Discussion

After estimating the match advantage of object orientation across 18 languages, no evidence for a global effect was found, but the meta-analysis and mixed-effect models indicated the marginal match advantage was present in the investigated languages that had at least two datasets. Traditional Chinese, especially, showed marginal results which are consistent with the findings of Chen et al. (2020). This suggests that the match advantage of object orientation for many languages are small. Thus large sample sizes are needed to determine if the match advantage for a language is significantly different from zero. One exception might be a language has some unique features that amplify the match advantage. This requirement is especially onerous in cross-linguistic studies when it is difficult to reach the desirable sample size. In sum, the present results put into question the robustness of cross-linguistic studies.

The second question addressed whether the mental simulation of object orientation could be predicted by mental rotation, which was operationalized as imagery scores. The mixed-effects models indicated that imagery scores were hardly affected by the different languages and data collection procedures. Regarding the planned regression analysis, the imagery scores underpredicted the effect of orientation match. In conclusion, the current findings barely confirm the predictions based on the mental simulation theory.

## 4.1 Measurement issues across platforms

Text

Description automatically generated

This paragraph should go because our new criterion solve this problem.

The precision of web-based experiments has been previously investigated (Anwyl-Irvine et al., 2020; Bridges et al., 2020). In the present study, the responses to the sentence-picture verification task collected on the web (*M* = 1,575, *SD* = 6,780) were roughly twice as long as those collected in labs (*M* = 716, *SD* = 289). Previous studies have also found online responses to be longer than on-site ones, but the current difference is larger. For instance, de Leeuw and Motz (2016) collected response times of just under 100 ms, and found that online responses were 10–40 ms longer than on-site ones. Our primary concern was whether OpenSesame could have caused a higher measurement error in the on-site data than in the web-based data. Of the frequently used desktop applications, OpenSesame Windows version has the highest precision and relatively low variation (see Table 2 in Bridges et al., 2020). Although Bridges et al. did not evaluate the performance of OSWeb, PsychoPy (Peirce et al., 2019), which is the basis for OpenSesame, had higher precision than OpenSesame desktop version. Specifically, it had a 25 to 50 ms lag (see Table 3 of Bridges et al., 2020) in many combinations of operating systems and web browsers. This lag is shorter than the response time difference between the data sources, suggesting that measurement precision was not the source of the timing discrepancies between on-site and web-based data.

According to our meta-analysis, the data from 23 teams revealed marginal match advantages: 16 of these teams collected data on site (e.g., NOR 003) and 9 collected data online (e.g., NZL 005). The former example of the online data is from a lab testing in Norwegian, a language in which the match advantage of orientation had not been studied before, to our knowledge. The latter example is from a lab testing in English, a language that has yielded marginal match advantages of object orientation before (Chen et al., 2020; Stanfield & Zwaan, 2001; Zwaan & Pecher, 2012).

## 4.2 Generalizability and Limitations

To acknowledge deviations from the preregistration, it must be noted that the final data included four more languages than were initially planned, and that some data were collected on the web due to the Covid-19 pandemic, instead of in labs as planned. We do not know of any research suggesting how these deviations could have affected our results. For instance, we reviewed research suggesting that on-site and web-based data need not substantially differ (e.g., de Leeuw & Motz, 2016).

This study reflected the difficulty of investigating cognition across languages, especially when dealing with effects that require large sample sizes (see Loken & Gelman, 2017; Vadillo et al., 2016). Indeed, a fundamental challenge for our project was substantial variation in the number of participants available for the languages we investigated. It must also be noted that the mixed-effects models could have been more conservative by prioritizing the maximal random-effects structure over the achievement of model convergence (Brauer & Curtin, 2018).

Some languages in this study had one participating team only or did not have sufficient data for the exploratory analysis: Arabic, Brazilian Portuguese, European Portuguese, Greek, Hebrew, Hindi, Hungarian, Polish, Serbian, Simplified Chinese and Thai. As a consequence, we do not know if these languages would show a robust match advantage if they were studied by more than two laboratories plus a sample size as large as that which we had for English. Researchers could use stimuli employed for the present study to launch new studies that focus on these specific languages. With rigorous power analyses and large multi-site collaborative projects, future research on specific languages has the potential to provide robust estimates of match advantages associated with a variety of object properties. Such research could serve as a firm foundation for developing a strong theory of mental simulation.

