Investigating Object Orientation Effects Across 14 Languages

Sau-Chin Chen[*1], Anna Szabelska[*2], Christopher R. Chartier[*3], Zoltan Kekecs[*4], Dermot Lynott[*5], Pablo Bernabeu[*5], Benedict Jones[*6], Lisa DeBruine[6], Carmel Levitan[*7], Kaitlyn M. Werner[*8], Kelly Wang[*8], Marina Milyavskaya[*8], Erica Musser[*9], Marietta Papadatou-Pastou[*10], Nicholas Coles[*11], Steve M. J. Janssen[*12], Asil Özdoğru[*13], Daniel Storage[*14], Harry Manley[*15], Benjamin Brown[*16], Krystian Barzykowski[*17], Thomas Rhys Evans[*18], Sara Álvarez Solas[*19], Manyu Li[*20], Leigh Ann Vaughn[*21], Balazs Aczel[*22], Attila Szuts[22], Carlota Batres[*23], William Chopik[*24], Kim Peters[*25], Jerome Olsen[*26], Martin Voracek[*26], Christian K. Tamnes[*27], Miroslav Sirota[*28], Dawn Liu[*28], Glenn P. Williams[*29], Arti Parganiha[*30], Babita Pande[*30], Sraddha Pradhan[*30], Noorshama Parveen[*30], Pratibha Kujur[*30], Priyanka Chandel[*30], Margaret Messiah Singh[*30], Chrystalle B. Y. Tan[31], John Protzko[32], Jack Arnal[34], Stefan Stieger[36], Marco Tullio Liuzza[37], Pavol Kačmár[38], Jozef Bavoľár[38], Monika Hricová[38], Gabriel Baník[39], Matúš Adamkovič[39], Ivan Ropovik[39], Peter Babinčák[39], Martin Seehuus[40], Vanja Ković[41], Kathleen Schmidt[42], & Elisabeth Oberzaucher[43]

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

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

[3] Department of Psychology, Ashland University, USA

[4] Department of Psychology, Lund University, Sweden

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

[6] Institute of Neuroscience & Psychology, University of Glasgow, Scotland, UK

[7] Department of Cognitive Science, Occidental College, USA

[8] Department of Psychology, Carleton University, Canada

[9] Department of Psychology, Florida International University, Miami, FL, USA

[10] National and Kapodistrian, University of Athens

[11] Department of Psychology, University of Tennessee, Knoxvile, TN, USA

[12] School of Psychology, University of Nottingham - Malaysia Campus, Malaysia

[13] Department of Psychology, Üsküdar University, Turkey

[14] Department of Psychology, University of Denver, USA

[15] Faculty of Psychology, Chulalongkorn University, Thailand

[16] School of Liberal Arts, Georgia Gwinnett College, USA

[17] Applied Memory Research Laboratory, Institute of Psychology, Jagiellonian University, Krakow, Poland

[18] School of Social, Psychological and Behavioural Sciences, Coventry University, United Kingdom

[19] Ecosystem Engineer, Universidad Regional Amazónica Ikiam, Tena, Ecuador

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

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

[22] Institute of Psychology, ELTE, Eotvos Lorand University, Hungary

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

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

[25] School of Psychology, University of Queensland, Australia

[26] Faculty of Psychology, University of Vienna, Austria

[27] Department of Psychology, University of Oslo, Norway

[28] Department of Psychology, University of Essex, United Kingdom

[29] Division of Psychology, School of Social and Health Sciences, Abertay University, Dundee, United Kingdom

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

[31] Department of Community and Family Medicine, Universiti Malaysia Sabah, Malaysia

[32] Department of Psychological and Brain Sciences, University of California, Santa Barbara, CA, USA

[34] Department of Psychology, McDaniel College, MD, USA

[36] Department of Psychology and Psychodynamics, Karl Landsteiner University of Health Sciences, Austria

[37] Catanzaro School of Medicine and Surgery, Magna Graecia University of Catanzaro, Italy

[38] Department of Psychology, Faculty of Arts, Pavol Jozef Šafarik University in Košice, Slovakia

[39] Faculty of Education, University of Presov, Slovakia

[40] Department of Psychology, Middlebury College, USA

[41] Department of Psychology, Faculty of Philosophy, University of Belgrade, Serbia

[42] Department of Psychology, Southern Illinois University Carbondale, USA

[43] Department of Evolutionary Anthropology, University of Vienna, Austria

Author Note

Author contributions: Sau-Chin Chen developed study concept and proposed the initial design. Sau-Chin Chen and Anna Szabelska wrote initial proposal draft. Dermot Lynott and Benedict C. Jones gave many suggestions on the improvement of study design and analysis plan. Zoltan Kekecs shared his opinions on the power analysis and sampling plan. Authors who coordinated proposal development are Christopher R. Chartier, Carmel A. Levitan, Kaitlyn M. Werner, Erica D. Musser, Marietta Papadatou-Pastou, Nicholas A. Coles, Pablo Bernabeu, Niklas K. Steffens, and Kelly Wang. Christopher R. Chartier organized and conducted pilot study. Nicholas A. Coles developed the codes for drawing the map of participated laboratories. Authors with asterisk contributed their efforts for the substantial proposal revisions. Sau-Chin Chen, Anna Szabelska and Christopher R. Chartier managed the project's OSF (view-only link: https://tinyurl.com/yb6sg2df).

Funding statement: Glenn P. Williams is supported by the Leverhulme Trust Research Project Grant (RPG-2016-093). Krystian Barzykowski is supported by National Science Centre, Poland (2015/19/D/HS6/00641). Erica D. Musser is supported by the National Institute on Mental Health (1R03MH110812-01). Nicholas Coles is supported by the National Science Foundation Graduate Research Fellowship (R010138018).

Ethical approval statement: Our leading researcher, Sau-Chin Chen, has obtained the approval from the Research Ethics Committee, National Taiwan University (NTU-REC No.: 201805ES025). The latest status of ethical approval for all the participating laboratories is available at our public sheet (https://tinyurl.com/ya2ssl4u).

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

Abstract

Mental simulation theories of language comprehension propose that people automatically create mental representations of real objects. Evidence from sentence-picture verification tasks has shown that people mentally represent various visual properties such as shape, color, and size. However, the evidence for mental simulations of object orientation is limited. By this multi-laboratory project across multiple languages and laboratories, we had the lessons on the estimation of match advantage of object orientation. At first the estimations varied among the langauges and experimental platforms, sites and Internet. On the other hand, the imagery scores summarized from picture-picture verifcation responses were nearly equal across langauges and platforms. When we considered the imager socres as the predictor of the match advantage, its acccountibility was able to be excluded. These lessons showed the robust estimations would be avaiable from the studies on one langauge rather than all langauges.

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

Word count: 6218

Investigating Object Orientation Effects Across 14 Languages

Introduction

Researchers have explored that how human mind represents objects through numerous methods and investigated the underlying cognitive processes of mental simulation during language comprehension. The sentence-picture verification task is a well-known method for assessing the mental simulation of an object's physical features during sentence reading (Connell, 2007; De Koning, Wassenburg, Bos, & Van der Schoot, 2017a; Stanfield & Zwaan, 2001; **ZwaanEmbodiedsentencecomprehension2005?**; **ZwaanRevisitingMentalSimulation2012?**; **ZwaanLanguageComprehendersMentally2002?**). The task requires participants to read a probe sentence (e.g., *He saw the eagle in the sky.*) and then verify if the target object in the picture was mentioned in the probe sentence. The target object (e.g., *eagle*) is presented in one of the two ways: with a matching feature (e.g., *flying eagle*) or with a mismatching feature (e.g., *standing eagle*). The match advantage refers to faster reaction time for verifying the target object. Because the match advantages resulted from the implied perceptual features (e.g., shape, color, and orientation) of the object while reading the sentence. These findings have been believed to serve as evidence for how the mental representations of objects are formed and activated during language processing (e.g., Barsalou, 1999, 2009).

