Investigating Object Orientation Effects Across 18 Languages

Sau-Chin Chen[1], Erin Buchanan[2], Zoltan Kekecs[3,66], Jeremy K. Miller[4], Anna Szabelska[5], Balazs Aczel[3], Pablo Bernabeu[6,67], Patrick Forscher[7,68], Attila Szuts[3], Zahir Vally[8], Ali H. Al-Hoorie[9], Mai Helmy[10,69], Caio Santos Alves da Silva[11], Luana Oliveira da Silva[11], Yago Luksevicius de Moraes[11], Rafael Ming Chi Santos Hsu[11], Anthonieta Looman Mafra[11], Jaroslava V. Valentova[11], Marco Antonio Correa Varella[11], Barnaby Dixson[12], Kim Peters[12], Nik Steffens[12], Omid Ghasemi[13], Andrew Roberts[13], Robert M. Ross[14], Ian D. Stephen[13,70], Marina Milyavskaya[15], Kelly Wang[15], Kaitlyn M. Werner[15], Dawn L. Holford[16], Miroslav Sirota[16], Thomas Rhys Evans[17], Dermot Lynott[18], Bethany M. Lane[19], Danny Riis[19], Glenn P. Williams[20], Chrystalle B. Y. Tan[21], Alicia Foo[22], Steve M. J. Janssen[22], Nwadiogo Chisom Arinze[23], Izuchukwu Lawrence Gabriel Ndukaihe[23], David Moreau[24], Brianna Jurosic[25], Brynna Leach[25], Savannah Lewis[26], Peter R. Mallik[27], Kathleen Schmidt[25], William J. Chopik[28], Leigh Ann Vaughn[29], Manyu Li[30], Carmel A. Levitan[31], Daniel Storage[32], Carlota Batres[33], Janina Enachescu[34], Jerome Olsen[34], Martin Voracek[34], Claus Lamm[35], Ekaterina Pronizius[35], Tilli Ripp[36], Jan Philipp Röer[36], Roxane Schnepper[36], Marietta Papadatou-Pastou[37], Aviv Mokady[38], Niv Reggev[38], Priyanka Chandel[39], Pratibha Kujur[39], Babita Pande[39], Arti Parganiha[39], Noorshama Parveen[39], Sraddha Pradhan[39], Margaret Messiah Singh[39], Max Korbmacher[40], Jonas R. Kunst[41], Christian K. Tamnes[41], Frederike S. Woelfert[41], Kristoffer Klevjer[42], Sarah E. Martiny[42], Gerit Pfuhl[42], Sylwia Adamus[43], Krystian Barzykowski[43], Katarzyna Filip[43], Patrícia Arriaga[44], Vasilije Gvozdenović[45], Vanja Ković[45], Tao-tao Gan[46], Hu Chuan-Peng[47], Qing-Lan Liu[46], Zhong Chen[48], Fei Gao[48], Lisa Li[48], Jozef Bavoľár[49], Monika Hricová[49], Pavol Kačmár[49], Matúš Adamkovič[50,71], Peter Babinčák[51], Gabriel Baník[51,52], Ivan Ropovik[52,72], Danilo Zambrano Ricaurte[53], Sara Álvarez Solas[54], Harry Manley[55,73], Panita Suavansri[55], Chun-Chia Kung[56], Asil Ali Özdoğru[57], Belemir Çoktok[57], Çağlar Solak[58], Sinem Söylemez[58], Sami Çoksan[59], İlker Dalgar[60], Mahmoud Elsherif[61], Martin Vasilev[62], Vinka Mlakic[63], Elisabeth Oberzaucher[64], Stefan Stieger[63], Selina Volsa[63], Janis Zickfeld[65], and & Christopher R. Chartier[25]

[1] Department of Human Development and Psychology

Tzu-Chi University

[2] Harrisburg University of Science and Technology

[3] Institute of Psychology

ELTE

Eotvos Lorand University

[4] Department of Psychology

Willamette University

[5] Institute of Cognition and Culture

Queen's University Belfast

[6] Department of Psychology

Lancaster University

[7] LIP/PC2S

Université Grenoble Alpes

[8] Department of Clinical Psychology

United Arab Emirates University

[9] Independent Researcher

[10] Psychology Department

College of Education

Sultan Qaboos University

[11] Department of Experimental Psychology
Institute of Psychology
University of Sao Paulo
[12] School of Psychology
University of Queensland
[13] Department of Psychology
Macquarie University
[14] Department of Philosophy
Macquarie University
[15] Department of Psychology
Carleton University
[16] Department of Psychology
University of Essex
[17] School of Social
Psychological and Behavioural Sciences
Coventry University
[18] Department of Psychology
Maynooth University
[19] Division of Psychology
School of Social and Health Sciences
Abertay University
[20] School of Psychology
Faculty of Health Sciences and Wellbeing
University of Sunderland
[21] School of Psychology and Vision Sciences
University of Leicester
[22] School of Psychology

University of Nottingham Malaysia
[23] Department of Psychology
Alex Ekwueme Federal University
[24] School of Psychology
University of Auckland
[25] Department of Psychology
Ashland University
[26] Department of Psychology
University of Alabama
[27] Hubbard Decision Research
[28] Department of Psychology
Michigan State University
[29] Department of Psychology
Ithaca College
[30] Department of Psychology
University of Louisiana at Lafayette
[31] Department of Cognitive Science
Occidental College
[32] Department of Psychology
University of Denver
[33] Department of Psychology
Franklin and Marshall College
[34] Faculty of Psychology
University of Vienna
[35] Department of Cognition
Emotion
and Methods in Psychology

Faculty of Psychology
University of Vienna
[36] Department of Psychology and Psychotherapy
Witten/Herdecke University
[37] School of Education
National and Kapodistrian University of Athens
[38] Department of Psychology and School of Brain Sciences and Cognition
Ben Gurion University
[39] School of Studies in Life Science
Pt. Ravishankar Shukla University
[40] Department of Health and Functioning
Western Norway University of Applied Sciences
[41] Department of Psychology
University of Oslo
[42] Department of Psychology
UiT - The Arctic University of Norway
[43] Institute of Psychology
Jagiellonian University
[44] Iscte-University Institute of Lisbon
[45] Laboratory for Neurocognition and Applied Cognition
Faculty of Philosophy
University of Belgrade
[46] Department of Psychology
Hubei University
[47] School of Psychology
Nanjing Normal University
[48] Faculty of Arts and Humanities

University of Macau
[49] Department of Psychology
Faculty of Arts
Pavol Jozef Šafarik University in Košice
[50] Institute of Social Sciences
CSPS
Slovak Academy of Sciences
[51] Institute of Psychology
University of Presov
[52] Institute for Research and Development of Education
Faculty of Education
Charles University
[53] Faculty of Psychology
Fundación Universitaria Konrad Lorenz
[54] Ecosystem Engineer
Universidad Regional Amazónica Ikiam
[55] Faculty of Psychology
Chulalongkorn University
[56] Department of Psychology
National Cheng Kung University
[57] Department of Psychology
Üsküdar University
[58] Department of Psychology
Manisa Celal Bayar University
[59] Department of Psychology
Erzurum Technical University
[60] Department of Psychology

Ankara Medipol University
[61] Department of Vision Sciences
University of Leicester
[62] Bournemouth University
Talbot Campus
[63] Department of Psychology and Psychodynamics
Karl Landsteiner University of Health Sciences
[64] Department of Evolutionary Anthropology
University of Vienna
[65] Department of Management Aarhus University
[66] Department of Psychology
Lund University
[67] Department of Language and Culture
UiT The Arctic University of Norway
[68] Busara Center for Behavioral Economics
[69] Psychology Department
Faculty of Arts
Menoufia University
[70] Department of Psychology
Nottingham Trent University
[71] University of Jyväskylä
Finland
[72] Faculty of Education
University of Presov
[73] Faculty of Behavioral Sciences
Education
& Languages

HELP University Subang 2

Author Note

Funding statement. Matúš Adamkovič was supported by APVV-20-0319; Robert M. Ross was supported by Australian Research Council (grant number: DP180102384) and the John Templeton Foundation (grant ID: 62631); Zoltan Kekecs was supported by János Bolyai Research Scholarship of the Hungarian Academy of Science; Mahmoud Elsherif was supported by Leverhulme Trust; Glenn P. Williams was supported by Leverhulme Trust Research Project Grant (RPG-2016-093); Ivan Ropovik was supported by NPO Systemic Risk Institute (LX22NPO5101); Krystian Barzykowski was supported by National Science Centre, Poland (2019/35/B/HS6/00528); Gabriel Baník was supported by PRIMUS/20/HUM/009; Patrícia Arriaga was supported by Portuguese National Foundation for Science and Technology (FCT UID/PSI/03125/2019); Monika Hricová was supported by VEGA 1/0145/23.

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 thank the suggestions from the editor and two reviewers on our first and second proposals. CB would like to thank Tyler McGee for help with data collection. PB would like to thank Liam Morgillo for help with data collection.

The authors made the following contributions. Sau-Chin Chen: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Supervision, Validation, Visualization, Writing - original draft, Writing - review & editing; Erin Buchanan: Formal analysis, Project administration, Resources, Software, Validation, Writing - review & editing; Zoltan Kekecs: Project administration, Writing - review &

editing; Jeremy K. Miller: Project administration, Resources, Supervision, Writing - review & editing; Anna Szabelska: Project administration, Writing - original draft, Writing - review & editing; Balazs Aczel: Investigation, Methodology, Resources, Writing - review & editing; Pablo Bernabeu: Investigation, Methodology, Visualization, Writing - review & editing; Patrick Forscher: Methodology, Writing - review & editing; Attila Szuts: Investigation, Methodology, Resources, Writing - review & editing; Zahir Vally: Investigation, Resources, Writing - review & editing; Ali H. Al-Hoorie: Investigation, Resources, Writing - review & editing; Mai Helmy: Investigation, Resources, Writing - review & editing; Caio Santos Alves da Silva: Investigation, Resources, Writing - review & editing; Luana Oliveira da Silva: Investigation, Resources, Writing - review & editing; Yago Luksevicius de Moraes: Investigation, Resources, Writing - review & editing; Rafael Ming Chi Santos Hsu: Investigation, Resources, Writing - review & editing; Anthonieta Looman Mafra: Investigation, Resources, Writing - review & editing; Jaroslava V. Valentova: Investigation, Resources, Writing - review & editing; Marco Antonio Correa Varella: Investigation, Resources, Writing - review & editing; Barnaby Dixson: Investigation, Writing - review & editing; Kim Peters: Investigation, Writing - review & editing; Nik Steffens: Investigation, Writing - review & editing; Omid Ghasemi: Investigation, Writing - review & editing; Andrew Roberts: Investigation, Writing - review & editing; Robert M. Ross: Investigation, Writing - review & editing; Ian D. Stephen: Investigation, Writing - review & editing; Marina Milyavskaya: Investigation, Writing - review & editing; Kelly Wang: Investigation, Writing - review & editing; Kaitlyn M. Werner: Investigation, Writing - review & editing; Dawn L. Holford: Investigation, Writing - review & editing; Miroslav Sirota: Investigation, Writing - review & editing; Thomas Rhys Evans: Investigation, Writing - review & editing; Dermot Lynott: Investigation, Writing - review & editing; Bethany M. Lane: Investigation, Writing - review & editing; Danny Riis: Investigation, Writing - review & editing; Glenn P. Williams: Investigation, Writing - review & editing; Chrystalle B. Y. Tan: Investigation, Writing - review & editing; Alicia