# 5 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>

Barsalou, L. W. (2019). Establishing generalizable mechanisms. *Psychological Inquiry*, *30*(4), 220–230. <https://doi.org/10.1080/1047840X.2019.1693857>

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*, 577–660. <https://doi.org/10.1017/S0140525X99002149>

Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*, 1281–1289. <https://doi.org/10.1098/rstb.2008.0319>

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>

Beilock, S. L., Lyons, I. M., Mattarella-Micke, A., Nusbaum, H. C., & Small, S. L. (2008). Sports experience changes the neural processing of action language. *Proceedings of the National Academy of Sciences*, *105*(36), 13269–13273. <https://doi.org/10.1073/pnas.0803424105>

Brauer, M., & Curtin, J. J. (2018). Linear mixed-effects models and the analysis of nonindependent data: A unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items. *Psychological Methods*, *23*(3), 389–411. <https://doi.org/10.1037/met0000159>

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>

Chu, M., & Kita, S. (2008). Spontaneous gestures during mental rotation tasks: Insights into the microdevelopment of the motor strategy. *Journal of Experimental Psychology: General*, *137*(4), 706–723. <https://doi.org/10.1037/a0013157>

Cohen, D., & Kubovy, M. (1993). Mental rotation, mental representation, and flat slopes. *Cognitive Psychology*, *25*, 351–382. <https://doi.org/10.1006/cogp.1993.1009>

Connell, L. (2007). Representing object colour in language comprehension. *Cognition*, *102*, 476–485. <https://doi.org/10.1016/j.cognition.2006.02.009>

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>

De Koning, B. B., Wassenburg, S. I., Bos, L. T., & Van der Schoot, M. (2017). Mental simulation of four visual object properties: Similarities and differences as assessed by the sentence-picture verification task. *Journal of Cognitive Psychology*, *29*(4), 420–432. <https://doi.org/10.1080/20445911.2017.1281283>

Engelen, J. A. A., Bouwmeester, S., de Bruin, A. B. H., & Zwaan, R. A. (2011). Perceptual simulation in developing language comprehension. *Journal of Experimental Child Psychology*, *110*(4), 659–675. <https://doi.org/10.1016/j.jecp.2011.06.009>

Frick, A., & Möhring, W. (2013). Mental object rotation and motor development in 8- and 10-month-old infants. *Journal of Experimental Child Psychology*, *115*(4), 708–720. <https://doi.org/10.1016/j.jecp.2013.04.001>

Ghandhari, M., Fini, C., Da Rold, F., & Borghi, A. M. (2020). Different kinds of embodied language: A comparison between Italian and Persian languages. *Brain and Cognition*, *142*, 105581. <https://doi.org/10.1016/j.bandc.2020.105581>

Glenberg, A., & Kaschak, M. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, *9*, 558–565. <https://doi.org/10.3758/bf03196313>

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>

Li, Y., & Shang, L. (2017). An ERP study on the mental simulation of implied object color information during Chinese sentence comprehension. *Journal of Psychological Science*, *40*(1), 29–36. <https://doi.org/10.16719/j.cnki.1671-6981.20170105>

Loken, E., & Gelman, A. (2017). Measurement error and the replication crisis. *Science*, *355*(6325), 584–585. <https://doi.org/10.1126/science.aal3618>

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>

Newman, J. (2002). 1. A cross-linguistic overview of the posture verbs ‘Sit,’ ‘Stand,’ and ‘Lie.’ In J. Newman (Ed.), *Typological Studies in Language* (Vol. 51, pp. 1–24). John Benjamins Publishing Company. <https://doi.org/10.1075/tsl.51.02new>

Norcliffe, E., Harris, A. C., & Jaeger, T. F. (2015). Cross-linguistic psycholinguistics and its critical role in theory development: Early beginnings and recent advances. *Language, Cognition and Neuroscience*, *30*(9), 1009–1032. <https://doi.org/10.1080/23273798.2015.1080373>

Ostarek, M., & Huettig, F. (2019). Six Challenges for Embodiment Research. *Current Directions in Psychological Science*, *28*(6), 593–599. <https://doi.org/10.1177/0963721419866441>

Pecher, D., van Dantzig, S., Zwaan, R. A., & Zeelenberg, R. (2009). Language comprehenders retain implied shape and orientation of objects. *The Quarterly Journal of Experimental Psychology*, *62*(6), 1108–1114. <https://doi.org/10.1080/17470210802633255>