Studied features include matching shape (i.e., *flying eagle* versus *standing eagle*) and color (i.e., *chocolate ice cream* versus *vanilla ice cream*) and matching size (i.e., *lipstick* versus *big lipstick*, followed by a picture matching with the latter). These features have been shown to produce match advantages in sentence-picture verification tasks for English (**ZwaanEmbodiedsentencecomprehension2005?**; **ZwaanRevisitingMentalSimulation2012?**), and across many languages, including Chinese (Li & Shang, 2017), Dutch (De Koning, Wassenburg, Bos, & Van der Schoot, 2017a; Engelen, Bouwmeester, de Bruin, & Zwaan, 2011; Pecher, van Dantzig, Zwaan, & Zeelenberg, 2009; **RommersObjectshapeorientation2013?**), German (Koster, Cadierno, & Chiarandini, 2018), Croatian (**SeticNumericalCongruencyEffect2017?**), and Japanese (Sato, Schafer, & Bergen, 2013). In comparison to these findings, the match advantages of orientation (i.e., *toothbrush in the cup* versus *toothbrush on the washbasin*, followed by a picture matching with the latter) are relatively small (**ZwaanRevisitingMentalSimulation2012?**) and inconsistent between languages (Chen, de Koning, & Zwaan, 2020). Considering the unequal strength of evidence among the features, theories of mental simulation hardly generate the predictions on the match advantage of the orientation for the particular language. Therefore we tested the match advantage of orientation in a crowd-sourced collaborative framework.

The match advantage of object orientation has been investigated to date using objects with particular characteristics. As in the example illustrated in 1, a picture of a "horizontal nail" matched the sentence "Tom hammered the nail into the wall" and a picture of a "vertical nail" matched the sentence "Tom hammered the nail into the floor." With this manipulation, the response data from two object orientations can be labeled "matching pairs" and "mismatching pairs," respectively. Specifically, all the materials of Stanfield and Zwaan (2001) were manipulable objects, such as nails, pens, and scissors. This choice restricted the object orientations to horizontal and vertical.

Investigations on intrinsic and extrinsic properties

Simulating object properties is a major topic in research examining embodied language understanding (Scorolli, 2014). Shape, color and size are identified as intrinsic properties which means these features are relatively independent of the observer's state. Regardless of the location or state of observers, the object tends to keep a certain shape, color, and size. Extrinsic properties, on the other hand, which have a prime example in object orientation, are more dependent on the observer's state. Koster, Cadierno, and

Chiarandini (2018) reviewed evidence suggesting that orientation effects are more consistent when the properties presented as stimuli as stimuli are intrinsic (e.g., shape, color and size) properties, rather than extrinsic (e.g., orientation). The inconsistent match advantages of orientation have raised the controversy from the studies focused on English (Zwaan & Pecher, 2012) and Dutch (De Koning, Wassenburg, Bos, & Van der Schoot, 2017b, 2017a; Rommers, Meyer, & Huettig, 2013). Recently Chen, de Koning, and Zwaan (2020) and Koster, Cadierno, and Chiarandini (2018) have emphasized that the extrinsic properties caused various results across languages in comparison with the intrinsic properties. Based on two experiments, Koster, Cadierno, and Chiarandini (2018) found a null effect of orientation from German speakers but a strong match advantage of size from Spanish speakers. Chen, de Koning, and Zwaan (2020) evaluated the match advantages of orientation between large and small objects among English, Chinese and Dutch. Although their meta analysis indicated a stronger orientation match advantage for the large objects, the match advantage in Dutch appeared only when the Dutch participants read English probes.

This project depended on crowd sourced data collection to systematically evaluate the orientation match advantages across languages. Our exploratory analysis aimed to explore two linguistic aspects. First is the way languages encode the motion of objects in sentences. The probe sentences of orientation studies always contain multiple instances of motion. For the three languages in the past studies, English, German and Dutch, there is a tendency to encode manner in the verb and leave path out to satellite adjuncts (e.g., 'The ant walked towards the pot of honey'). Specifically, Verkerk (2014) found that these languages behave similarly in their use of manner-only constructions (e.g., 'The ant swam') and manner plus path constructions (e.g., 'The ant swam in the honey'). Verkerk only reports a difference in terms of the use of path-only constructions (e.g., 'The ant went to the feast'), which are more common in English. As such, the lexical encoding of motion does not seem to account for the cross-linguistic differences in the match advantages.

The lexical encoding of placement is the other language aspect this project explored. Chen, de Koning, and Zwaan (2020) and Koster, Cadierno, and Chiarandini (2018) noted that some Germanic languages, such as German and Dutch, often make the orientation of objects more explicit than English does. Specifically, whereas in English one could use the verb "put" in both "She put the book on the table" and "She put the bottle on the table," in both Dutch and German, one could instead say "She laid the book on the table," and "She stood the bottle on the table." In these literal translations from German and Dutch, the verb "lay" encodes a horizontal orientation, whereas the verb "stand" encodes a vertical orientation. This distinction extends to verbs indicating existence. As Newman (2002) exemplified, an English speaker would be likely to say "There's a lamp in the corner," whereas a Dutch speaker would be more likely to say "There 'stands' a lamp in the corner." However genuine these differences across the languages, it would be difficult for us to make any predictions for the experiment on this basis because placement verbs do not seem to be widespread enough in the stimuli of these experiments.

Progressing Investigations on Match Advantage

Stanfield and Zwaan (2001) ensured that the participants sufficiently understood the probe sentences and found significant match advantages of object orientation. Later studies on this topic have examined the association between the match advantage and certain cognitive abilities. Spatial cognition is one of the relevant areas, which may be measured with mental rotation tasks. Studies have suggested that mental rotation tasks offer valid reflections of previous spatial experience (Frick & Möhring, 2013) and of current spatial cognition (Chu & Kita, 2008; Pouw, de Nooijer, van Gog, Zwaan, & Paas, 2014). Chen, de Koning, and Zwaan (2020) investigated the relationship between the match advantage and mental rotation across three languages: English, Dutch, and Chinese. They introduced the picture-picture verification task to examine how individuals process the target pictures with the same mental image, regardless of their native language. This picture-picture verification task was a modified form of the mental rotation paradigm (Cohen & Kubovy, 1993). In each trial of this task, two pictures appeared on opposite sides of the screen. Participants verify whether the pictures represent identical or different objects. The verification times for pictures of identical objects presented in the same orientation (that is, two identical pictures presented in horizontal orientation or vertical orientation) were shorter than those presented in different orientations (one horizontal; one vertical).

Several explanations have been proposed for the inconsistent findings on the orientation match advantage, including the procedures in the later studies. Unlike the original study of Stanfield and Zwaan (2001), the latter studies did not require participants to recognize the probe sentences they had read. Without this memory task during verification trials, participants could pay less attention to the meaning of the probe sentence , in which case they were less likely to form a mental representation of the objects (e.g., Zwaan & van Oostendorp, 1993). In this regard, it is relevant to acknowledge the large effect that may be caused by variations among experiments on the same topic. Barsalou (2019) underscores the variability caused by different trials, individuals, and experiments, referring to the great variation caused in terms of a quantum mechanism. As a result, Barsalou argues that replicating effects should be regarded with surprise.

On the investigations of Englihs, Dutch, and Chinese, Chen, de Koning, and Zwaan (2020) found that the match advantages were inconsistent among languages and the effects of mental rotation were consistent across languages. These findings revealed two primary questions to be answered through this project: (1) “How much of the match advantage of object orientation can be obtained within different languages?” and (2) “What is the correlation between the mental rotation of objects and the match advantage across different languages?” In the search of the clues to answer the two questions, researchers could look into the evidence of each language and examine the language aspects that might support or violate the predictions of mental simulation theories.