Foo: Investigation, Writing - review & editing; Steve M. J. Janssen: Investigation, Writing - review & editing; Nwadiogo Chisom Arinze: Investigation, Writing - review & editing; Izuchukwu Lawrence Gabriel Ndukaihe: Investigation, Writing - review & editing; David Moreau: Investigation, Writing - review & editing; Brianna Jurosic: Investigation, Writing - review & editing; Brynna Leach: Investigation, Writing - review & editing; Savannah Lewis: Investigation, Writing - review & editing; Peter R. Mallik: Investigation, Writing - review & editing; Kathleen Schmidt: Investigation, Resources, Writing - review & editing; William J. Chopik: Investigation, Writing - review & editing; Leigh Ann Vaughn: Investigation, Writing - review & editing; Manyu Li: Investigation, Writing - review & editing; Carmel A. Levitan: Investigation, Writing - review & editing; Daniel Storage: Investigation, Writing - review & editing; Carlota Batres: Investigation, Writing - review & editing; Janina Enachescu: Investigation, Resources, Writing - review & editing; Jerome Olsen: Investigation, Resources, Writing - review & editing; Martin Voracek: Investigation, Resources, Writing - review & editing; Claus Lamm: Investigation, Resources, Writing - review & editing; Ekaterina Pronizius: Investigation, Writing - review & editing; Tilli Ripp: Investigation, Writing - review & editing; Jan Philipp Röer: Investigation, Writing - review & editing; Roxane Schnepper: Investigation, Writing - review & editing; Marietta Papadatou-Pastou: Investigation, Resources, Writing - review & editing; Aviv Mokady: Investigation, Resources, Writing - review & editing; Niv Reggev: Investigation, Resources, Writing - review & editing; Priyanka Chandel: Investigation, Resources, Writing - review & editing; Pratibha Kujur: Investigation, Resources, Writing - review & editing; Babita Pande: Investigation, Resources, Supervision, Writing - review & editing; Arti Parganiha: Investigation, Resources, Supervision, Writing - review & editing; Noorshama Parveen: Investigation, Resources, Writing - review & editing; Sraddha Pradhan: Investigation, Resources, Writing - review & editing; Margaret Messiah Singh: Investigation, Writing - review & editing; Max Korbmacher: Investigation, Writing - review & editing; Jonas R. Kunst: Investigation, Resources, Writing - review & editing; Christian K. Tamnes:

Investigation, Resources, Writing - review & editing; Frederike S. Woelfert: Investigation, Writing - review & editing; Kristoffer Klevjer: Investigation, Writing - review & editing; Sarah E. Martiny: Investigation, Writing - review & editing; Gerit Pfuhl: Investigation, Resources, Writing - review & editing; Sylwia Adamus: Investigation, Resources, Writing - review & editing; Krystian Barzykowski: Investigation, Resources, Supervision, Writing - review & editing; Katarzyna Filip: Investigation, Resources, Writing - review & editing; Patrícia Arriaga: Funding acquisition, Investigation, Resources, Writing - review & editing; Vasilije Gvozdenović: Investigation, Resources, Writing - review & editing; Vanja Ković: Investigation, Resources, Writing - review & editing; Tao-tao Gan: Investigation, Writing - review & editing; Hu Chuan-Peng: Investigation, Writing - review & editing; Qing-Lan Liu: Investigation, Writing - review & editing; Zhong Chen: Investigation, Writing - review & editing; Fei Gao: Investigation, Resources, Writing - review & editing; Lisa Li: Investigation, Resources, Writing - review & editing; Jozef Bavoľár: Investigation, Resources, Writing - review & editing; Monika Hricová: Investigation, Resources, Writing - review & editing; Pavol Kačmár: Investigation, Resources, Writing - review & editing; Matúš Adamkovič: Investigation, Writing - review & editing; Peter Babinčák: Investigation, Writing - review & editing; Gabriel Baník: Investigation, Writing - review & editing; Ivan Ropovik: Investigation, Writing - review & editing; Danilo Zambrano Ricaurte: Investigation, Resources, Writing - review & editing; Sara Álvarez Solas: Investigation, Resources, Writing - review & editing; Harry Manley: Investigation, Resources, Writing - review & editing; Panita Suavansri: Investigation, Resources, Writing - review & editing; Chun-Chia Kung: Investigation, Resources, Writing - review & editing; Asil Ali Özdoğru: Investigation, Resources, Writing - review & editing; Belemir Çoktok: Investigation, Resources, Writing - review & editing; Çağlar Solak: Investigation, Writing - review & editing; Sinem Söylemez: Investigation, Writing - review & editing; Sami Çoksan: Investigation, Resources, Writing - review & editing; İlker Dalgar: Resources, Writing - review & editing; Mahmoud Elsherif: Writing - review & editing; Martin Vasilev: Writing -

review & editing; Vinka Mlakic: Resources, Writing - review & editing; Elisabeth Oberzaucher: Resources, Writing - review & editing; Stefan Stieger: Resources, Writing - review & editing; Selina Volsa: Resources, Writing - review & editing; Janis Zickfeld: Resources, Writing - review & editing; Christopher R. Chartier: Investigation, Project administration, Writing - review & editing.

Correspondence concerning this article should be addressed to Sau-Chin Chen. 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. Mental representation is often measured with the sentence-picture verification task, wherein participants first read a sentence that implies the object property (i.e., shape and orientation). Participants then respond to an image of an object by indicating whether it was an object from the sentence or not. Previous studies have shown matching advantages for shape, but findings concerning object orientation have not been robust across languages. This registered report investigated the match advantage of object orientation across 18 languages in nearly 4,000 participants. The preregistered analysis revealed no compelling evidence for a match advantage for orientation across languages. Additionally, the match advantage was not predicted by mental rotation scores. Overall, the results did not support current mental simulation theories.

Keywords: cross-lingual research, language comprehension, mental rotation, mental simulation

Investigating Object Orientation Effects Across 18 Languages

Mental simulation of object properties is a major topic in conceptual processing research (Ostarek & Huettig, 2019; Scorolli, 2014). Theoretical frameworks of conceptual processing describe the integration of linguistic representations and situated simulation (e.g., reading about bicycles integrates the situation in which bicycles would be used, Barsalou, 2008; Zwaan, 2014a). Proponents of situated cognition contend that perceptual representations can be generated during language processing (Barsalou, 1999; Wilson, 2002), as cognition is thought to be an interaction of the body, environment, and processing (Barsalou, 2020). Given this definition of situated cognition, it is important to investigate previously established embodied cognition effects across multiple environments (in this case, languages and cultures), especially as the credibility revolution has indicated that not all published findings are replicable (Vazire, 2018).

One empirical index of situated simulation is the mental simulation effect measured in the sentence-picture verification task (see Figure 1). This task requires participants to read a probe sentence displayed on the screen. On the following screen, participants see a picture of an object and must verify whether the object was mentioned in the probe sentence. Verification response times are used to test the mental simulation effect, which occurs when people are faster to respond to pictures that match the properties implied by the probe sentences. For example, the orientation implied by the sentence *Tom hammered the nail into the wall* would be matched if the following picture showed a horizontally-oriented nail rather than a vertically-oriented one. The opposite would be true of the sentence *Tom hammered the nail into the floor plank*.

Mental simulation effects have been demonstrated for object shape (Zwaan et al., 2002), color (Connell, 2007), and orientation (Stanfield & Zwaan, 2001). Subsequent replication studies revealed consistent results for shape but inconsistent findings for color and orientation effects (De Koning et al., 2017; Rommers et al., 2013; Zwaan & Pecher,

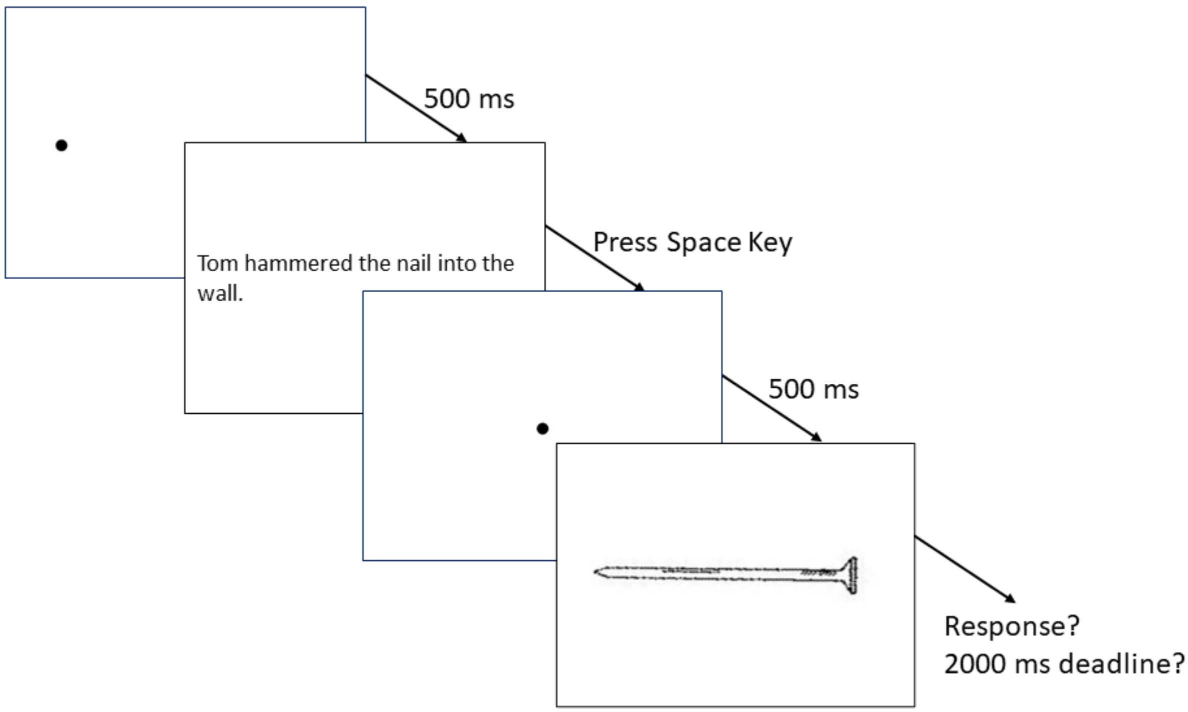

Figure 1

Procedure of the sentence-picture verification task, with an example of matching orientation.

2012). Existing theoretical frameworks do not provide much guidance regarding the potential causes for this discrepancy. With the accumulating concerns about the lack of reproducibility (e.g., Kaschak & Madden, 2021), researchers have found it challenging to reconcile the theory of mental simulation with the failures to replicate some of the effects (e.g., Morey et al., 2022). In an empirical discipline like cognitive science, a theory requires the support of reproducible results.

The reliability of match advantage effects seems to vary depending on both the object properties and the languages under study. Mental simulation effects for object shape have consistently been found in English (Zwaan et al., 2017; 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: namely, positive evidence in English (Stanfield & Zwaan, 2001; Zwaan & Pecher, 2012) and in Chinese (Chen et al., 2020), and null evidence in Dutch (De Koning et al., 2017; Rommers et al., 2013) and in German as second language (Koster et al., 2018). Among studies on shape and orientation, the effects of object orientation have been smaller than those of object shape (e.g., $d = 0.10$ vs. 0.17 in Zwaan & Pecher, 2012; $d = 0.07$ vs. 0.27 in De Koning et al., 2017). To understand the causes for the discrepancies among object properties and languages, it is imperative to consider the cross-linguistic and experimental factors of the sentence-picture verification task.

