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. We report a study that investigates the match advantage of object orientation across speakers of different languages. This multi-laboratory project aims to achieve two objectives. First, we examine the replicability of the match advantage of object orientation across multiple languages and laboratories. Second, we will use a mental rotation task to measure participants' mental imagery after the sentence-picture verification task. The relationship between the participants' performance of the two tasks will provide a cross-linguistic examination of perceptual simulation processes. With the evaluation of individual mental imagery ability and potential linguistic moderators, we expect a robust estimation of match advantage of object orientation.

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

Word count: 3931

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 object's physical features during sentence reading (Connell, 2007; de Koning, Wassenburg, Bos, & Schoot, 2017; Stanfield & Zwaan, 2001; Zwaan & Madden, 2005; Zwaan & Pecher, 2012; Zwaan, Stanfield, & Yaxley, 2002). 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).

Among the studied features, objects with matching shape (i.e., *flying eagle* versus *standing eagle*) and color (i.e., *chocolate ice cream* versus *vanilla ice cream*) have been shown to produce match advantages across many languages, including English (Zwaan & Madden, 2005; Zwaan & Pecher, 2012), Dutch (de Koning et al., 2017; Rommers, Meyer, & Huettig, 2013), German (Koster, Cadierno, & Chiarandini, 2018), Croatian (Šetić & Domijan, 2017), and Chinese (Li & Shang, 2017). In contrast, match advantages of object orientation have been found in English (Stanfield & Zwaan, 2001; Zwaan & Pecher, 2012) and German (Koster et al., 2018), but was absent in the other languages, such as Dutch (de Koning et al., 2017; Engelen, Bouwmeester, de Bruin, & Zwaan, 2011; Rommers et al., 2013). Several explanations have been proposed to explain these inconsistent findings, including the different experimental procedures in the earlier mentioned studies. Unlike the original study (Stanfield & Zwaan, 2001), the former 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 (Zwaan, 2014), in which case they were less likely to form a mental representation of the objects (e.g., Zwaan & van Oostendorp, 1993).

Extended Studies on Embodied Cognition

Among studies which focused on the conceptual representations of perception and action, there is a shared assumption that linguistic cues activate sensory and motor information processing outside the non-linguistic context (for a review, see Zwaan & Pecher, 2012; for a critique, see Mahon & Caramazza, 2008). A highly cited research topic related to it is the action–sentence compatibility effect that is measured by the sensibility judgment task (Glenberg & Kaschak, 2002). This sensibility judgment task requires participants to verify the sensibility of sentences by pressing the response buttons toward or away from their bodies. The sentences implied the direction of body movement (e.g., *You gave the earring to Susan*; *Joe sends the card to you*). Participants made faster judgements of sentences that implied the congruent direction. Since the first study of action–sentence compatibility effect was published, researchers valiantly tested the related hypotheses and investigated similar effects with other methods. For example, Beilock, Lyons, Mattarella-Micke, Nusbaum, and Small (2008) found that specialised motor experience (via sport) enhances the mental processing of sentences that are related to the specific sport practised.

As like the action-sentence compatibility effect, some researchers have suggested that the match advantage of object orientation would associate with the activated motor information (Connell, Lynott, & Dreyer, 2012; Ping, Goldin-Meadow, & Beilock, 2014). Ping et al. (2014), for example, exposed individuals to the implied gesture video or to the blank screen between hearing the sentence and watching the object picture. They found that the facilitation effect appeared only in the condition where participants heard a sentence and watched the progressive action in the video and not in the condition where they heard a sentence only. Although these findings provide initial support for the association between mental representations of body movements and non-verbal communications (Alibali & Nathan, 2012; Novack & Goldin-Meadow, 2015, 2017), the tasks across these studies substantially differed. Without evaluating the true effect size of the match advantage, the results of subsequent studies would not only add little to the overall understanding of how body movements are mentally represented but also will not explain the inconsistency in the orientation effects across the studies.

Therefore, it is necessary to precisely estimate the true effect size in the original paradigm, before commencing with the theoretical and empirical probing for the orientation effect. Because single laboratory studies (de Koning et al., 2017; Rommers et al., 2013; Zwaan & Pecher, 2012) hardly address the possible causes of inconsistent findings, it is imperative to collect data across multiple languages and laboratories as many as possible.This project accumulates materials translated into many languages by the labs recruited from a distributed laboratory network (Moshontz et al., 2018), using the materials and procedure created by Stanfield and Zwaan (2001).