Methods

Hypotheses and Design

Both the sentence-picture verification task and the picture-picture verification task have two independent variables - a shared between-participant independent variable and a task-specific within-participant independent variable. The shared variable is the primary languages of the participating laboratories. The task-specific variable in sentence-picture verification is whether or not sentence and picture have matching object orientations; the variable in picture-picture verification is whether picture and picture have identical object orientations. The dependent variable for both tasks is the response time the participant verify the target object.

In the sentence-picture verification task, we expect response time to be shorter when object orientations are matching than when mismatching. We expect to see the match advantage within each language but we make no hypotheses about the specific effect sizes in each language. In the picture-picture verification task, we expect shorter response time when the two pictures have identical orientations than different orientations. In addition, we compute an imagery score by taking the elapsed verification time between the orientation settings of critical object pictures (identical, different). If the mental rotation were the general cognitive aspect, we expect imagery scores to be the same across laboratories and languages. If the mental simulation shared the cognitive processing of mental rotation, the imagery scores could be the predictor as critical as the languages.

General Procedure

Participating laboratories will conduct both the sentence-picture verification task and the picture-picture verification task, taking 10 to 15 minutes in total (see Figure 2 for an outline of the general procedure). The sentence-picture verification task will start before or after the survey of Phills and Kekecs (in preparation). In the beginning of the sentence-picture verification task, participants will be instructed with an example. Each trial starts with a left-justified and horizontally centered fixation displayed point for 1000ms, immediately followed by the probe sentence. The sentence is presented until the participant presses the space key, thus acknowledging that they understood the sentence. Then the object picture is presented in the center of the screen until the participant responds. If no response is made, the object picture disappears from the screen after 2 seconds. Participants are instructed to verify the object picture mentioned in the probe sentence as quickly and accurately as they can. Following the original study (Stanfield & Zwaan, 2001), a test to recognize the presented probe sentence will be conducted after every three to eight trials of sentence-picture verification. This is to make sure that the participants have read each sentence carefully.

The picture-picture verification task will use the same object pictures. The procedure is like the sentence-picture verification task, with one exception. In each trial, two pictures of objects will appear beside the central fixation point until either the participant indicates that the pictures display the same object or two different objects, or the time dedicated for the trial (2 seconds) has elapsed. Two pictures showing the same critical object will appear in each “yes” trial; two pictures showing two different objects from the filler items will appear in each “no” trial. All the procedures are compiled in OpenSesame scripts (Mathôt, Schreij, & Theeuwes, 2012).

Before the pandemic outbreaks happened in 2020 spring, 2/3 of participating laboratories had completed the studies. The rest of teams had to stop their data collection plans because of the local lockdown. The core project team decided to migrate the study from in the laboratories to on the Internet. To decrease the differences between laboratories and Internet circumstances, we migrated the original python codes to javascript codes. After months of tests, the revised script was able to collect the data through JATOS server(Lange, Kühn, & Filevich, 2015). The rest of teams conducted the Internet study between February to June 2021. The changes of the procedure had been priorly approved by Journal editor and reviewers in October, 22, 2020.

Participant Characteristics

This study were conducted within the Psychological Science Accelerator (PSA, Moshontz et al., 2018), a network of globally distributed psychological science laboratories. Before the pandemic outbreak, 2340 participants (1104 females; averagely 21.46 years old) from 33 laboratories joined and finished the study. After the study migrated to Internet, 1403 participants (926 females; averagely 23.75 years old) completed the study. Internet participants at the beginning heard the auditory instruction and had to correctly answer at least 2 of 3 questions. All the participated laboratories had the approve from their local/institutional ethics review board or committee before the data collection. Appendix 1 summarized the average characteristics by laboratory.

Material

The latest instructions and experimental scripts are available at the project OSF repository (https://osf.io/e428p/). For the study on Internet, a recorded verbal brief played at first. Participants soon had a test for the confirmation they art the native speakers of the targeted language. All the verbal briefs were packaged in each language-specific scripts. Appendix (#1) described the deployment of the scripts and the results of participants' fluency test.

Results

No matter which data sources, the studies were conducted in site or on Internet, the sequential analysis were managed as the preregistered plan. Among the studies conducted in sites, 1979 participants finished sentence-picture verification task and passed the preregistered exclusion criterion (accuracy percentile > 70%); 2007 participants finished picture-picture verification task. -28 participants' rawdata files were lost because the experimenters did not upload the files or submitted the wrong files. Among the studies conduced on Internet, 1337 participants finished sentence-picture verification task and passed the preregistered exclusion criterion; 1402 participants finished picture-picture verification task. We excluded 1 participant from one laboratory (USA_033) because this participant did not complete the picture-picture verification. All the analysis scripts are available in the source files of this article (OSF link).

Intra-lab analysis during data collection

During the data collection, the sample size justification of each laboratory depended on the Bayesian sequential analysis (Schönbrodt, Wagenmakers, Zehetleitner, & Perugini, 2017). Once a laboratory replied their progress, the latest results were updated on the public website. Till the end of data collection, only two laboratories (HUN_001, TWN_001) stopped the data collection because the sequential analysis indicated the latest collected data reached the preregistered criterion ($BF_{10} = 10 \; or - 10$). Many laboratories stopped the data collection when they reached their registered sample size. Some laboratories did not reach the minimal criterion because of the following reasons: (1)their works were interrupted by the pandemic outbreak; (2)the participants performed worse through the online study; (3) the laboratories allocated the seats for the foreign participants; (4) A language (Serbian) has a various of scripts. At the end of data collection, there are 18 languages registered in this project. The collected data were from 47 laboratories and 3316 participants. The details of the public website are available in Appendix 2.

Inter-lab analysis after data collection

In addition to the preregistered variables included languages and laboratories, we also evaluated the data sources, in site/on Internet, in each preregistered analysis protocol. In the preregistered plan we decided two approaches to estimate the match advantages. The first approach summarized the median reaction times by condition as the original studies (Stanfield & Zwaan, 2001; Zwaan & Pecher, 2012). The planned meta-analysis evaluated the global effect size based on the medians. The second approach used the raw response time then computed the fixed effects in the mixed-effect models (Baayen, Davidson, & Bates, 2008).

Identify the outliers in each data set. By each laboratory data set, we summarized how many response time points beyond the third quantile in that data set. 0 laboratories had no outliers. Among the data sets showed outliers, the averaged proportion of outliers was 0.25. Appendix 1 Table S4 illustrated the distribution of outliers by laboratory. In sum of the data excluded the outliers, Table 1 and Table 2 respectively summarized the match advantages collected in sites and on Internet by each language.

(Insert Table 1 about here)

(Insert Table 2 about here)

Meta analysis of match advantages across laboratories. Because the first preregistered analysis plan did not consider the data collected on Internet, we conducted three parts of meta-analysis to evaluate the global effect size. In the first part the data sources were combined. In the second and third parts we conducted the meta analysis on the data in sites and data on Internet respectively. In each part we computed the effect size by data set then estimated the global effect size. Considering the small sample size may bias the estimated effect size, 9 data sets were excluded from the analysis because the available sample size was less than 25.

Meta-analysis across all the data sets showed a positive but weak effect size. Figure 3 showed the larger variances happened to some laboratories in comparison to the others.

(Insert Figure 3 about here)

When we focused on the data collected in sites, the variation among the laboratories was small although the overall effect size was positive but weak. Figure 4 showed that only one laboratory (HUN_001) found a true positive match advantage.

(Insert Figure 4 about here)

When we focused on the data collected on Internet, the variations among laboratories were obvious larger than the data collected in sites. The overall effect size tended to be negative. Figure 5 showed that only one laboratory (NZL_005) found a true positive match advantage.