Cross-linguistic, Methodological, and Cognitive Factors

Several factors might contribute to cross-linguistic differences in the match advantage of object orientation. First, languages differ in how they encode motion and placement events in sentences (Newman, 2002; Verkerk, 2014). Second, the potential role of mental rotation as a confound has been considered (Rommers et al., 2013). We expand on how linguistic, methodological, and cognitive factors hinder the improvement of theoretical frameworks below.

Linguistic 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*). Languages encode motion events in different ways, and grammatical differences between lexical encodings could explain different match advantage results. According to Verkerk (2014), Germanic languages (e.g., Dutch, English, and 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 the path in the verb (e.g., *crossing*, *exiting*). Crucially, 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 (2014) 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 contain 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. In English, for example, the verb *put* does not convey a specific orientation in the sentences *She put the book on the table* and *She put the bottle on the table.* However, in German and Dutch, speakers preferred the verbs *laid* or *stood* in the above sentences. In this case, 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 the current theories (e.g., Language and Situated Simulation, Barsalou, 2008) have not addressed the potential influence of linguistic aspects such as the lexical encoding of placement.

Methodological factors. Inconsistent findings on the match advantage of object orientation may be due to variability in task design. For example, studies failing to detect the match advantage may not have required participants to verify the probe sentence after the response to the target picture (see Zwaan, 2014a). Without such a verification, participants might have paid less attention to the meaning of the probe sentences, in which they would have been less likely to form a mental representation of the objects (e.g., Zwaan & van Oostendorp, 1993). In this regard, variability originating from differences in the characteristics of experiments can substantially influence the results (Barsalou, 2019; Kaschak & Madden, 2021).

Cognitive Factors. Since Stanfield and Zwaan (2001) showed a match advantage of object orientation, later studies on this topic have examined the association between the match advantage and alternative cognitive mechanisms rather than situated simulation. One of these potential mechanisms is spatial cognition, which can be measured with mental rotation tasks. Indeed, 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). Some previous studies have drawn on mental rotation to study mental simulation. For instance, De Koning et al. (2017) observed that the effectiveness of mental rotation increased with the size of the depicted object. Chen et al. (2020) examined the implication of this finding for the match advantage of object orientation (Stanfield & Zwaan, 2001), and implemented a picture-picture verification task using the mental rotation paradigm (D. Cohen & Kubovy, 1993). In each trial, two pictures appeared on opposite sides of the screen. Participants had to verify whether the pictures represented identical or different objects.

Chen et al. (2020) not only revealed shorter verification times for matching orientations (i.e., two identical pictures presented in horizontal or vertical orientation) but also replicated the larger effect for larger objects (i.e., pictures of bridges versus pictures of pens). The results were consistent across the three languages investigated: English, Dutch and Chinese. Compared to the results of sentence-picture verification and picture-picture verification, Chen et al. (2020) converted the picture-picture verification times to the mental rotation scores that were the discrepancy of verification times between the identical and different orientations[1]. Their analysis showed that mental rotation affected the Dutch participants' sentence-picture verification performance. With the measurement of mental rotation scores, we explore the association of spatial cognition and the effect of orientation in comprehension across the investigated languages.

Purposes of this study

To scrutinize the discrepancies in findings across languages and cognitive factors, we examined the reproducibility of the object orientation effect in a multi-lab collaboration. Our pre-registered plan aimed at detecting a general match advantage of object orientation across languages and evaluated the magnitude of match advantage in each specific language. Additionally, we examined whether the match advantages were related to the mental rotation index. 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 are differences in the mental rotation index associated with the match advantage across languages?

Method

Hypotheses and Design

The study design for the sentence-picture and picture-picture verification tasks 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 response time. The time difference between match conditions in each task is the measurement of mental simulation effects (for the sentence-picture task) and mental rotation scores (for the picture-picture task). We did

[1] In the pre-registered plan, we used the term "imagery score" but this term was confusing. Therefore, we used "mental rotation scores" instead of "imagery scores" in the final report.

not select languages systematically, but instead based on our collaboration recruitment with the Psychological Science Accelerator (PSA, Moshontz et al., 2018).

We pre-registered the following hypotheses:

(1) In the sentence-picture verification task, we expected response times 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) Based on the assumption that 'mental rotation is a general cognitive function', we expect equal mental rotation scores across languages, but no association between mental rotation scores and mental simulation effects (see Chen et al., 2020).

Participants

We performed a pre-registered power analysis, which sought to achieve a power of 80% in a directional one-sample t-test. When an effect size of $d = 0.20$ was hypothesized, a sample size of $N = 156$ was required. Instead, for a hypothetical effect size $d = 0.10$, a sample size of $N = 620$ was required.'In addition, a power analysis tailored to mixed-effects models was performed. The effect size hypothesized in this analysis was equal to that observed by Zwaan and Pecher (2012), and the number of items was 100 (i.e., 24 planned items nested within at least five languages). The result revealed that a sample size of $N =$ 400 would be required to achieve a power of 90%. We expected laboratories to show differences in orientation effects, and therefore, the mixed effect analysis treated the laboratories as a random variable to account for different analyses. The laboratories were allowed to follow a secondary plan: a team collected at least their pre-registered 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 $BF_{10} = 10$ or 0.10)[2].

We collected data in 18 languages from 50 laboratories. Each laboratory chose a maximal sample size and an incremental n for sequential analysis before their data collection. Because the pre-registered power analysis did not match the final analysis plan, we additionally completed a sensitivity analysis to ensure the sample size was adequate to detect small effects, and the results indicated that if the observe orientation difference in reaction time between the different orientations was overall 2.36 ms or higher, the effect would be detected as significant. Appendix A summarizes the details of the sensitivity analysis.

The original sample sizes are presented in Table 1 for the teams that provided raw data to the project. Data collection proceeded in two broad stages: initially we collected data in the laboratory. However, when the global COVID-19 pandemic made this practice impossible to continue, we moved data collection online. In total, 4,248 unique participants completed the present study with 2,843 completing the in-person version and 1,405 completing the online version[3]. The in-person version included 35 research teams and the online version included 19 with 50 total teams across both data collection methods (i.e., some labs completed both in-person and online data collection). Based on

[2] See details of power analysis in the pre-registered plan, pp. 13 - 15. https://psyarxiv.com/t2pjv/

[3] Data for this study was collected together with another unrelated study (Phills et al., 2022) during the same data collection session, with the two studies using different data collection platforms. The demographic data was collected within the platform of the other study during the in-person sessions. Some participants only completed the Phills et al. study and dropped out without completing the present study, and there were also some data entry errors in the demographic data. Thus, the demographic data of some participants who took the present study are missing or unidentifiable ($n = 39$ cannot be matched to a lab, $n =$ 2,053 were missing gender information, and $n = 332$ were missing age information). Importantly, this does not affect the integrity of the experimental research data.

recommendations from the research teams (TUR_007, TWN_002), two sets of data were excluded from all analyses due to participants being non-native speakers. Figure 2 provides a flow chart for participant exclusion and inclusion for analyses. 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://codeocean.com/capsule/3994879). Appendix B summarizes the average characteristics by language and laboratory.

Materials

Sentences. 24 critical sentence pairs (48 total sentences) were included in this study following Stanfield and Zwaan (2001). Each pair consisted of versions that differed in their implied orientation of the object embedded in the sentence. For instance, the sentence *The librarian put the book back on the table* - which implies a horizontal orientation - had a counterpart in the sentence *The librarian put the book back on the shelf* - which implies a vertical orientation. Another two sets of 24 sentences were included as filler sentences for the task demand. These sentences were not matched to any particular orientation but included a potential object for depiction. For example, *After a week the painting arrived by mail*, and *The flowers that were planted last week had survived the storm* were included as filler sentences. Each participant was shown 24 critical sentences and 24 filler sentences in the study. The filler sentences were included to counterbalance the number of yes-no answers to create an even 50% ratio.

Pictures. The study included 24 critical matched pictures that only varied in their orientation (vertical/horizontal) for a total of 48 critical pictures (from Zwaan & Pecher, 2012). These pictures were matched to their respective sentences for implied orientation. *The librarian put the book back on the table* was matched with a horizontally-oriented book, while *The librarian put the book back on the shelf* was matched with a vertically-oriented book. For counterbalancing, the mismatch between picture orientation and sentence was

Table 1

Demographic and Sample Size Characteristics

Language	SP_{Trials}	PP_{Trials}	SP_N	PP_N	$Demo_N$	$Female_N$	$Male_N$	M_{Age}	SD_{Age}
Arabic	2544	2544	106	106	107	42	12	32.26	18.59
Brazilian Portuguese	1200	1200	50	50	50	36	13	30.80	8.73
English	45189	45312	1884	1888	2055	1360	465	21.71	3.85
German	5616	5616	234	234	248	98	26	22.34	3.40
Greek	2376	2376	99	99	109	0	0	33.86	11.30
Hebrew	3576	3571	149	149	181	0	0	24.25	9.29
Hindi	1896	1896	79	79	86	57	27	21.66	3.46
Magyar	3610	3816	151	159	168	3	1	21.50	2.82
Norwegian	3576	3576	149	149	154	13	9	25.22	6.40
Polish	1368	1368	57	57	146	0	0	23.25	7.96
Portuguese	1488	1464	62	61	55	26	23	30.74	9.09
Serbian	3120	3120	130	130	130	108	21	21.38	4.50
Simplified Chinese	2040	2016	85	84	96	0	1	21.92	4.68
Slovak	3881	3599	162	150	325	1	0	21.77	2.33
Spanish	3120	3096	130	129	146	0	0	21.73	3.83
Thai	1200	1152	50	48	50	29	9	21.54	3.81
Traditional Chinese	3600	3600	150	150	186	69	46	20.89	2.44
Turkish	6456	6432	269	268	274	36	14	21.38	4.59

Note. SP = Sentence Picture Verification, PP = Picture Picture Verification. Sample sizes for demographics may be higher than the sample size for the this study, as participants could have only completed the bundled experiment. Additionally, not all entries could be unambigously matched by lab ID, and therefore, demographic sample sizes could also be less than data collected.