Progressing Investigations on Match Advantage

The match advantage of object orientation has been investigated to date using objects with particular characteristics. As in the example illustrated in Figure 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 the “matching pairs” and “mismatching pairs,” respectively. Specifically, all the materials of Stanfield and Zwaan (2001) were manipulable objects, such as nails, and pens, and scissors. This choice restricted the object orientations to horizontal and vertical.

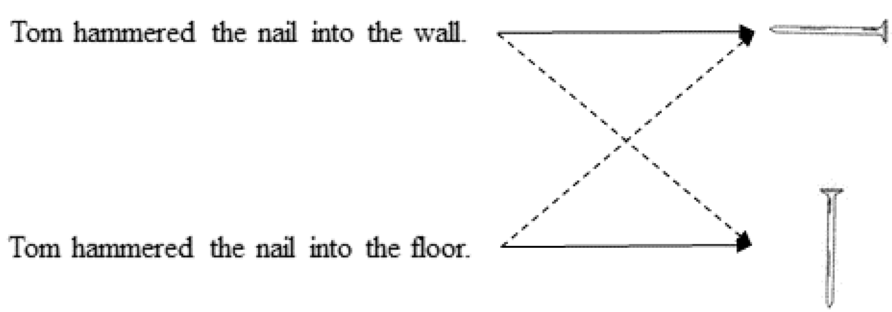

Figure 1. Example of matching (solid lines) and mismatching (dashed lines) sentence-picture pairs.

Stanfield and Zwaan (2001) ensured that the participants sufficiently understood the probe sentences and showed significant match advantages of object orientation. However, it is still unknown whether the match advantage associated with the general cognitive ability. Some further studies of mental simulation have suggested the relationship of mental rotation and spatial cognition (Chu & Kita, 2008; Pouw, de Nooijer, van Gog, Zwaan, & Paas, 2014). Chen et al. (2018b) 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 that 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).

To evaluate the relationship between match advantage and mental rotation, Chen et al. (2018b) employed larger (e.g., *anchor*, *gate arms*) and smaller (e.g., *pen*, *nail*) objects. According to the findings about the relationship of mental rotation and spatial cognition (Chu & Kita, 2008; Pouw et al., 2014), the authors assumed that response times would be longer for larger objects in both the sentence-picture verification task and picture-picture verification task. The results of the two tasks showed that large objects presented in different orientations take longer time to verify than small objects (Chen et al., 2018b). However, the meta-analysis of Chen et al. (2018b) indicated variation in match advantages across different languages but a consistent effect of mental rotation across different languages. Their study left 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?"

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 (n.d.). 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, 2011). The instructions and experimental scripts are available in the project's repository (Chen et al., 2018a).

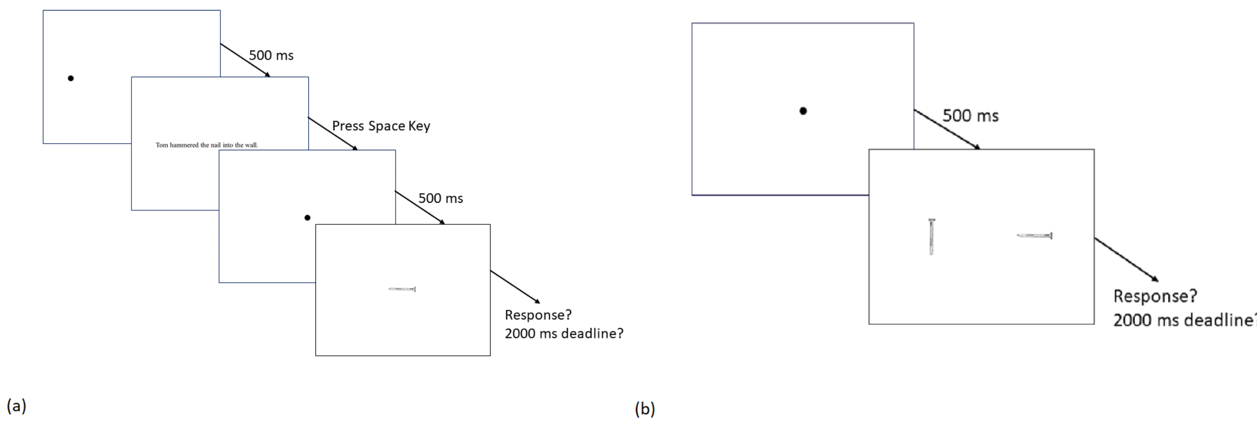

Figure 2. Procedures of stimuli presentation and response. (a) Sentence-picture verification task; (b) Picture-picture verification task.