(Insert Figure 5 about here)

Evaluate match advantage in mixed-effect models. As with the original plan for the mixed-effect models, all our works started from the model having the matching condition as the only one fixed effect. Because the final data came from sites and Internet respectively, we had to evaluate if one mixed-effect model sufficiently fit all the data. Otherwise we had to build the models for each data set. The results indicated that the data collected from sites and Internet had to be analyzed separately. Based on the last recommended practices (Bates, Kliegl, Vasishth, & Baayen, 2015; Brauer & Curtin, 2018), we decided the fitted model in terms of the maximal random structure and the least convergence problems. The final models indicated the interaction of languages and match advantages from each data source. This section summarized the two final models from the two data sources as below. All the evaluated mixed-effect models are reported in Appendix 3.

With the same sample size justification principle as the meta analysis, we excluded the languages with more than 25 participants in each data source before we conducted the mixed-effect models. Portuguese in the sites data and Norwegian in the Internet data were excluded from the analysis of final models. In each data set, we compared the fitness among the models with and without the random slope of matching condition. Both results indicated the model without the random slope had the best fitness. The model from the sites data indicated the fixed effect of match advantage was positive but insignificant: coefficient = 5.73 ms, $p = 03$. However, the model from Internet data indicated the fixed effect of match advantage was negative but insignificant: -3.92 ms, $p = 13$.

In terms of the final models, we plotted the predicted response times of matching condition in the function of languages. Figure 6 illustrated the estimated responses times based on the sites data. Greek tended to have the detectable match advantage although the responses times were the longest among languages. Figure 7 illustrated the estimated responses times based on the sites data. Serbian showed a weak match advantage but had the longest responses times. Many languages had inconsistent results between the studies in sites and on Internet, such as English, German, and Traditional Chinese.

(Insert Figure 6 about here)

(Insert Figure 7 about here)

Analysis on imagery scores. Imagery scores are the dependent measurement of picture-picture verification task. Dependent measurements were summarized by the elapsed responses times from different to identical orientation settings. Prior to the data collection, we assumed the imagery scores of every language group would be nearly equal. Because the real data came from sites and Internet, we managed the mixed-effect models to verify our prior assumption and to evaluate the confounding of data sources and laboratories. Each final model had the better fitness in consideration of the maximal random structure and the least convergence problem. The models indicated that the data sources and laboratories affected little on the imagery scores. In addition to summarize the fitted model of orientation setting and languages interaction as below, all the evaluated mixed-effect are reported in Appendix 4.

The best fitted model had the random intercepts of participants, targets, and laboratories, but had no slopes of orientation setting. The fixed effect of orientation setting was significant: coefficient = 33.35 ms, p $<$ .001. The estimated response times illustrated in Figure 8 indicated that the imagery scores measured in each language study were consistently positive. English result also showed the large sample size decreased the variations.

The last planned analysis attempted to build the regression model that the imagery scores is the predictor of the match advantage. The above analyses confirmed that data sources did not alter the imagery scores but alter the match advantage. Therefore we evaluated the regression models respectively. Because the random slopes of items in the analyses of the match advantage were zero, the integrated data for the regression models were the summarized data by participant.

The regression models should have the match advantage as the dependent variable. The independent variables in the fitted model should have the languages and the imagery scores. If the imagery scores could predict the match advantage, the model with two independent variables has to fit the data better than the model with language only. For the analysis based on the sites data, the model with language only has a better fitness than the model of two independent variables. For the analysis on the Internet data, on the contrary, the model of two independent variables has a better fitness than the model with language only. However, the coefficient of imagery scores was below the significant level, $b = -0.28$, 95% CI $[-1.68, 1.12]$, $t(1132) = -0.40$, $p = .691$. Appendix (#5) summarized the coefficients of the models included in this analysis.

Discussion

By the first question, we attempted to detect the robust match advantage in some languages. The overall results showed the positive but weak match advantage of object orientation, and no language result. None of the planned analyses indicated some languages had the larger advantage than the other languages. On the other hand, because the final data were collected from more than registered languages and from Internet, the analyses indicated that these unexpected changes altered the pattern of match advantages.

By the second question, we were curious if mental simulation of object orientations associated with the mental rotation mechanism. This study measured the mental rotation processing by the imagery scores summarized from the picture-picture verification responses. Unlike the match advantage of object orientation, the languages and data collection procedures affected little on the imagery scores. In the planned regression analysis based on the data collected in sites, the imagery scores failed to account the match advantage. However, although the intercept of imagery scores was insignificant, the best fitted regression model based on the data collected from Internet had the languages and imagery scores as the independent variables.

In this discussion we summarized plausible causes that the data collection procedure and the language aspects altered the match advantage. These causes could constrain the evidence level to evaluate the true effect size of match advantage. Without the precise estimation of true effect, there would be the difficulties to increase the verisimilitude of mental simulation theory (van Rooij & Baggio, 2021). Thus we suggested the recommendations to accumulate the data for the estimation of true effect.

Measurment issues across platforms

Averagely the sentence-picture verification responses collected from Internet was 878 ms longer than those collected in sites. In consideration of the pre-analysis plans, we did not evaluate the elapsed time between the data collection procedures in use the methods more than the meta analysis and the mixed-effect models. Although the extended replications and advanced meta-analysis are required, the current results pointed out the factors that could delay the responses on the web pages.

Measurement precision and accuracy in web-based experimental applications has been the frequently discussed issues since the experimental researchers are used to collect the response time data from Internet. Our primary concern is if OpenSesame caused higher measurement bias in sites than on the web pages. In sum of the frequently used desktop applications, OpenSesame windows version has the fairly fine precision and relatively low variation in compared to the others (see Table 2 of Bridges, Pitiot, MacAskill, & Peirce, 2020). Although Bridges, Pitiot, MacAskill, and Peirce (2020) did not evaluate OSWEB performance, we could take the results of PsychoPy (Peirce et al., 2019) because it is the core of OpenSesame. In many combinations of operating systems and web browsers, PsychoPy had higher precision than OpenSesame desktop version but had 25 to 50 ms lag (see Table 3 of Bridges, Pitiot, MacAskill, & Peirce, 2020). This lag obviously is shorter than the elapsed time between our data collection sources.

By the instruction of sentence-picture verification task, the studies that showed the significant match advantages always had the settings assist the participants maintain awareness on the sentences they read. These studies were conducted in sites(Stanfield & Zwaan, 2001) and on the web pages(Chen, de Koning, & Zwaan, 2020; Zwaan & Pecher, 2012). In the consideration of the resources each site managed, each team fellows just followed the minimal instructions on the data collection. Therefore we could not confirm each participant maintained their attention as awaken as the previous studies. According to our meta analysis, however, there were some teams found the significant match advantages in site (NOR 003) and on web page (NZL 005).

True effect of match advantage for single language

Although it is difficult to conclude the overall true effect of match advantage, we could ask the feasibility to estimate the true effect of a language. To have the estimation of true effect, the power analysis in the preregistered plan revealed that the required participants and items were amount to 400 and 100. When a study employed 24 items only, we would require more than 1200 participants. Therefore only the English data has the nearly sufficient sample size for the evaluation of true effect. In addition, this study collected the

English from many countries and from two data collection methods. This extended analysis has to consider the English systems (American, British) and the data collection methods. The participants who are not native English speakers can not be attributed to any English system. Thus the analysis excluded the data from the non-English countries.

Based on the data from 1216 English participants, we tested the mixed-effect model having matching condition, English systems, and data sources as the fixed effects. In terms of the recommended practices (Brauer & Curtin, 2018), we converted English systems and data sources to standardized coding. The fitted model indicated the data sources as the only one significant coefficient: -869.70 ms, $p < .01$. Although the fixed effect of matching condition was insignificant, this model indicated the interaction of matching condition and English systems: -446.26 ms, $p = 03$. Figure 9 showed the match advantages happened in the Internet data but only for American English.