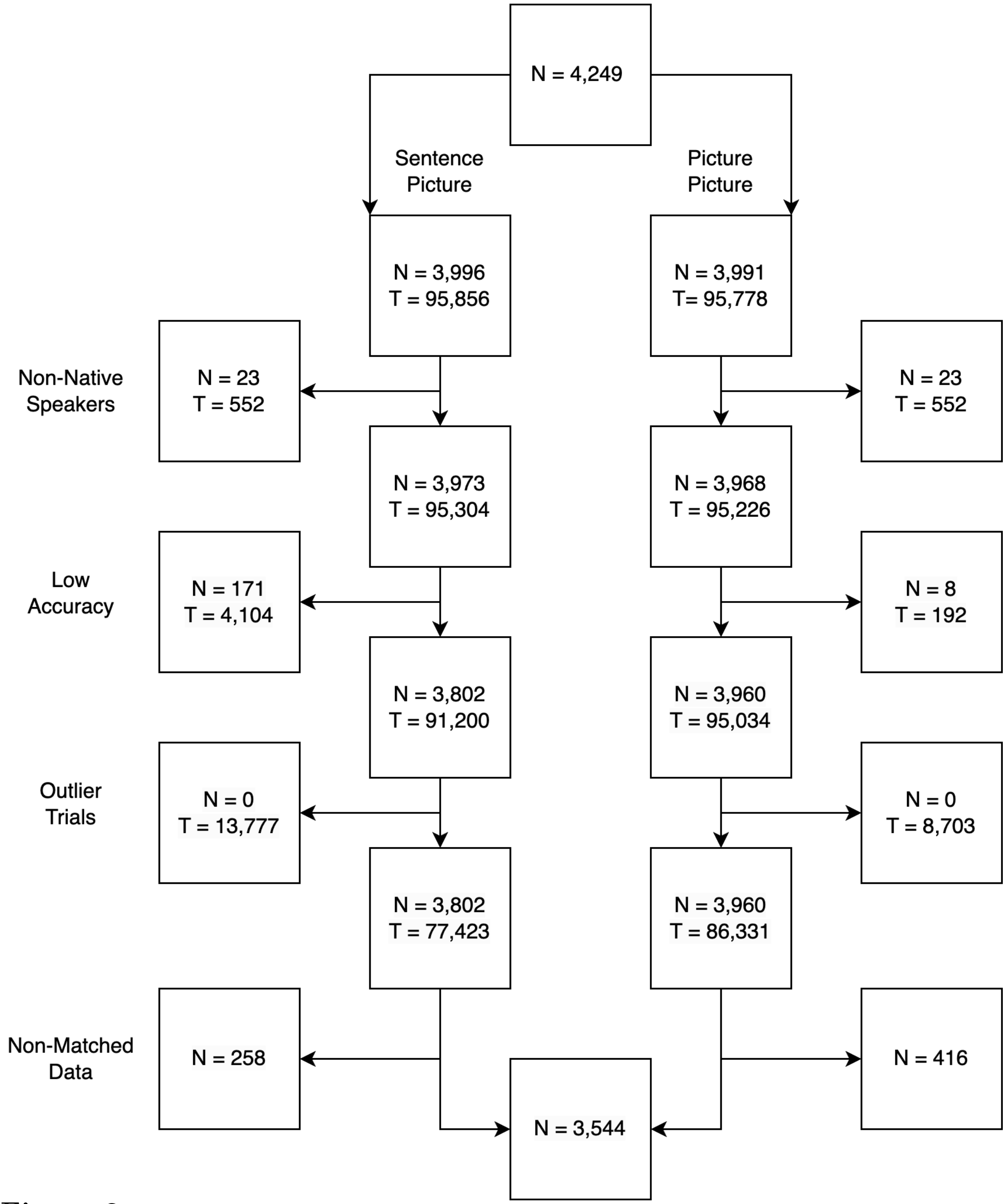

Figure 2

Sample size and exclusions. N = number of unique participants, T = number of trials. The final combined sample was summarized to a median score for each match/mismatch condition, and therefore, includes one summary score per person.

Table 2

Trial conditions for the Sentence-Picture and Picture-Picture Verification Task

Condition	Item 1	Item 2	Answer	Number
Sentence-Picture Critical Match	Critical Sentence: Horizontal	Critical Picture: Horizontal	Yes	6
Sentence-Picture Critical Match	Critical Sentence: Vertical	Critical Picture: Vertical	Yes	6
Sentence-Picture Critical Mismatch	Critical Sentence: Horizontal	Critical Picture: Vertical	Yes	6
Sentence-Picture Critical Mismatch	Critical Sentence: Vertical	Critical Picture: Horizontal	Yes	6
Sentence-Picture Filler	Sentence	Picture	No	24
Picture-Picture Critical Match	Critical Picture: Horizontal	Critical Picture: Horizontal	Yes	6
Picture-Picture Critical Match	Critical Picture: Vertical	Critical Picture: Vertical	Yes	6
Picture-Picture Critical Mismatch	Critical Picture: Horizontal	Critical Picture: Vertical	Yes	6
Picture-Picture Critical Mismatch	Critical Picture: Vertical	Critical Picture: Horizontal	Yes	6
Picture-Picture Filler	Picture	Picture	No	24

created, and the book would be shown in the respective opposite orientation (see orientation pairs at https://osf.io/utqxb). Another 48 pictures were included for the fillers which were unrelated to the corresponding sentence. Therefore, the answer to critical pairs was always “yes”, while the filler sentence-picture combinations answer was always “no”.

Picture-Picture Trials. The picture-picture verification task used the same object pictures as the above task. The 24 critical picture pairs were included as match trials and were counterbalanced such as half the time they appeared with the same object and orientation (i.e., the same picture), and half the time with the opposite orientation (i.e., horizontal and vertical). The filler pictures were randomly paired to create mismatch trials. Table 2 shows the counterbalancing and combinations for trials.

Procedure

Sentence-Picture Task. The sentence-picture verification task was always administered first. This task began with six practice trials. Each trial started with a left-aligned vertically-centered fixation point displayed for 1,000 ms, immediately followed by the probe sentence. The sentence remained on the screen until the participant pressed the space key, acknowledging that they had read the sentence. Then, the object picture was presented in the center of the screen until the participant responded, or it disappeared after two seconds. Participants were instructed to verify, as quickly and accurately as possible, whether the object on screen had been mentioned in the probe sentence. Following Stanfield and Zwaan (2001), a memory check test was carried out after every three to eight trials to ensure that participants had read each sentence carefully.

As shown in Table 2, the trials for the sentence-picture task were created by counterbalancing the sentence implied orientation (vertical, horizontal) by the pictured object orientation creating a fully-crossed combination between matching sentences and objects. Therefore, each participant only saw one of the four possible combinations (sentence orientation 2 x object orientation 2). For the filler items, sentences and pictures were randomly assigned in two separate patterns, and these were included with the critical pairs. Stimuli lists were created in Excel, and this information can be found at https://osf.io/utqxb.

Translation of Sentences. The translation of probe sentences followed our pre-registered plan. Every non-English language coordinator was required to recruit at least four translators who were fluent in both English and the target language. Every language coordinator supervised the translators using the Psychological Science Accelerator guidelines (https://psysciacc.org/translation-process/). In addition, the coordinator and participating laboratories consulted about each of the following points:

1) Four translators could denote the items that are unfamiliar to a particular language based on object familiarity ratings. The two forward translators would suggest alternative probe sentences and object pictures to replace the unfamiliar objects. The two backward translators would evaluate the suggested items.

2) Some objects in a particular language have different spellings or pronunciations among countries and geographical zones due to dialect. For example, American

speakers tend to write *tire* whereas British speakers tend to write *tyre*. Every coordinator would mark these local translations in the final version of translated materials. Participating laboratories could replace the names to match the local dialect.

Picture-Picture Task. Next, the picture-picture verification task was administered. 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 two seconds elapsed. As shown in Table 2, four possible combinations of critical orientations could be shown with the picture (same, different) by orientation (same, different). Each participant only saw one of the critical combinations, and filler items were randomly paired in two combinations to match. The stimuli lists can be found at https://osf.io/utqxb.

Software Implementation. The study was executed using OpenSesame software for millisecond timing (Mathôt et al., 2012). After data collection moved 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; Mathôt & March, 2022). We proceeded with the online study from February to June 2021 after the changes in the procedure were approved by the journal editor and reviewers. Following the literature, we did not anticipate any theoretically important differences between the two data collection methods (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

To test Hypothesis 1, our first planned analysis[4] used a random-effects meta-analysis model that estimated the match advantage across laboratories and languages. The meta-analysis summarized the median reaction times by match condition to determine the effect size by laboratory. The following formula was used:

$$d = \frac{Mdn_{Mismatch} - Mdn_{Match}}{\sqrt{MAD^2_{Mismatch} + MAD^2_{Match} - 2 \times r \times MAD_{Mismatch} \times MAD_{Match}}} \times \sqrt{2 \times (1 - r)}$$

where d is Cohen's d (Fritz et al., 2012), Mdn is Median, MAD is median absolute deviation, and r is correlation between match and mismatch condition. Meta-analytic effect sizes were computed for those languages that had data from more than one team.

Continuing to test Hypothesis 1, next, we ran planned mixed-effects models using each individual response time from the sentence-picture verification task as the dependent variable. In each analysis, we first built a simple linear regression model with a fixed intercept-only. Then, we systematically added random intercepts and fixed effects, arriving at the final model. First, the random intercepts were added to the model one-by-one in the following order: participant ID, target, laboratory ID, and finally language. See below section for decision criteria for determining the final random-effect structure. Then, the fixed effect of matching condition (match vs. mismatch) was added to the model. Language-specific mixed-effects models were conducted in the same way if the meta-analysis showed a significant orientation effect.

According to the pre-registration, we planned to test Hypothesis 2 by first evaluating the equality of mental rotation scores across languages using an ANOVA.

[4] See the analysis plan in the pre-registered plan, pp. 19 - 20, https://psyarxiv.com/t2pjv/. This plan was changed to a random-effects model to ensure that we did not assume the exact same effect size for each language and lab.

However, this plan was updated to use mixed models instead due to the nested structure of the data (Gelman, 2006). The same analysis plan was used for model building and selection as described above for the sentence-picture verification task.

To further assess Hypothesis 2, the last planned analysis was to use mental rotation scores for the prediction of mental stimulation with an interaction between language and mental rotation scores computed from the picture-picture task to determine if there were differences in prediction of match advantage in the sentence-picture task. Here, we used a mixed-effects model as well to control for the random effect of the data collection lab, and with language, mental rotation score, and their interaction as fixed effect predictors.

Decision criterion for model selection and hypothesis testing. The inclusion of both random and fixed effects in models was assessed using model comparison based on the Akaike information criterion (AIC). While this method is less conservative than alternatives such as the likelihood ratio test (Matuschek et al., 2017), the AIC was deemed appropriate due to the modest effect sizes that tend to be produced by mental simulation effects, and the limited sample sizes in the present study (albeit larger samples than those of most previous studies). Models with lower AIC were preferred over models with higher AIC, and in cases where the difference in AIC did not reach 2 (Burnham & Anderson, 1998), the model with fewer parameters was preferred.

p-values for each effect were calculated using the Satterthwaite approximation for degrees of freedom for individual predictor coefficients and meta-analysis (Luke, 2017). *p*-values were interpreted using the pre-registered α level of .05.

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 pre-registered protocol for labs applying sequential analysis established that they could stop data collection upon reaching the pre-registered criterion ($BF_{10} = 10$ or .10), or the maximal sample size. Each laboratory chose a fixed sample size and an incremental n for sequential analysis before their data collection. Two laboratories (HUN_001, TWN_001) stopped data collection at the pre-registered criterion, $BF_{10} = .10$. Fourteen laboratories did not finish the sequential analysis because (1) twelve laboratories were interrupted by the pandemic outbreak; (2) two laboratories (TUR_007E, TWN_002E) recruited English-speaking participants to comply with institutional policies. Lab-based records were reported on a public website as each laboratory completed data collection (details are available in Appendix C).

Results

Data Screening

As shown in Figure 2, participants' data were deleted listwise from the sentence-picture and picture-picture tasks if they did not perform with at least 70% accuracy. Next, the data were screened for outliers. Our pre-registered plan excluded 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. The minimum response time was set to 160 ms based on Hick's Law (Kvålseth, 2021; Proctor & Schneider, 2018). The maximum response latency was calculated as two times the mean absolute deviation plus the median calculated separately for each participant. Exclusions were performed at the trial level for these outlier response times.

To ensure equivalence between data collection methods, we evaluated the response times predicted by the fixed effects of the interaction between match (match vs. mismatch) and data collection source (in-person vs. online). We included random intercepts for participants, lab, language, and random slopes for source by lab and source by language. This analysis showed no difference between data sources: b = 2.41, SE = 2.77, t(73729.28) = 0.87, p = .385. Therefore, the following analyses did not separate in-person and online data. Table 3 provides a summary of the match advantage by language for the sentence-picture verification task.