Participant Characteristics

This study will be conducted within the Psychological Science Accelerator (PSA, Moshontz et al., 2018), a network of globally distributed psychological science laboratories. Every laboratory registered for this project will conduct this study using the protocol described in the general procedure. Upon registration, every participating laboratory will indicate the primary language for data collection and the participants in the study will be asked to specify their first language. If many of the participants within a given laboratory are familiar with more than two languages, the laboratories may collect additional data to perform secondary analyses based on their specific hypothesis. To maximize the diversity of languages, this study encourages the participation of laboratories from non-English speaking countries. According to the submission policy, the principal investigator (Sau-Chin Chen) has secured approval of this study by the Research Ethics Committee, National Taiwan University (NTU-REC No.: 201805ES025) before the submission. Before data collection, participating laboratories are required to obtain the approval from their local/institutional ethics review board or committee.

Power Analysis and Sampling plan

Because this project used the materials of the replication study (Zwaan & Pecher, 2012), we used two approaches of power analysis based on the archived data. Our first approach estimated the minimum sample size in terms of the average effect size of the match advantage. The Cohen's d of Zwaan and Pecher (2012) was 0.10, which was smaller than the original study, 0.13 (Stanfield & Zwaan, 2001). To reach the effect size 0.1 and 0.2 at a statistical power of 80% for a one-tailed test, the required sample size estimated by G*Power (Faul, Erdfelder, Buchner, & Lang, 2009) are 620 and 156 participants, respectively. In use of Bayesian meta-analysis (Morey, Rouder, Love, & Marwick, 2015), the effect size of 0.1 is equal to $BF_{10} = 21$. Because $BF > 10$ has been suggested as strong evidence and many laboratories will conduct this study at their first time, we recommend that each laboratory uses the uninformative prior ($\delta = 0.707$, see Morey & Rouder, 2011; Rouder, Speckman, Sun, Morey, & Iverson, 2009) to evaluate the evidence value during data collection.

Our second approach considered the match advantage as the fixed effect, and participants and items as the random effects in a mixed-effect model (Brysbaert & Stevens, 2018). At present, this approach could be implemented via the PANGEA website (Westfall, 2016) and `simr` package (Green & MacLeod, 2016). We generated the mixed-effect model for the above applications based on the raw data of 165 participants shared by Zwaan and Pecher (2012). The power estimated by PANGEA and `simr` were 0.13 and 0.29. Table 1 summarizes the estimated power as function of numbers of participants and items; Appendix 3 presents the details of the mixed-effect model. It shows that we have to increase both the number of the participants and the number of stimuli in the following study to reach an effect size equal to Zwaan and Pecher (2012).

Table 1

Achieved power estimated by PANGEA and simr.

		Estimated Power	
Number of Participants	Number of Items	PANGEA	simr
200	24	0.136	0.326
	48	0.208	0.525
	100	0.311	0.736
400	24	0.151	0.412
	48	0.247	0.647
	100	0.401	0.901
800	24	0.161	0.469
	48	0.274	0.748
	100	0.471	0.943
1200	24	0.165	0.480
	48	0.285	0.779
	100	0.500	0.965

Note. The settings of PANGEA is available at https://osf.io/mxnrb/

In consideration of the feasibility of this project among the laboratories, each participating laboratory will set up the sampling plan in line with the first approach. The participating laboratories will at first decide the minimal number of participants to be tested (between 100 and 160) in consideration of the available resources. Each laboratory will stop data collection when the number of participants has reached the pre-registered sample size.

In addition, the principal investigator will implement the Bayesian sequential analysis (Schönbrodt, Wagenmakers, Zehetleitner, & Perugini, 2017). Each laboratory could stop the data collection once the results reach the Bayes Factor beyond 10, even if the sample size has not reached the pre-registered number. For power analysis using the mixed-effect model, we strongly recommend that the data will be collected in at least eight different languages. There are 24 critical items in the material set of a language. Therefore the total number of items of 8 languages would be 196. The mixed-effect model in the inter-lab analysis will evaluate the effect size of the orientation effect when the languages are treated as a random effect.

Data Collection Plan

This project follows the protocols of multi-lab registered replication reports (e.g., Cheung et al., 2016; Hagger et al., 2016) and the workflow of Psychological Science Accelerator (see Moshontz et al., 2018, Figure 2). Until the submission of this proposal, there have been 40 participating laboratories from 14 languages (See registered laboratories and expected sample size in Figure 3 and Table 2). Each registered laboratory will upload their publicly accessible pre-registered plan on the OSF (see the form in Appendix 1). The principal investigators and the network committees will use the plans to guide the preparation of both stimuli and scripts.