(Insert Figure 9 about here)

Although the analyses on the English data showed the discrepancy between the studies in sites and on Internet, the variations may be within the range the past results summarized. Before the current study, the English studies that reported the significant match advantages had the median response times were between 800 ms and 1100 ms(see Table 1 of Zwaan & Pecher, 2012). On the other hand, the English study that failed to report the significant result had the median response time lower than 700 ms (see Table 2 of Chen, de Koning, & Zwaan, 2020). So far the current findings can not sufficiently support the idea that the match advantage of object orientation usually happen in the slow responses. The task settings obviously changed the participants' performance then determined the availability of match advantage.

Limitations and Generalizability

This study launched the platform for many languages collected their first data set in the literature of mental simulation studies, although we initially attempted to find out the global effect for the mental simulation theories. The current results revealed the difficulties to investigate the mental simulation effects across languages, especially the targeted effect was too small to be detected. Researchers who are developing the mental simulation theories should learn when we had the positive evidence in one language, we may find the negative results in another language. Here are the two suggestions to whom intend to break the limitations on accumulating the reproducible results and to increase the generalizability of mental simulation theories: (1) in a language researchers have to exhaustively evaluate any aspect altered the match advantage; (2) in a language the researchers accumulate the empirical results in terms of a well organized protocol.

For the inclusion of the languages as many as possible, the original plan never consider the particular aspects in any languages. Our concerns were to lease the burden on the volunteer translators and to increase the participation of fellows who are not psycholinguistic specialists. This arrangement may cause the non-English stimuli sentences contained the aspects inhibited or facilitated the responses to the matched pictures. Chen, de Koning, and Zwaan (2020) had figured out that Dutch verb inflections explicitly described the object orientations. In this study many teams had no positive results till their maximal sample size, but there were non-English teams (Chinese, Thai, and Hungarian) stopped the data collection before they reached the maximal sample size. For these languages researchers are recommended to find out any aspect could increase the plausibility to detect the match advantage of object orientations.

Because of the unexpected situation, the data collection had to be conducted on Internet instead of in the sites. This lesson revealed the naive issue if the estimation of true effect could be reproduced, all the data collection procedure had to follow the preregistered method. The discussions about this issue must be proceed in each language which could collect sufficient amount of samples. Form this study, we learned the practices researchers could follow when they aim to some mental simulation effect of a language. (1) The preparation of stimuli follows a reproducible protocol. This protocol should be based on the researchers' consensus. (2) When the researchers could lead their participants into the experimental circumstance in the same manner, the data collection could be conducted either in sites or on Internet. (3) In a language the data could be collected in different timings, but the sample size should be decided by the theoretical interests.

Conclusion

On estimating the true effect of match advantage of object orientation, this study has yet concluded either the global effect or the effect in a particular language. The regression analysis indicated that the affection of mental rotation on the match advantage was not as important as the languages. Rather than the estimation of global mental simulation effect, the current works revealed the feasibility to estimate the match advantage of object aspects in a language. The mental simulation theory would provide the particular assumptions and predictions in terms of the particular language aspects. Our understandings of mental simulation may improve with the research quality in each corresponding language studies rather than the scale of a single massive collaborative research project.

Acknowledgement

Contributions: Chris Chartier and Jeremy Miller managed and monitor the progress. Erin Buchanan provide guides to improve the progress website management and manage JATOS server. Marton Kovacs and Balazs Aczel helped update the scripts of sequential analysis.

References

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

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

Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. *Philos Trans R Soc Lond B Biol Sci*, *364*, 1281–1289. https://doi.org/10.1098/rstb.2008.0319

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

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious Mixed Models. Retrieved from http://arxiv.org/abs/1506.04967

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

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

Chen, S.-C., de Koning, B. B., & Zwaan, R. A. (2020). Does Object Size Matter With Regard to the Mental Simulation of Object Orientation? *Experimental Psychology*, *67*(1), 56–72. https://doi.org/10.1027/1618-3169/a000468

Chu, M., & Kita, S. (2008). Spontaneous gestures during mental rotation tasks: Insights into the microdevelopment of the motor strategy. *Journal of*

Experimental Psychology: General, *137*(4), 706–723. https://doi.org/10.1037/a0013157

Cohen, D., & Kubovy, M. (1993). Mental Rotation, Mental Representation, and Flat Slopes. *Cognitive Psychology*, *25*, 351–382. https://doi.org/10.1006/cogp.1993.1009

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

De Koning, B. B., Wassenburg, S. I., Bos, L. T., & Van der Schoot, M. (2017a). 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 Koning, B. B., Wassenburg, S. I., Bos, L. T., & Van der Schoot, M. (2017b). Size Does Matter: Implied Object Size is Mentally Simulated During Language Comprehension. *Discourse Processes*, *54*(7), 493–503. https://doi.org/10.1080/0163853X.2015.1119604

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

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

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

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 ERPs 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

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

Moshontz, H., Campbell, L., Ebersole, C. R., IJzerman, H., Urry, H. L., Forscher, P. S., … 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). Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/tsl.51.02new

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

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., … Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy.

Behavior Research Methods, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Phills, C., & Kekecs, Z. (in preparation). *Gendered Social Category Representations*.

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

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

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

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 *The Routledge handbook of embodied cognition* (pp. 145–156). Routledge.

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

van Rooij, I., & Baggio, G. (2021). Theory Before the Test: How to Build High-Verisimilitude Explanatory Theories in Psychological Science. *Perspectives on Psychological Science*, *16*(4), 682–697.

https://doi.org/10.1177/1745691620970604

Verkerk, A. (2014). *The Evolutionary Dynamics of Motion Event Encoding.* Radboud Universiteit Nijmegen.

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., & 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

tags: `PSA002`

Table 1

Summarized match advantages by langauge. Data collected in sites.

Language	N	Mismatching	Matching	match advantage
English	473	567(93.73)	561(94.63)	6.00
German	74	553(95.5)	555(95.61)	-2.00
Greek	73	701(90.53)	692(91.1)	9.00
Hebrew	109	556(97.09)	568(95.57)	-12.00
Hindi	59	594(87.57)	627(91.24)	-32.50
Magyar	97	607(95.45)	614(95.62)	-7.50
Norwegian	92	577(96.29)	581(96.47)	-3.75
Polish	37	561(94.82)	585(95.72)	-24.00
Portuguese	5	588(96.67)	580(98.33)	8.00
Simple Chinese	60	632(91.67)	619(92.78)	13.75
Slovak	103	589(96.6)	606(94.82)	-17.00
Spanish	95	646(92.63)	644(92.54)	2.00
Thai	37	637(90.77)	596(87.61)	41.50
Traditional Chinese	78	611(94.23)	616(93.27)	-5.25
Turkish	137	631(95.26)	618(94.89)	12.50

Table 2

Summarized match advantages by langauge. Data collected on Internet.