Although we combined the two data sets in the final data analysis, it is worth considering that online participants' attention may be easily distracted given the lack of environmental control and experimenter overview. However, this secondary task revealed that online participants had a higher percent correct than in-person participants, t(3,214.86) = 35.77, p < .001, M_{online} = 85.46 (SD = 14.20) and $M_{in-person}$ = 67.71 (SD = 16.26).

Hypothesis 1: Meta-Analysis of the Orientation Effect

The planned meta-analysis examined the effect overall and within languages wherein at least two laboratories had collected data (Arabic, English, German, Norway, Simplified Chinese, Traditional Chinese, Slovakian, and Turkish). Figure 3 showed a significant positive orientation effect across German laboratories (b = 16.68, 95% CI [7.75, 25.62]) but did not reveal a significant overall effect (b = 2.05, 95% CI [-2.71, 6.82]). Also, a significant negative orientation effect was found in the Hungarian (b = -20.00, 95% CI [-29.60, -10.40]) and the Serbian laboratory (b = -17.25, 95% CI [-32.26, -2.24]), although in these languages only a single laboratory participated, so no language-specific meta-analysis was conducted.

Hypothesis 1: Mixed-Linear Modeling of the Orientation Effect

First, an intercept-only model of response times with no random intercepts was computed for comparison purposes AIC = 1008828.79. The model with the random intercept by participants was an improvement over this model, AIC = 971783.32. The addition of a target random intercept improved model fit over the participant intercept-only model, AIC = 969506.32. Data collection lab was then added to the model

Table 3

Descriptive Summary of Sentence-Picture Verification Task by Language

Language	Accuracy Percent	Mismatching	Matching	Match Advantage
Arabic	90.65	580.25 (167.53)	581.00 (200.89)	-0.75
Brazilian Portuguese	94.87	641.00 (136.40)	654.50 (146.78)	-13.50
English	95.04	576.75 (124.17)	578.75 (127.87)	-2.00
German	96.53	593.00 (106.75)	576.00 (107.12)	17.00
Greek	92.35	753.50 (225.36)	728.50 (230.91)	25.00
Hebrew	96.73	569.50 (98.59)	574.50 (110.45)	-5.00
Hindi	91.32	638.50 (207.19)	662.00 (228.32)	-23.50
Hungarian	96.47	623.00 (111.94)	643.00 (129.73)	-20.00
Norwegian	96.93	592.50 (126.39)	612.00 (136.03)	-19.50
Polish	96.11	601.00 (139.36)	586.00 (108.23)	15.00
Portuguese	95.01	616.50 (144.55)	607.00 (145.29)	9.50
Serbian	94.78	617.75 (158.64)	635.00 (168.28)	-17.25
Simplified Chinese	92.39	655.00 (170.50)	642.50 (158.64)	12.50
Slovak	96.45	610.50 (125.28)	607.25 (117.87)	3.25
Spanish	94.32	663.00 (147.52)	676.00 (154.19)	-13.00
Thai	93.92	652.50 (177.91)	637.75 (130.10)	14.75
Traditional Chinese	94.41	625.00 (139.36)	620.00 (123.06)	5.00
Turkish	95.38	654.50 (146.04)	637.00 (126.02)	17.50

Note. Average accuracy percentage, Median response times and median absolute deviations (in parentheses) per match condition (Mismatching, Matching); Match advantage (difference in response times).

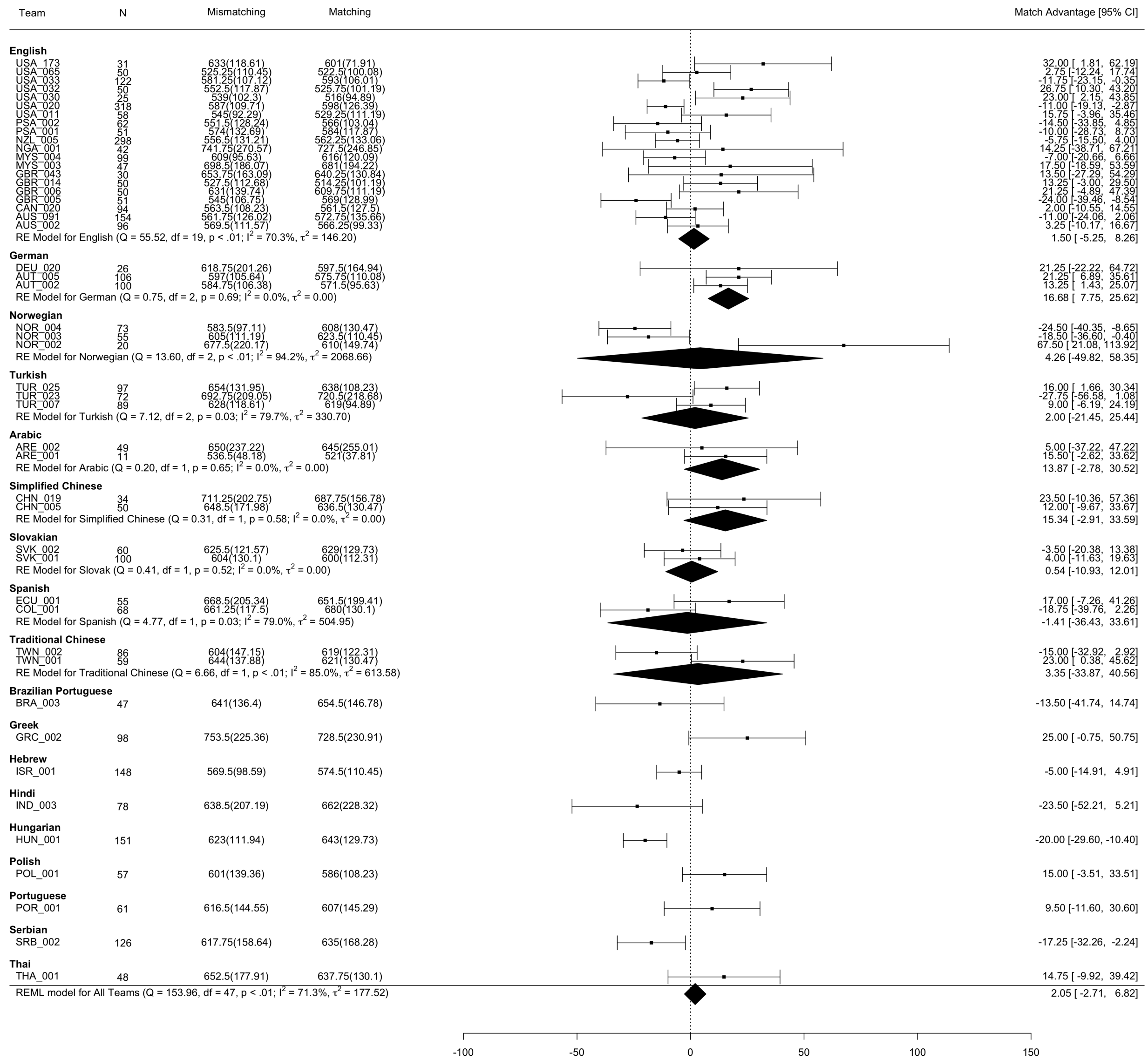

Figure 3

Meta-analysis on match advantage of object orientation for all languages. Diamonds indicate summary estimates, the midpoint of the diamond indicating the point estimate, and the left and right endpoints indicating the lower and upper bounds of the confidence interval of the estimated effect size. The lowermost diamond represents the estimate derived from the whole dataset.

as a random intercept, also showing model improvement, AIC = 969265.28, and the random intercept of language was added last, AIC = 969263.66 which did not show model improvement at least 2 points change. Last, the fixed effect of match advantage was added with approximately the same fit as the three random-intercept model, AIC = 969265.06. This model did not reveal a significant effect of match advantage: b = -0.17, SE = 1.20, t(69830.14) = -0.14, p = .887.

We conducted an exploratory mixed-effects model on German data as this was the only language indicating a significant match advantage in the meta-analysis. An intercept-only model with random effects for participants, target, and lab was used as a comparison, AIC = 55828.57. The addition of the fixed effect of match showed a small improvement over this random-intercept model, AIC = 55824.52. Whereas the AIC values indicated a significant change, the p-value did not reveal a significant effect of match advantage: b = 4.84, SE = 4.12, t(4085.71) = 1.17, p = .241. All the details of the above fixed effects and random intercepts are summarized in Appendix D.

Hypothesis 2: Mental Rotation Scores

Using the same steps as described for the sentence-picture verification mixed model, we first started with an intercept-only model with no random effects for comparison, AIC = 1029362.78. The addition of random intercepts by subject, AIC = 979873.47, by item, AIC = 977037.64, by lab, AIC = 976721.45, and by language, AIC = 976717.46, all subsequently improved model fit. Next, the match effect for object orientation was entered as the fixed effect for mental rotation score, AIC = 973054.93, which showed improvement over the random intercepts model. This model showed a significant effect of object orientation, b = 32.30, SE = 0.53, t(79585.24) = 61.23, p < .001, such that identical orientations were processed faster than rotated orientations. The point estimates of the orientation effect varied between 23.79–40.24, revealing a range of 14 ms across languages. The coefficients of all mixed-effects models are reported in Appendix E, along with all effects presented by language.

Hypothesis 2: Prediction of Match Advantage

The last analysis included a mixed effects regression model using the interaction of language and mental rotation scores to predict match advantage in the sentence-picture task. First, an intercept-only model was calculated for comparison, AIC = 42678.66, which was improved slightly by adding a random intercept by data collection lab, AIC = 42677.80. The addition of the fixed effects interaction of language and mental rotation score improved the overall model, AIC = 42633.44. English was used as the comparison group for all language comparisons. Neither the mental rotation score nor the interaction of mental rotation score and language were significant, and these results are detailed in Appendix E.

Discussion

This study aimed to test a global object orientation effect and to estimate the magnitude of object orientation effect in each particular language. The findings of our study did not support the existence of the object orientation effect as an outcome of general cognitive function. Furthermore, our data failed to replicate the effects in English and Chinese, languages in which the effect has been reported previously (Chen et al., 2020; Stanfield & Zwaan, 2001; Zwaan & Pecher, 2012). The only language in which we found an indication of the orientation effect in the predicted direction was German, but this effect was evident only in the meta-analysis and not in the mixed-effects model approach. Although tangential to our topic, an effect of mental rotation was observed, such that identical orientations were processed faster than rotated orientations. However, the mental rotation score did not predict the object orientation effects nor interact with language. Overall, the failure to replicate the previously reported object orientation effects casts doubt on the existence of the effect as a language-general phenomenon (Kaschak & Madden, 2021). Below, we summarize the lessons and limitations of the methodology and analysis, and discuss theoretical issues related to the orientation effect as an effective probe to investigate the mental simulation process.

Methodological Considerations

By examining the failed replications of the object orientation effect in the English-language labs (see Figure 3), researchers can further identify the possible factors that may have contributed to the discrepancies between the results of this project and the original studies. Although our project had a larger sample of English-speaking participants compared to the original studies (i.e., Stanfield & Zwaan, 2001; Zwaan & Pecher, 2012), our English-speaking participants came from multiple countries where the participants' lexical knowledge is not completely consistent with American English. Although we prepared an alternative version of the stimuli for British English, these two versions of English stimuli did not cover all English language backgrounds, such as participants from Malaysia and Africa. Despite the overall non-significant effect in all English-language data, the meta-analysis indicated three significant positive team-based effects (USA_173, USA_030 and USA_032, see Figure 3) but also three significant negative effects (USA_33, USA_20, and GBR_005, see Figure 3). Future cross-linguistic studies should attempt to balance sample sizes across languages to allow reliable cross-linguistic comparisons.