Every participating laboratory has to pre-register the data collection plan and obtain the approval of the local research ethics board before the data collection. The project's OSF repository (Chen et al., 2018a) provided the standard instruction, according to outputs obtained from the pilot study (see Appendix 2), for the preparation of the experimental scripts and the coding scheme of filenames. The anonymized data of each participant will be managed by the laboratory members on the OSF repository.

Table 2

Summary of registered participating laboratories in the function of primary languages and sample size.

Language	Number of Laboratories	Registered Sample Size
English	23	1930
German	3	400
Greek	1	100
Hindi	1	200
Hungarian	1	200
Italian	1	50
Norwegian	1	100
Polish	1	100
Serbian	1	150
Slovak	2	300
Spanish	2	100
Thai	1	50
Traditional Chinese	1	100
Turkish	1	100

Note. Recruitment of the participating laboratories is going on while we submitted this proposal. Last updated: 2018/11/02

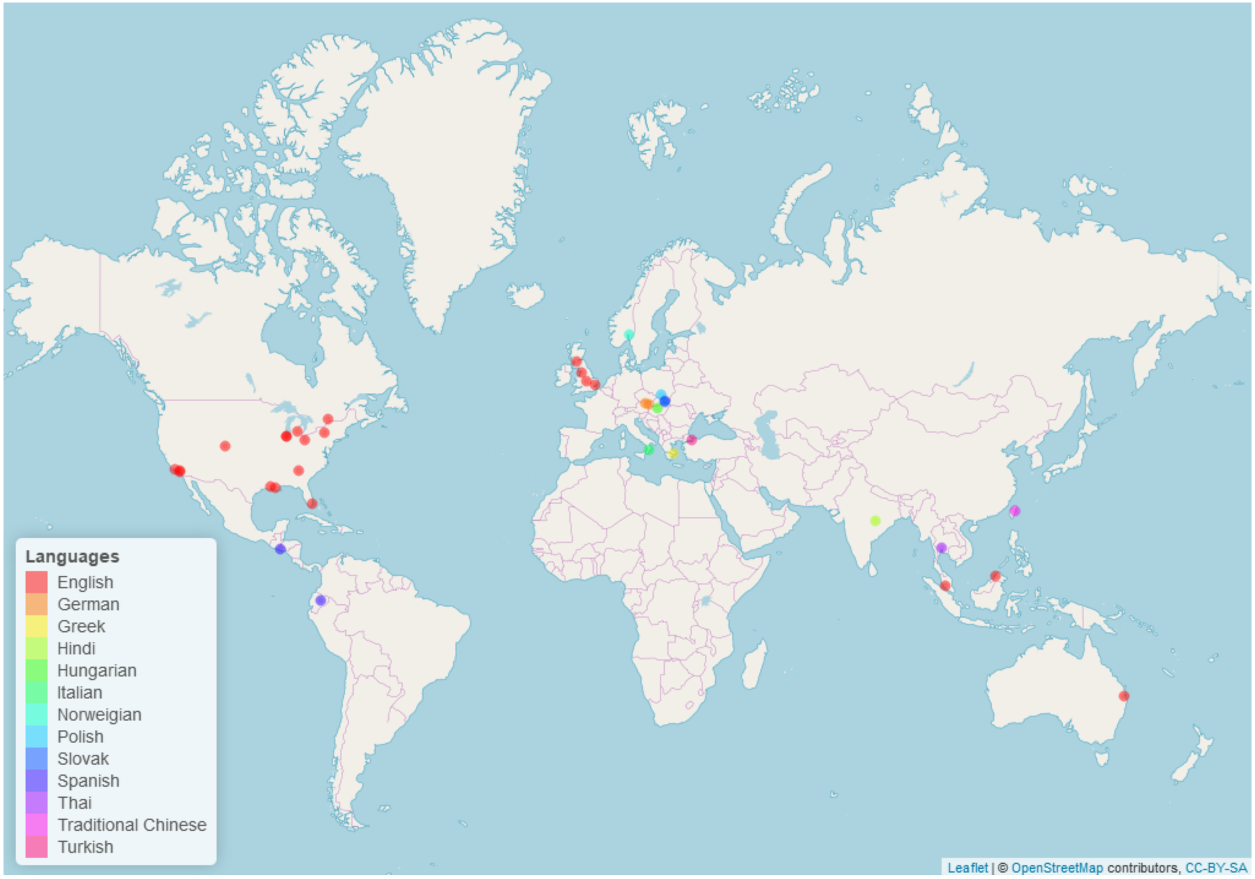

Figure 3. Distribution of registered laboratories and languages to be investigated.