Language	N	Mismatching	Matching	match advantage
Arabic	79	522(77.95)	544(76.69)	-21.50
Brazilian Portuguese	50	778(94)	779(94)	-0.75
English	630	644(93.16)	645(93.51)	-1.00
German	127	657(95.6)	668(95.21)	-11.50
Norwegian	16	844(96.88)	636(93.75)	208.25
Portuguese	40	672(96.46)	664(94.79)	9.00
Serbian	129	712(93.48)	722(94.38)	-9.50
Traditional Chinese	34	616(91.18)	664(95.34)	-47.75
Turkish	59	779(92.8)	780(90.68)	-1.00

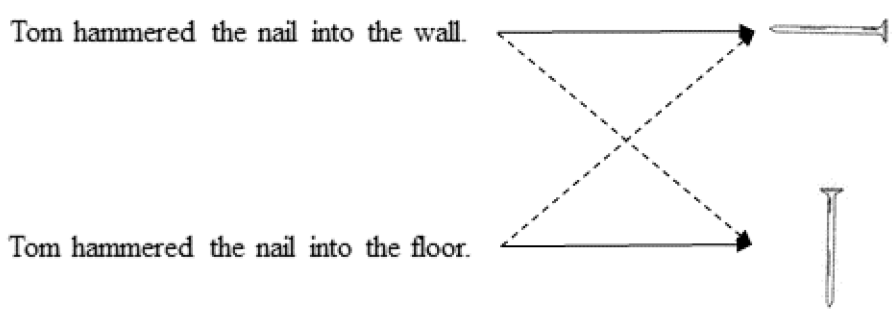
Tom hammered the nail into the wall.
Tom hammered the nail into the floor.

Figure 1

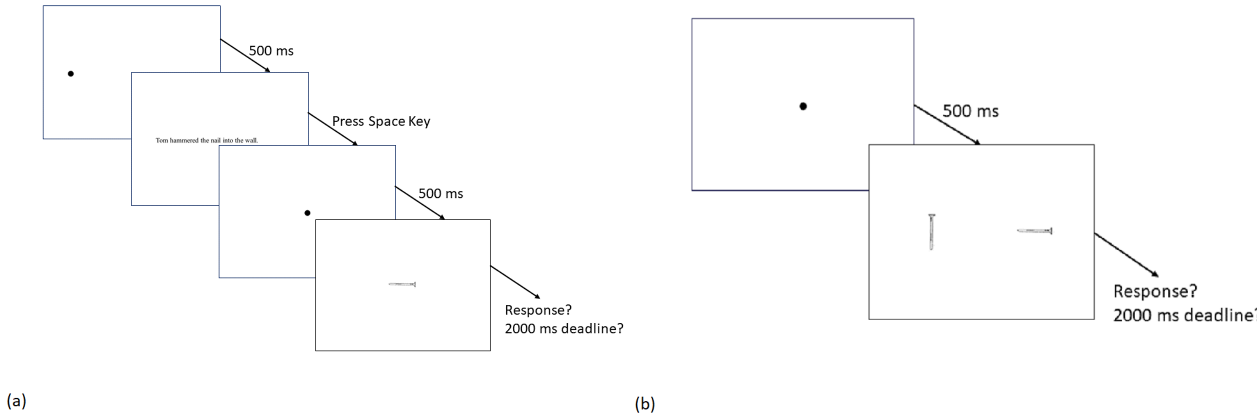

Figure 2. Procedures of stimuli presentation and response. (a) Sentence-picture verification task; (b) Picture-picture verification task.