Regarding the failed replication of Chinese orientation effects, the past study used simpler sentence content compared to this project. Chen et al. (2020) used the probe sentences in which the target objects were the subject of sentences (e.g., *The* ***nail*** *was hammered into the wall*; bold added to mark the subject noun). The Chinese probe sentences in this project were translated from the English sentences used in Stanfield and Zwaan (2001), in which the target objects are the object of sentences (e.g., *The carpenter hammered the* ***nail*** *into the wall*; bold added to mark object noun). It is possible that the object orientation effect may be present or stronger when the target objects are the subject of the sentence, rather than the direct object, and future studies could explore this distinction.

Lastly, past studies that employed a secondary task among the experimental trials (Chen et al., 2020; Kaschak & Madden, 2021; Stanfield & Zwaan, 2001) showed a positive object orientation effect. In our study, the memory check did not increase the likelihood to detect the mental simulation effects. In addition, we did not find that mental imagery predicted match advantage, which implies that this strategy to ensure linguistic processing had limited influence in our study.

Analysis Considerations

The orientation effects were analyzed using a meta-analytic approach and mixed-effects models. Neither approach revealed an overall effect of object orientation. In the language-by-language analysis, a significant orientation effect was found in the German language data in the meta-analysis. The mixed model analysis did not confirm this result because the effect in the German data was not significant according to our pre-registered test criteria. There is considerable debate in the statistical community regarding the precision of the p values computed for linear mixed models (Bolker, 2015). One alternative, less-conservative approach to testing the significance of a fixed effect predictor is assessing the difference in the AIC model fit index between a model that contains a fixed effect predictor and one that does not (Matuschek et al., 2017). Using this approach in an exploratory analysis, we found that the effect of orientation in the German language data was not negligible, rendering this result compatible with the result obtained for German in the meta-analysis. However, considered in the general context including all the other results, the present exploratory result for German could stem from measurement error (Loken & Gelman, 2017) or from family-wise error (Armstrong, 2014).

When a topic area yields inconsistent or small effects, some researchers have questioned the utility of further research (Brysbaert, 2020; Sala & Gobet, 2017). However, research on embodied cognition should continue with the aim of determining the factors behind the variability of the effects. One of these factors could be the nature of the variables used - for instance, categorical versus continuous. The object orientation design is a factorial, congruency paradigm, based on congruent (matching) and incongruent (mismatching) conditions. Another paradigm of similar characteristics, namely the action sentence compatibility effect, similarly failed to replicate in a large-scale study (Morey et al., 2022). Whereas factorial paradigms require the use of categorical variables, other studies have operationalized sensorimotor information using continuous variables, and observed significant effects (Bernabeu, 2022; Lynott et al., 2020; Petilli et al., 2021). Since continuous variables contain more information, they may afford more statistical power (J. Cohen, 1983). Furthermore, in addition to categorical versus continuous predictors, sensorimotor effects are likely to be moderated by factors influencing participants' attention during experiments (Barsalou, 2019; Noah et al., 2018). Last, due to publication bias, the true size of sensorimotor effects is likely to be smaller than that observed in small-sample studies (Vasishth & Gelman, 2021). Indeed, studying these effects reliably may require samples exceeding 1,000 participants (Bernabeu, 2022). In summary, addressing the above issues may permit the analytic sensitivity needed to observe the presence and causes of object orientation effects.

Theoretical Considerations

Scholars interested in mental simulation have investigated whether the human mind processes linguistic content as abstract symbols or as grounded mental representations (Barsalou, 1999, 2008; Zwaan, 2014b). Some of the tasks used to test these theories-such as the sentence-verification task-rely on priming-based logic, whereby a designed sentence generates representations along some dimension (such as orientation) that facilitates or interferes with the processing of the subsequent stimulus (Roelke et al., 2018). Furthermore, embodied cognition theories suggest that the reading of the sentence will activate perceptual experience, thus facilitating a matching object picture and causing interference for a mismatching picture (Kaschak & Madden, 2021; McNamara, 2005). To scrutinize these effects, future studies could augment the sentence-picture verification task to compare the degree of priming based on object orientation with the priming based on other semantic information. The present study constitutes the first large-scale, cross-linguistic approach to the object orientation effect. Cross-linguistic studies are rare in the present topic, and generally in the topic of conceptual/semantic processing. In future studies, the basis for cross-linguistic comparisons in conceptual processing should be expanded, for instance, by studying the lexicosemantic features of the stimuli used, how those differ across languages, and how those differences may influence psycholinguistic processing. For the development of this founding work, the field of linguistic relativity may be useful as a model (e.g., Athanasopoulos, 2023).

In addition, further research should compare the size of mental simulation effects with the size of effects that are associated with the symbolic account of conceptual processing. The symbolic account posits that conceptual processing (i.e., the comprehension of the meaning of words) depends on the abstract symbols (e.g., propositions and production rules). So far, some of these comparisons have supported both accounts. However, in some studies, the effects of the symbolic account have been larger than those of the embodied account (Bernabeu, 2022; Louwerse et al., 2015), whereas the reverse has been true in other studies (Fernandino et al., 2022; Tong et al., 2022).

Limitations

This study reflects the challenges to assess the mental simulation of object orientation across languages, especially when dealing with effects that require large sample sizes (see Loken & Gelman, 2017; Vadillo et al., 2016). Our data collection deviated from the pre-registered plan because of the COVID-19 pandemic. Due to the lack of participant monitoring online, and an inspection of the data, we post-hoc used filtering on outliers in terms of participants' response times for both too fast (< 160 ms) and too slow responses (2 MAD beyond the median for each participant individually). After these exclusions, a mixed-effects model confirmed no difference of response times between in-person and online data. Future studies could evaluate how the task environments alter the magnitude of the orientation effect.

Conclusion

Based on the results of this project, we did not find evidence for a general object orientation effect across languages. Our findings on the orientation effects question the theoretical importance of mental simulation in linguistic processing, but they also provide directions for new avenues of investigation.

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

Athanasopoulos, P. (2023). *Linguistic relativity* (J. Culpeper, Ed.; pp. 469–477). Routledge.

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

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

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

Barsalou, L. W. (2020). Challenges and opportunities for grounding cognition. *Journal of Cognition*, *3*(1), 31. https://doi.org/10.5334/joc.116

Bernabeu, P. (2022). *Language and sensorimotor simulation in conceptual processing: Multilevel analysis and statistical power*. https://doi.org/10.17635/LANCASTER/THESIS/1795

Bolker, B. M. (2015). *Linear and generalized linear mixed models* (G. A. Fox, S. Negrete-Yankelevich, & V. J. Sosa, Eds.; pp. 309–333). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199672547.003.0014

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

Brysbaert, M. (2020). Power considerations in bilingualism research: Time to step up our game. *Bilingualism: Language and Cognition*, *24*(5), 813–818.

https://doi.org/10.1017/s1366728920000437

Burnham, K. P., & Anderson, D. R. (1998). *Practical use of the information-theoretic approach* (pp. 75–117). Springer New York. https://doi.org/10.1007/978-1-4757-2917-7_3

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

Cohen, J. (1983). The Cost of Dichotomization. *Applied Psychological Measurement*, *7*(3), 249–253. https://doi.org/10.1177/014662168300700301

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

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

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

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

Fernandino, L., Tong, J.-Q., Conant, L. L., Humphries, C. J., & Binder, J. R. (2022). Decoding the information structure underlying the neural representation of concepts. *Proceedings of the National Academy of Sciences*, *119*(6), e2108091119. https://doi.org/10.1073/pnas.2108091119

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

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, *141*(1), 2–18. https://doi.org/10.1037/a0024338

Kaschak, M. P., & Madden, J. (2021). Embodiment in the lab: Theory, measurement, and reproducibility. In M. D. Robinson & L. E. Thomas (Eds.), *Handbook of Embodied Psychology* (pp. 619–635). Springer International Publishing. https://doi.org/10.1007/978-3-030-78471-3_27

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

Kvålseth, T. O. (2021). Hick's law equivalent for reaction time to individual stimuli. *British Journal of Mathematical and Statistical Psychology*, *74*(S1), 275–293. https://doi.org/10.1111/bmsp.12232

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

Louwerse, M. M., Hutchinson, S., Tillman, R., & Recchia, G. (2015). Effect size matters: the role of language statistics and perceptual simulation in conceptual processing. *Language, Cognition and Neuroscience*, *30*(4), 430–447. https://doi.org/10.1080/23273798.2014.981552

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

Lynott, D., Connell, L., Brysbaert, M., Brand, J., & Carney, J. (2020). The Lancaster Sensorimotor Norms: multidimensional measures of perceptual and action strength for 40,000 English words. *Behavior Research Methods*, *52*(3), 1271–1291. https://doi.org/10.3758/s13428-019-01316-z

Mathôt, S., & March, J. (2022). Conducting linguistic experiments online with OpenSesame and OSWeb. *Language Learning*, *72*(4), 1017–1048. https://doi.org/10.1111/lang.12509

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

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

McNamara, T. P. (2005). *Semantic priming: Perspectives from memory and word recognition*. Psychology Press.

Morey, R. D., Kaschak, M. P., Díez-Álamo, A. M., Glenberg, A. M., Zwaan, R. A., Lakens, D., Ibáñez, A., García, A., Gianelli, C., Jones, J. L., Madden, J., Alifano, F., Bergen, B., Bloxsom, N. G., Bub, D. N., Cai, Z. G., Chartier, C. R., Chatterjee, A., Conwell, E., … Ziv-Crispel, N. (2022). A pre-registered, multi-lab non-replication of the action-sentence compatibility effect (ACE). *Psychonomic Bulletin & Review*, *29*(2), 613–626. https://doi.org/10.3758/s13423-021-01927-8

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

Noah, T., Schul, Y., & Mayo, R. (2018). When both the original study and its failed replication are correct: Feeling observed eliminates the facial-feedback effect. *Journal of Personality and Social Psychology*, *114*(5), 657–664. https://doi.org/10.1037/pspa0000121

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

Petilli, M. A., Günther, F., Vergallito, A., Ciapparelli, M., & Marelli, M. (2021). Data-driven computational models reveal perceptual simulation in word processing. *Journal of Memory and Language*, *117*, 104194. https://doi.org/10.1016/j.jml.2020.104194

Phills, C., Kekecs, Z., & Chartier, C. (2022). *Pre-Registration.*

https://doi.org/10.17605/OSF.IO/4HMGS

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

Proctor, R. W., & Schneider, D. W. (2018). Hick's law for choice reaction time: A review. *Quarterly Journal of Experimental Psychology*, *71*(6), 1281–1299. https://doi.org/10.1080/17470218.2017.1322622

Roelke, A., Franke, N., Biemann, C., Radach, R., Jacobs, A. M., & Hofmann, M. J. (2018). A novel co-occurrence-based approach to predict pure associative and semantic priming. *Psychonomic Bulletin & Review*, *25*(4), 1488–1493. https://doi.org/10.3758/s13423-018-1453-6

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

Sala, G., & Gobet, F. (2017). Does Far Transfer Exist? Negative Evidence From Chess, Music, and Working Memory Training. *Current Directions in Psychological Science*, *26*(6), 515–520. https://doi.org/10.1177/0963721417712760

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

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.