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. <https://doi.org/10.3758/s13428-018-01193-y>

Pouw, W. T. J. L., de Nooijer, J. A., van Gog, T., Zwaan, R. A., & Paas, F. (2014). Toward a more embedded/extended perspective on the cognitive function of gestures. *Frontiers in Psychology*, *5*. <https://doi.org/10.3389/fpsyg.2014.00359>

Rommers, J., Meyer, A. S., & Huettig, F. (2013). Object shape and orientation do not routinely influence performance during language processing. *Psychological Science*, *24*(11), 2218–2225. <https://doi.org/10.1177/0956797613490746>

Sato, M., Schafer, A. J., & Bergen, B. K. (2013). One word at a time: Mental representations of object shape change incrementally during sentence processing. *Language and Cognition*, *5*(04), 345–373. <https://doi.org/10.1515/langcog-2013-0022>

Schilling, H. E. H., Rayner, K., & Chumbley, J. I. (1998). Comparing naming, lexical decision, and eye fixation times: Word frequency effects and individual differences. *Memory & Cognition*, *26*(6), 1270–1281. <https://doi.org/10.3758/BF03201199>

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>

Scorolli, C. (2014). Embodiment and language. In L. Shapiro (Ed.), *The Routledge handbook of embodied cognition* (pp. 145–156). Routledge.

Seidenberg, M. (1985). The time course of phonological code activation in two writing systems. *Cognition*, *19*(1), 1–30. <https://doi.org/10.1016/0010-0277(85)90029-0>

Šetić, M., & Domijan, D. (2017). Numerical Congruency Effect in the Sentence-Picture Verification Task. *Experimental Psychology*, *64*(3), 159–169. <https://doi.org/10.1027/1618-3169/a000358>

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>

Tainturier, M. (1992). Educational level and the word frequency effect: A lexical decision investigation\*1. *Brain and Language*, *43*(3), 460–474. <https://doi.org/10.1016/0093-934X(92)90112-R>

Vadillo, M. A., Konstantinidis, E., & Shanks, D. R. (2016). Underpowered samples, false negatives, and unconscious learning. *Psychonomic Bulletin & Review*, *23*(1), 87–102. <https://doi.org/10.3758/s13423-015-0892-6>

Verkerk, A. (2014). *The evolutionary dynamics of motion event encoding.* [Radboud Universiteit Nijmegen]. <https://repository.ubn.ru.nl/handle/2066/127455>

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>

Yap, M. J., Sibley, D. E., Balota, D. A., Ratcliff, R., & Rueckl, J. (2014). Responding to nonwords in the lexical decision task: Insights from the English Lexicon Project. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, No Pagination Specified. <https://doi.org/10.1037/xlm0000064>

Yaxley, R. H., & Zwaan, R. A. (2007). Simulating visibility during language comprehension. *Cognition*, *105*(1), 229–236. <https://doi.org/10.1016/j.cognition.2006.09.003>

Zwaan, R. A. (2014). Replications Should Be Performed With Power and Precision: A Response to Rommers, Meyer, and Huettig (2013). *Psychological Science*, *25*(1), 305–307. <https://doi.org/10.1177/0956797613509634>

Zwaan, R. A., & Madden, C. J. (2005). Embodied sentence comprehension. In D. Pecher & R. A. Zwaan (Eds.), *Grounding cognition: The role of perception and action in memory, language, and thinking* (pp. 224–245). Cambridge University Press.

Zwaan, R. A., & Pecher, D. (2012). Revisiting mental simulation in language comprehension: Six replication attempts. *PLoS ONE*, *7*, e51382. <https://doi.org/10.1371/journal.pone.0051382>

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science*, *13*, 168–171. <https://doi.org/10.1111/1467-9280.00430>

Zwaan, R. A., & van Oostendorp, H. (1993). Do readers construct spatial representations in naturalistic story comprehension? *Discourse Processes*, *16*(1-2), 125–143. <https://doi.org/10.1080/01638539309544832>

1. See details of power analysis in the preregistered plan, p. 13 ~ 15. <https://psyarxiv.com/t2pjv/> [↑](#footnote-ref-1)
2. See the analysis plan in the preregistered plan, p. 19 ~ 20. <https://psyarxiv.com/t2pjv/> [↑](#footnote-ref-2)
3. See the analysis plan in the preregistered plan, p. 21. <https://psyarxiv.com/t2pjv/> [↑](#footnote-ref-3)