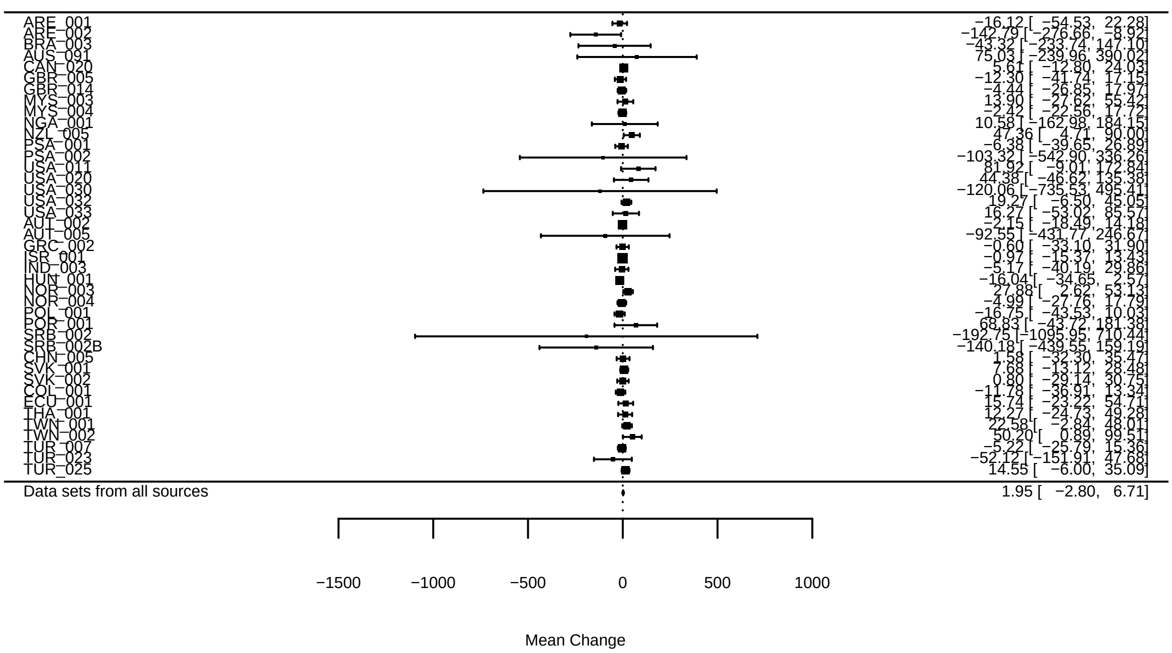

Figure 3. Meta analysis for all data

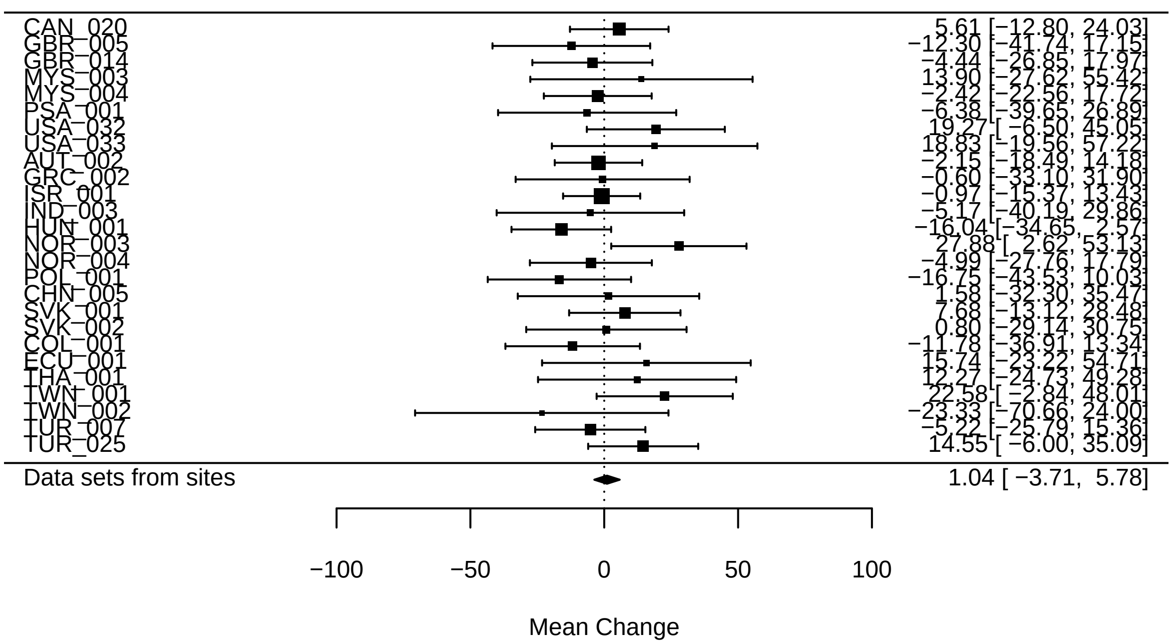

Figure 4. Meta analysis for site data

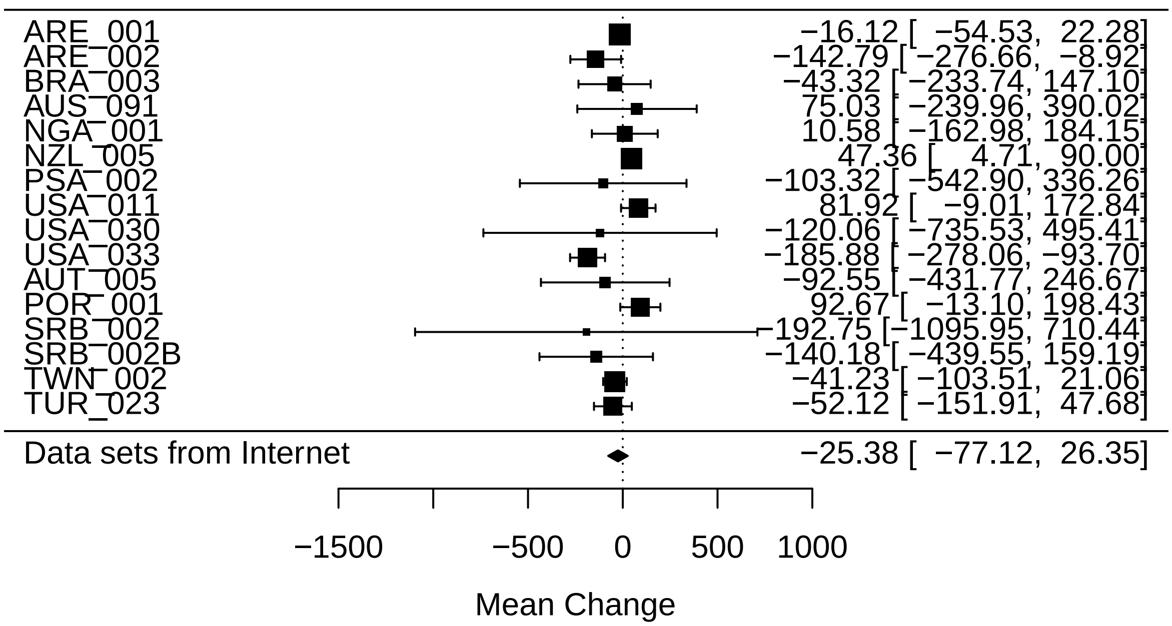

Figure 5. Meta analysis for Internet data

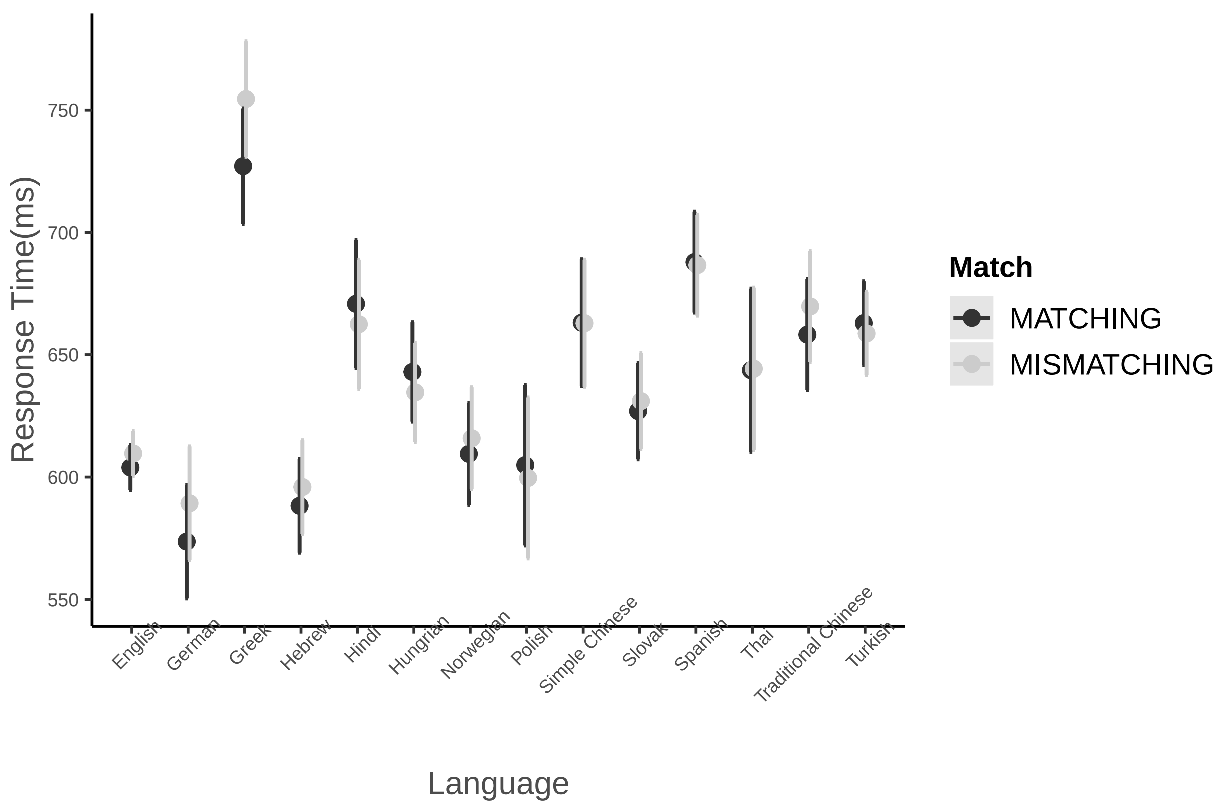

Figure 6. Estimated response times and standard error of sentence-picture verification in the function of langauges (by sites data)

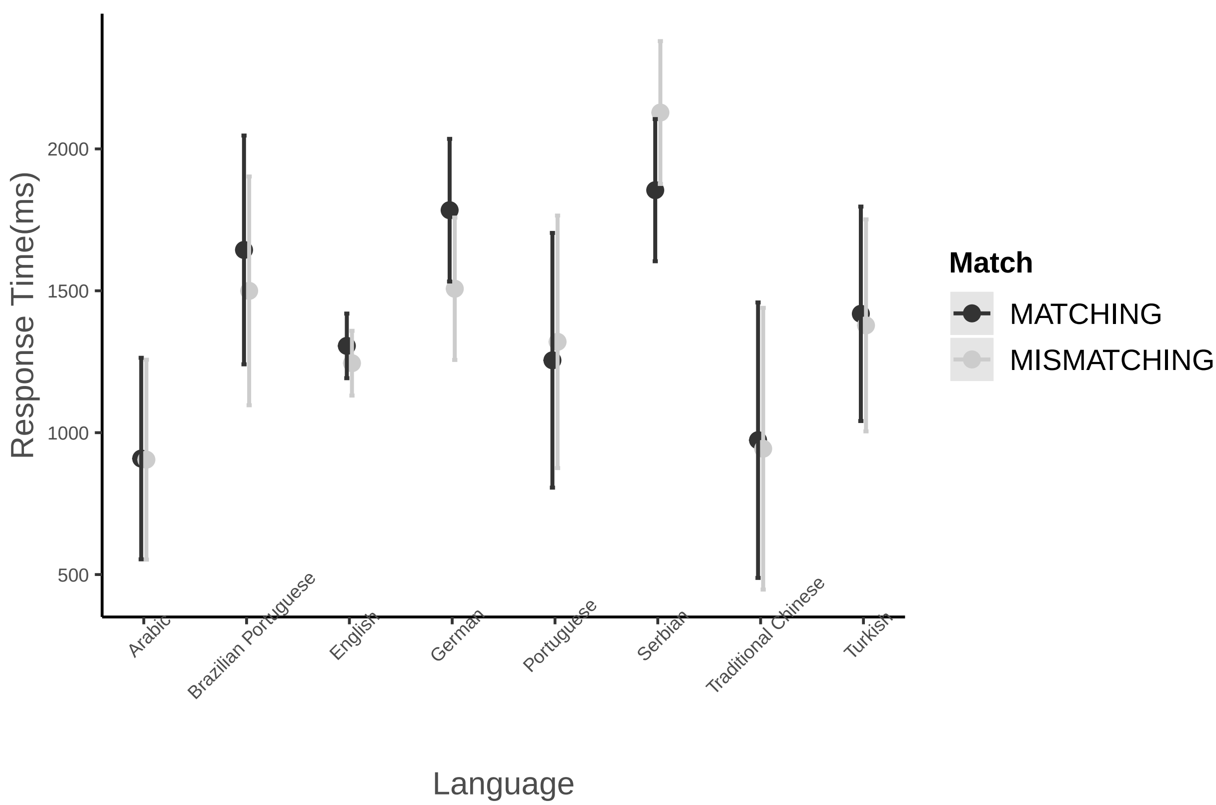

Figure 7. Estimated response times and standard error of sentence-picture verification in the function of langauges (by Internet data)