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

Tong, J., Binder, J. R., Humphries, C., Mazurchuk, S., Conant, L. L., & Fernandino, L. (2022). A distributed network for multimodal experiential representation of concepts. *The Journal of Neuroscience*, *42*(37), 7121–7130. https://doi.org/10.1523/JNEUROSCI.1243-21.2022

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

Vasishth, S., & Gelman, A. (2021). How to embrace variation and accept uncertainty in linguistic and psycholinguistic data analysis. *Linguistics*, *59*(5), 1311–1342. https://doi.org/10.1515/ling-2019-0051

Vazire, S. (2018). Implications of the credibility revolution for productivity, creativity, and progress. *Perspectives on Psychological Science*, *13*(4), 411–417. https://doi.org/10.1177/1745691617751884

Verkerk, A. (2014). The correlation between motion event encoding and path verb lexicon size in the indo-european language family. *Folia Linguistica*, *35*(1). https://doi.org/10.1515/flih.2014.009

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, *9*(4), 625–636. https://doi.org/10.3758/BF03196322

Zwaan, R. A. (2014a). Embodiment and language comprehension: Reframing the discussion. *Trends in Cognitive Sciences*, *18*(5), 229–234. https://doi.org/10.1016/j.tics.2014.02.008

Zwaan, R. A. (2014b). 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., Diane Pecher, Paolacci, G., Bouwmeester, S., Verkoeijen, P., Dijkstra, K., & Zeelenberg, R. (2017). Participant Nonnaiveté and the reproducibility of cognitive psychology. *Psychonomic Bulletin & Review*, 1–5. https://doi.org/10.3758/s13423-017-1348-y

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

Appendix A

Sensitivity Analyses

The R codes for the sensitivity analysis on the trial level were written by Erin M. Buchanan.

Load data and run models

The data for the sensitivity analysis shared the same exclusion criterion for the pre-registered mixed-effects models. The first step is to determine if there is a minimum number of trials required for stable results.

View the Results

b values

These values represent the *b* values found for each run of 3 up to 12 trials.

-0.17, -0.17, -0.17, -0.17, -0.18, -0.12, 0.49, -0.14, 0.67, and 3.11

p values

These values represent the *p* values found for each run of 7 up to 12 trials.

.887, .887, .887, .890, .880, .918, .687, .913, .647, .150

As we can see, the effect is generally negative until participants were required to have 7-12 correct trials. When participants accurately answer all 12 trials the effect is approximately 3 ms. Examination of the *p*-values indicates that no coefficients would have been considered significant.

Calculate the Sensitivity

Given we used all data points, the smallest detectable effect with our standard error and degrees of freedom would have been:

```
## [1] 2.356441
```

Appendix B

Data Collection Logs

The log website was initiated since the data collection began. The public logs were updated when a laboratory updated their data for the sequential analysis. The link to access the public site is: https://scgeeker.github.io/PSA002_log_site/index.html

If you want to check the sequential analysis result of a laboratory, at first you have to identify the ID and language of this laboratory from "Overview" page. Next you will navigate to the language page under the banner "Tracking Logs". For example, you want to see the result of "GBR_005". Navigate "Tracking Logs -> English". Search the figure by ID "GBR_005".

The source files of the public logs are available in the github repository: https://github.com/SCgeeker/PSA002_log_site

All the raw data and log files are compressed in the project OSF repository: https://osf.io/e428p/

The R code to conduct the Bayesian sequential analysis is available at "data_seq_analysis.R". This file can be found at: https://github.com/SCgeeker/PSA002_log_site

Note 1 USA_067, BRA_004 and POL_004 were unavailable because the teams withdrew.

Note 2 Some mistakes happened between the collaborators' communications and required advanced data wrangling. For example, some AUS_091 participants were assigned to NZL_005. The Rmd file in NZL_005 folder were used to identify the AUS_091 participants' data then move them to AUS_091 folder.

Datasets

Complete data can be found online with this manuscript or on each collaborators OSF page. Please see the `Lab_Info.csv` on https://osf.io/e428p/.

Flunecy test for the online study

At the beginning of the online study, participants will hear the verbal instruction narrated by a native speaker. The original English transcript is as below:

"In this session you will complete two tasks. The first task is called the sentence picture verification task. In this task, you will read a sentence. You will then see a picture. Your job is to verify whether the picture represents an object that was described in the sentence or not. The second task is the picture verification task. In this task you will see two pictures on the screen at the same time and determine whether they are the same or different. Once you have completed both tasks, you will receive a completion code that you can use to verify your participation in the study."

The fluency test are three multiple choice questions. The question text and the correct answers are as below:

- How many tasks will you run in this session?

  A: 1 *B: 2 C: 3

- When will you get the completion code?

  A: Before the second task B: After the first task *C: After the second task

- What will you do in the sentence-picture verification task?

  A: Confirm two pictures for their objects

  *B: Read a sentence and verify a picture C: Judge sentences for their accuracy

Distributions of scripts

The instructions and experimental scripts are available at the public OSF folder (https://osf.io/e428p/ “Materials/js” folder in Files). To upload to a jatos server, a script had to be converted to the compatible package. Researchers could do this conversion by “OSWEB” package in OpenSesame. We rent an remote server for the distributions during the data collection period. Any researcher would distribute the scripts on a free jatos server such as MindProbe (https://www.mindprobe.eu/).

Appendix C

Demographic Characteristics by Language

Table C1

Demographic and Sample Size Characteristics by Language Part 1

Language	SP_{Trials}	PP_{Trials}	SP_N	PP_N	$Demo_N$	$Female_N$	$Male_N$	M_{Age}	SD_{Age}
Arabic	1248	1248	52	52	53	0	0	38.00	NaN
Arabic	1296	1296	54	54	54	42	12	26.51	18.59
Brazilian Portuguese	1200	1200	50	50	50	36	13	30.80	8.73
English	2376	2376	99	99	103	46	37	20.14	3.32
English	3840	3840	160	160	160	127	25	26.03	11.55
English	2352	2376	98	99	104	54	40	20.26	3.66
English	1272	1272	53	53	76	57	13	19.96	3.90
English	1200	1200	50	50	51	37	13	20.14	2.46
English	1200	1200	50	50	58	46	11	18.74	1.62
English	720	720	30	30	32	15	11	25.70	9.40
English	1200	1224	50	51	52	38	11	22.56	3.90
English	2400	2400	100	100	109	65	30	20.73	2.00
English	1248	1248	52	52	52	24	22	23.94	11.29
English	7680	7680	320	320	320	244	56	23.21	5.43
English	1248	1272	52	53	71	50	12	18.89	0.95

Note. SP = Sentence Picture Verification, PP = Picture Picture Verification. Sample sizes for demographics may be higher than the sample size for the this study, as participants could have only completed the bundled experiment. Additionally, not all entries could be unambigously matched by lab ID, and therefore, demographic sample sizes could also be less than data collected. Each row represents a single lab.

Table C2

Demographic and Sample Size Characteristics by Lab Part 2

Language	SP_{Trials}	PP_{Trials}	SP_N	PP_N	$Demo_N$	$Female_N$	$Male_N$	M_{Age}	SD_{Age}
English	1536	1536	64	64	102	79	11	19.82	2.42
English	264	264	11	11	12	9	2	20.36	1.91
English	288	288	12	12	12	6	5	21.17	1.19
English	1512	1512	63	63	63	30	23	22.34	11.55
English	7980	8064	333	336	403	258	76	19.63	2.12
English	648	648	27	27	31	20	3	36.00	0.96
English	1209	1224	51	51	51	30	21	19.29	1.51
English	3000	3024	125	126	129	90	25	20.06	1.36
English	1200	1200	50	50	61	35	15	18.86	1.63
English	816	744	34	31	3	0	3	19.67	0.58
German	2400	2400	100	100	114	0	1	20.94	2.56
German	2592	2592	108	108	108	80	22	22.18	4.26
German	624	624	26	26	26	18	3	23.88	3.39
Greek	2376	2376	99	99	109	0	0	33.86	11.30
Hebrew	3576	3571	149	149	181	0	0	24.25	9.29
Hindi	1896	1896	79	79	86	57	27	21.66	3.46
Magyar	3610	3816	151	159	168	3	1	21.50	2.82
Norwegian	504	504	21	21	21	12	8	30.10	8.58
Norwegian	1320	1320	55	55	53	1	1	23.55	6.25
Norwegian	1752	1752	73	73	80	0	0	22.00	4.38

Note. SP = Sentence Picture Verification, PP = Picture Picture Verification. Sample sizes for demographics may be higher than the sample size for the this study, as participants could have only completed the bundled experiment. Additionally, not all entries could be unambigously matched by lab ID, and therefore, demographic sample sizes could also be less than data collected.

Table C3

Demographic and Sample Size Characteristics by Lab Part 3

Language	SP_{Trials}	PP_{Trials}	SP_N	PP_N	$Demo_N$	$Female_N$	$Male_N$	M_{Age}	SD_{Age}
Polish	1368	1368	57	57	146	0	0	23.25	7.96
Portuguese	1488	1464	62	61	55	26	23	30.74	9.09
Serbian	3120	3120	130	130	130	108	21	21.38	4.50
Simplified Chinese	1200	1200	50	50	57	0	0	18.66	3.92
Simplified Chinese	840	816	35	34	39	0	1	25.17	5.44
Slovak	2419	2400	101	100	103	1	0	21.59	2.51
Slovak	1462	1199	61	50	222	0	0	21.96	2.14
Spanish	1680	1656	70	69	70	0	0	21.36	3.36
Spanish	1440	1440	60	60	76	0	0	22.10	4.30
Thai	1200	1152	50	48	50	29	9	21.54	3.81
Traditional Chinese	1440	1440	60	60	70	45	14	20.73	1.21
Traditional Chinese	2160	2160	90	90	116	24	32	21.04	3.66
Turkish	2184	2184	91	91	93	0	0	20.92	2.93
Turkish	1896	1896	79	79	80	36	14	21.58	8.64
Turkish	2376	2352	99	98	101	0	0	21.63	2.19

Note. SP = Sentence Picture Verification, PP = Picture Picture Verification. Sample sizes for demographics may be higher than the sample size for the this study, as participants could have only completed the bundled experiment. Additionally, not all entries could be unambigously matched by lab ID, and therefore, demographic sample sizes could also be less than data collected.

Appendix D

Model Estimates for Mental Simulation

All model estimates are given below for the planned mixed linear model to estimate the matching effect for object orientation in the sentence picture verification task.

Note. Fixed indicates fixed parameters in multilevel models, while "ran_pars" indicates the random intercepts included in the model.