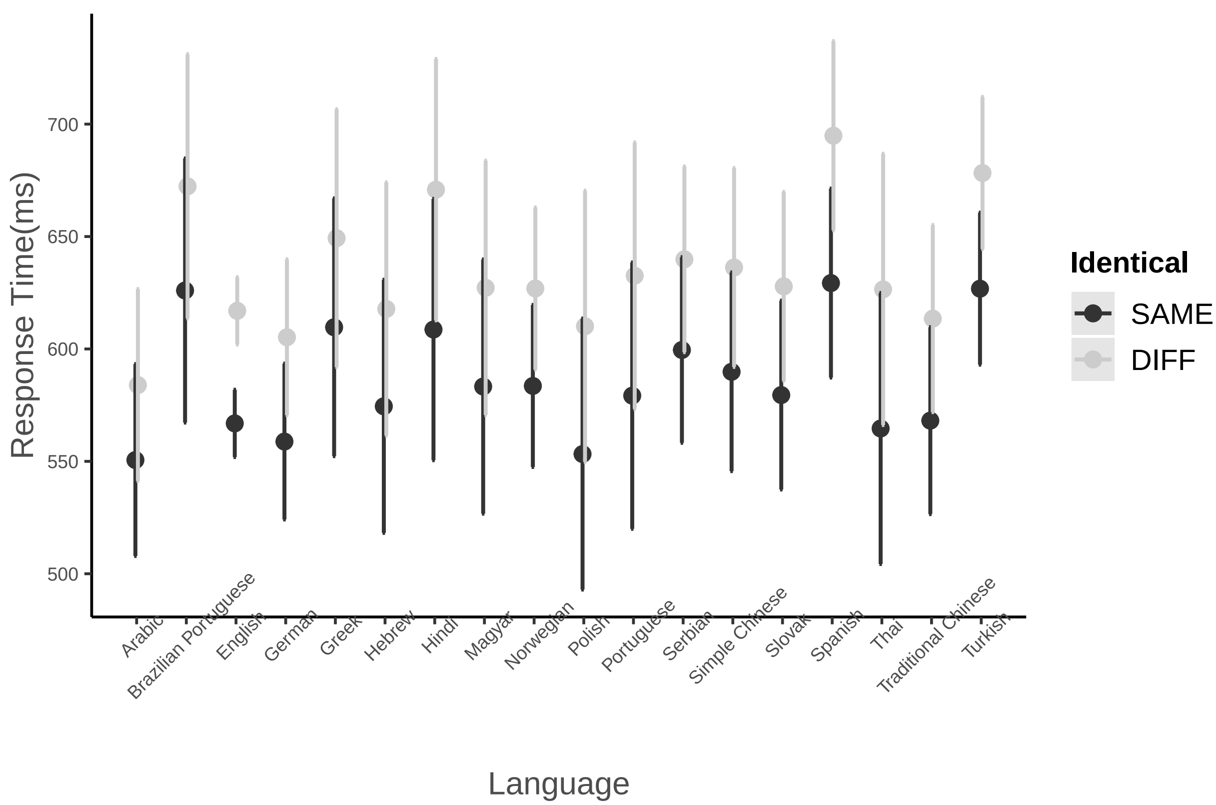

Figure 8. Estimated response times and standard error of picture-picture verification in the function of langauges (by all data)

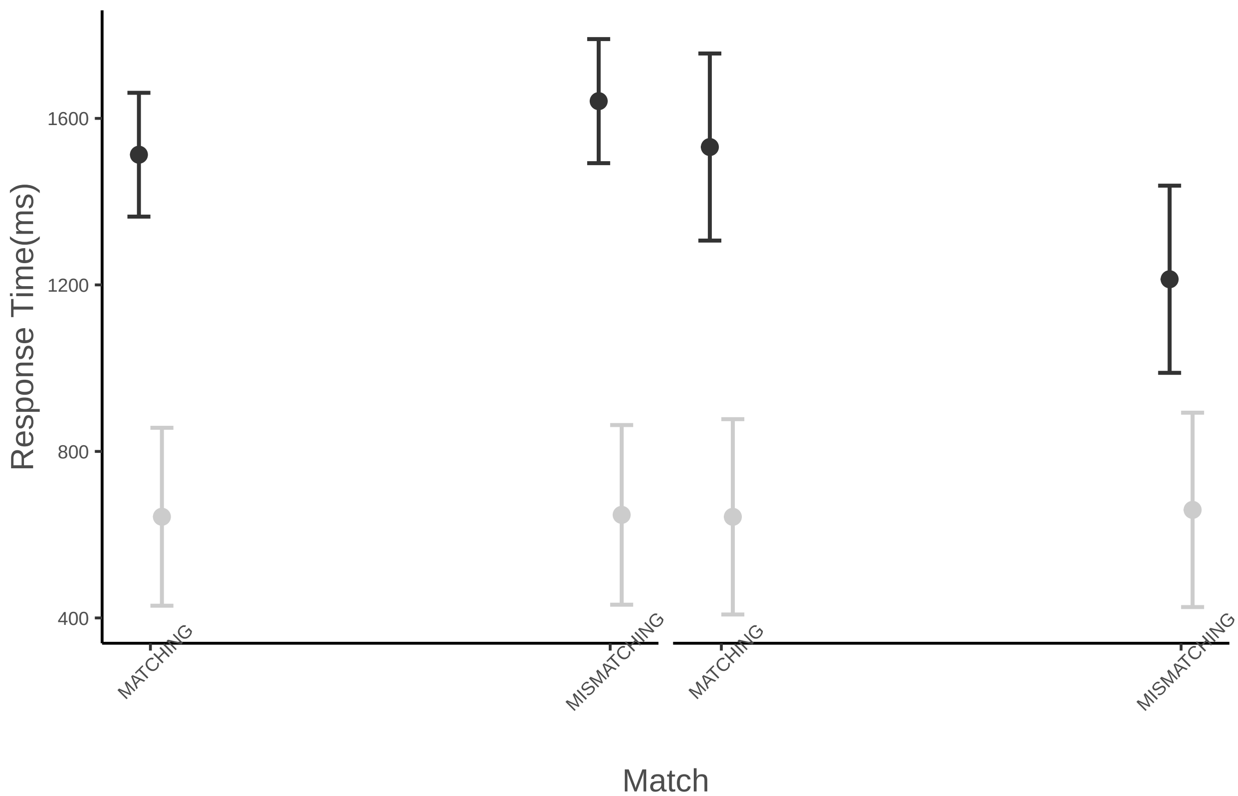

Figure 9. Estimated response times and standard error of sentence-picture verification in the function of English systems (left panel: UK; right panel: US) and data sources(black lines: Sites; grey lines: Internet).