Table D1

Intercept Only Object Orientation Results

Term	Estimate (b)	SE	t	p
(Intercept)	654.71	0.84	775.11	$< .001$

Table D2

Subject-Random Intercept Object Orientation Results

Effect	Group	Term	Estimate (b)	SE	t	df	p
fixed	NA	(Intercept)	654.26	2.69	243.34	3,787.12	$< .001$
ran_pars	Subject	sd___(Intercept)	161.40	NA	NA	NA	
ran_pars	Residual	sd___Observation	165.05	NA	NA	NA	

Table D3

Subject and Item-Random Intercept Object Orientation Results

Effect	Group	Term	Estimate (b)	SE	t	df	p
fixed	NA	(Intercept)	655.47	5.16	126.97	84.63	$< .001$
ran_pars	Subject	sd___(Intercept)	161.37	NA	NA	NA	
ran_pars	Target	sd___(Intercept)	30.54	NA	NA	NA	
ran_pars	Residual	sd___Observation	162.17	NA	NA	NA	

Table D4

Subject, Item, and Lab-Random Intercept Object Orientation Results

Effect	Group	Term	Estimate (b)	SE	t	df	p
fixed	NA	(Intercept)	659.63	9.67	68.24	65.54	$< .001$
ran_pars	Subject	sd___(Intercept)	153.76	NA	NA	NA	
ran_pars	Target	sd___(Intercept)	30.56	NA	NA	NA	
ran_pars	PSA_ID	sd___(Intercept)	55.76	NA	NA	NA	
ran_pars	Residual	sd___Observation	162.17	NA	NA	NA	

Table D5

Subject, Item, Lab, and Language-Random Intercept Object Orientation Results

Effect	Group	Term	Estimate (b)	SE	t	df	p
fixed	NA	(Intercept)	672.75	11.88	56.61	23.94	$< .001$
ran_pars	Subject	sd___(Intercept)	153.78	NA	NA	NA	
ran_pars	Target	sd___(Intercept)	30.56	NA	NA	NA	
ran_pars	PSA_ID	sd___(Intercept)	48.60	NA	NA	NA	
ran_pars	Language	sd___(Intercept)	25.52	NA	NA	NA	
ran_pars	Residual	sd___Observation	162.17	NA	NA	NA	

Table D6

Fixed Effects Object Orientation Results

Effect	Group	Term	Estimate (b)	SE	t	df	p
fixed	NA	(Intercept)	659.71	9.68	68.12	66.04	$< .001$
fixed	NA	MatchN	-0.17	1.20	-0.14	69,830.14	.887
ran_pars	Subject	sd___(Intercept)	153.76	NA	NA	NA	
ran_pars	Target	sd___(Intercept)	30.56	NA	NA	NA	
ran_pars	PSA_ID	sd___(Intercept)	55.76	NA	NA	NA	
ran_pars	Residual	sd___Observation	162.17	NA	NA	NA	

Table D7

Random Effects German Object Orientation Results

Effect	Group	Term	Estimate (b)	SE	t	df	p
fixed	NA	(Intercept)	631.96	19.65	32.15	1.74	.002
ran_pars	Subject	sd___(Intercept)	129.89	NA	NA	NA	
ran_pars	Target	sd___(Intercept)	33.04	NA	NA	NA	
ran_pars	PSA_ID	sd___(Intercept)	28.11	NA	NA	NA	
ran_pars	Residual	sd___Observation	134.89	NA	NA	NA	

Table D8

Fixed Effects German Object Orientation Results

Effect	Group	Term	Estimate (b)	SE	t	df	p
fixed	NA	(Intercept)	629.52	19.74	31.90	1.77	.002
fixed	NA	MatchN	4.84	4.12	1.17	4,085.71	.241
ran_pars	Subject	sd__(Intercept)	129.90	NA	NA	NA	
ran_pars	Target	sd__(Intercept)	33.06	NA	NA	NA	
ran_pars	PSA_ID	sd__(Intercept)	28.05	NA	NA	NA	
ran_pars	Residual	sd__Observation	134.88	NA	NA	NA	

Appendix E

Model Estimates for Mental Rotation

All model estimates are given below for the mixed linear model for the prediction of mental rotation scores by orientation, and the effects of predicting mental simulation effects (object orientation) with the mental rotation scores.

Note. Fixed indicates fixed parameters in multilevel models, while "ran_pars" indicates the random intercepts included in the model.

Table E1

Intercept Only Mental Rotation Results

Term	Estimate (b)	SE	t	p
(Intercept)	589.10	0.40	1,485.40	$< .001$

Table E2

Subject-Random Intercept Mental Rotation Results

Effect	Group	Term	Estimate (b)	SE	t	df	p
fixed	NA	(Intercept)	588.28	1.35	436.44	3,957.55	$< .001$
ran_pars	Subject	sd___(Intercept)	83.04	NA	NA	NA	
ran_pars	Residual	sd___Observation	79.04	NA	NA	NA	

Table E3

Subject and Item-Random Intercept Mental Rotation Results

Effect	Group	Term	Estimate (b)	SE	t	df	p
fixed	NA	(Intercept)	589.21	2.70	217.83	79.98	$< .001$
ran_pars	Subject	sd___(Intercept)	82.94	NA	NA	NA	
ran_pars	Picture1	sd___(Intercept)	16.26	NA	NA	NA	
ran_pars	Residual	sd___Observation	77.56	NA	NA	NA	

Table E4

Subject, Item, and Lab-Random Intercept Mental Rotation Results

Effect	Group	Term	Estimate (b)	SE	t	df	p
fixed	NA	(Intercept)	590.18	5.16	114.34	69.45	$< .001$
ran_pars	Subject	sd___(Intercept)	78.36	NA	NA	NA	
ran_pars	Picture1	sd___(Intercept)	16.32	NA	NA	NA	
ran_pars	PSA_ID	sd___(Intercept)	30.12	NA	NA	NA	
ran_pars	Residual	sd___Observation	77.56	NA	NA	NA	

Table E5

Subject, Item, Lab, and Language-Random Intercept Mental Rotation Results

Effect	Group	Term	Estimate (b)	SE	t	df	p
fixed	NA	(Intercept)	596.71	6.98	85.47	22.17	$< .001$
ran_pars	Subject	sd___(Intercept)	78.37	NA	NA	NA	
ran_pars	Picture1	sd___(Intercept)	16.32	NA	NA	NA	
ran_pars	PSA_ID	sd___(Intercept)	24.00	NA	NA	NA	
ran_pars	Language	sd___(Intercept)	19.29	NA	NA	NA	
ran_pars	Residual	sd___Observation	77.56	NA	NA	NA	

Table E6

Fixed Effects Mental Rotation Results

Effect	Group	Term	Estimate (b)	SE	t	df	p
fixed	NA	(Intercept)	581.52	7.02	82.82	22.57	$< .001$
fixed	NA	IdenticalN	32.30	0.53	61.23	79,585.24	$< .001$
ran_pars	Subject	sd__(Intercept)	78.24	NA	NA	NA	
ran_pars	Picture1	sd__(Intercept)	16.89	NA	NA	NA	
ran_pars	PSA_ID	sd__(Intercept)	24.01	NA	NA	NA	
ran_pars	Language	sd__(Intercept)	19.33	NA	NA	NA	
ran_pars	Residual	sd__Observation	75.80	NA	NA	NA	

Table E7

Language Specific Mental Rotation Results

Language	Coefficient (b)	SE
Arabic	28.27	3.36
English	33.02	0.77
German	31.38	1.91
Brazilian Portuguese	23.79	4.85
Simplified Chinese	32.40	3.64
Spanish	40.24	3.75
Greek	30.59	3.67
Hungarian	25.43	2.57
Hindi	35.83	3.86
Hebrew	29.02	2.43
Norwegian	28.12	2.59
Polish	38.74	3.51
Portuguese	34.67	4.05
Serbian	25.93	2.75
Slovak	33.34	2.61
Thai	34.99	4.13
Turkish	37.46	2.29
Traditional Chinese	30.31	2.45

Table E8

Intercept Only Predicting Mental Simulation Results

Term	Estimate (b)	SE	t	p
(Intercept)	-0.74	1.67	-0.44	.661

Table E9

Lab-Random Intercept Predicting Mental Simulation Results

Effect	Group	Term	Estimate (b)	SE	t	df	p
fixed	NA	(Intercept)	-0.74	1.67	-0.44	3,543.00	.661
ran_pars	PSA_ID	sd___(Intercept)	0.00	NA	NA	NA	
ran_pars	Residual	sd___Observation	99.67	NA	NA	NA	

Table E10

Fixed Effects Interaction Language and Rotation Predicting Mental Simulation Results Part 1

Effect	Group	Term	Estimate (b)	SE	t	df	p
fixed	NA	(Intercept)	-3.43	3.07	-1.12	3,510.00	.264
fixed	NA	LanguageArabic	16.27	16.34	1.00	3,510.00	.319
fixed	NA	LanguageBrazilian Portuguese	-17.00	16.63	-1.02	3,510.00	.307
fixed	NA	LanguageGerman	20.65	9.17	2.25	3,510.00	.024
fixed	NA	LanguageGreek	7.52	12.58	0.60	3,510.00	.550
fixed	NA	LanguageHindi	1.24	15.39	0.08	3,510.00	.936
fixed	NA	LanguageHungarian	-10.30	10.01	-1.03	3,510.00	.304
fixed	NA	LanguageNorwegian	19.66	10.31	1.91	3,510.00	.057
fixed	NA	LanguagePolish	-5.67	17.87	-0.32	3,510.00	.751
fixed	NA	LanguagePortuguese	-15.90	17.21	-0.92	3,510.00	.356
fixed	NA	LanguageSerbian	-0.48	11.20	-0.04	3,510.00	.966
fixed	NA	LanguageSimplified Chinese	22.70	14.45	1.57	3,510.00	.116
fixed	NA	LanguageSlovak	-0.34	11.58	-0.03	3,510.00	.976
fixed	NA	LanguageSpanish	5.39	11.51	0.47	3,510.00	.640
fixed	NA	LanguageThai	27.53	21.22	1.30	3,510.00	.195
fixed	NA	LanguageTraditional Chinese	-8.24	11.20	-0.74	3,510.00	.462
fixed	NA	LanguageTurkish	10.66	8.57	1.24	3,510.00	.214
fixed	NA	Imagery	0.04	0.05	0.79	3,510.00	.432

Table E11

Fixed Effects Interaction Language and Rotation Predicting Mental Simulation Results Part 2

Effect	Group	Term	Estimate (b)	SE	t	df	p
fixed	NA	LanguageArabic:Imagery	-0.37	0.24	-1.55	3,510.00	.122
fixed	NA	LanguageBrazilian Portuguese:Imagery	0.18	0.29	0.62	3,510.00	.536
fixed	NA	LanguageGerman:Imagery	-0.37	0.19	-1.96	3,510.00	.050
fixed	NA	LanguageGreek:Imagery	-0.07	0.20	-0.36	3,510.00	.718
fixed	NA	LanguageHindi:Imagery	-0.36	0.27	-1.34	3,510.00	.181
fixed	NA	LanguageHungarian:Imagery	0.10	0.22	0.45	3,510.00	.653
fixed	NA	LanguageNorwegian:Imagery	-0.07	0.19	-0.35	3,510.00	.726
fixed	NA	LanguagePolish:Imagery	0.29	0.31	0.91	3,510.00	.363
fixed	NA	LanguagePortuguese:Imagery	-0.05	0.32	-0.15	3,510.00	.884
fixed	NA	LanguageSerbian:Imagery	-0.12	0.22	-0.56	3,510.00	.576
fixed	NA	LanguageSimplified Chinese:Imagery	-0.32	0.24	-1.32	3,510.00	.187
fixed	NA	LanguageSlovak:Imagery	0.12	0.21	0.57	3,510.00	.568
fixed	NA	LanguageSpanish:Imagery	0.05	0.17	0.28	3,510.00	.781
fixed	NA	LanguageThai:Imagery	-0.50	0.38	-1.32	3,510.00	.186
fixed	NA	LanguageTraditional Chinese:Imagery	0.13	0.23	0.59	3,510.00	.556
fixed	NA	LanguageTurkish:Imagery	-0.15	0.14	-1.09	3,510.00	.274
ran_pars	PSA_ID	sd___(Intercept)	0.00	NA	NA	NA	
ran_pars	Residual	sd___Observation	99.73	NA	NA	NA	