Material

Zwaan and Pecher (2012) have shared the 24 critical object pictures, 48 filler pictures and the corresponding probe sentences in their OSF repository (https://osf.io/2dtrh/). The assignment of stimuli in each laboratory study followed the method of Zwaan and Pecher (2012). The original pictures came in two formats, hand drawings and camera photos. During the pilot study (see Appendix 2), colleagues from the collaborating laboratories suggested the unified format. There were four photographed objects presented in the original collection: *caterpillar*, *fire extinguisher*, *flashlight*, and *tire*. The photos of these objects were replaced by hand drawings before this submission. All the pictures are accessible in the project's OSF repository (Chen et al., 2018a).

Object pictures justification. We expect that at least two laboratories will contribute to the data collection in each language to be included in the main study. Along with the recommended 8 different languages, that amounts to at least 16 laboratories. For non-English languages, the participant laboratories will invite at least 10 volunteers to rate the object familiarity based on their native language knowledge. Every volunteer will rate the familiarity of 24 object pictures on a 7-point scale (1: not familiar - 7: very familiar). If a rater gives an object picture a familiarity score lower than 5, we will recommend they propose an object to replace the original one (Zwaan & Pecher, 2012). This object familiarity rating was conducted using *formr* survey platform (Arslan & Tata, 2018). So far the data are collected in six languages. The link to the survey, the survey scripts, and data summary are accessible in our project's OSF repository (Chen et al., 2018a).

Translation workflow. The translated materials for the participants of a particular language are conducted by at least four translators who are fluent in English and the local language. Every language coordinator must supervise the translators using the Psychological Science Accelerator guidelines (https://psysciacc.org/translation-process/). In addition, the coordinator and participating laboratories could make a decision at two optional parts. (1) According to the results of the object familiarity rating, the translators could isolate the items that are unfamiliar for a particular language. In this case, the translators who conducted the forward translation would suggest the probe sentences and object pictures to replace the isolated objects. The translators who conducted the backward translation would evaluate the suggested items. (2) Some objects in a particular language have different spellings or pronunciations among countries and geographical zones. For example, Americans say *tire*; British say *tyre*. Every coordinator would mark these objects in the final version of translated materials. Participating laboratories could replace the names in consideration of the local culture. In case of any replacement made then the laboratory plan has to be register. All the translated stimuli will be shared in the project's OSF repository (Chen et al., 2018a).

Analysis Plan

Intra-lab analysis during data collection

The intra-lab analysis will be used to assess the progress of data collection. According to the power analysis mentioned earlier, the Bayes Factor shows the moderate evidence value of the orientation effect ($BF_{10} = 21$). Therefore, every participating laboratory could stop the data collection at the BF_{10} of match advantage beyond 10. For each dataset from a participating laboratory, the median reaction times will be summarized in terms of the sentence-picture matching. Each laboratory, following the step-by-step instructions available at the project's OSF repository (Chen et al., 2018a), will upload the raw data to the OSF repository before data processing. At first we will excluded participants who made more than 30% of incorrect responses in the sentence-picture verification. Then we will conduct a Bayesian paired t-test on the participants' median response times across matching/mismatching object orientations for each laboratory. Along with the practices of Bayesian sequential analysis (Schönbrodt et al., 2017), a participating laboratory will collect the data until the preregistered sample size or the latest match advantage reaches the recommended Bayes Factor. Every laboratory has the flexibility to decide in which sample size they have to run the Bayesian sequential analysis. The statement of sequential analysis has to be made in the preregistered plan (see Appendix 1).

Inter-lab analysis after data collection

All the data included in the intra-lab analysis will enter the inter-lab analysis. Because the analytical unit is the participants' single response, we will at first use the shrinkage estimation method (see Baayen, 2008, sec. 7.3) to identify the outliers. In the data set of a laboratory, response times of matching and mismatching trials will construct the respective mixed-effect models in the function of the trial sequence. Each model includes the laboratory's slope (grand mean) and generates the participant's slope (by-participant means) respectively. If some participants' slopes are more or less 1 SD than their laboratory's slope, their response data will be excluded from the inter-lab analysis.

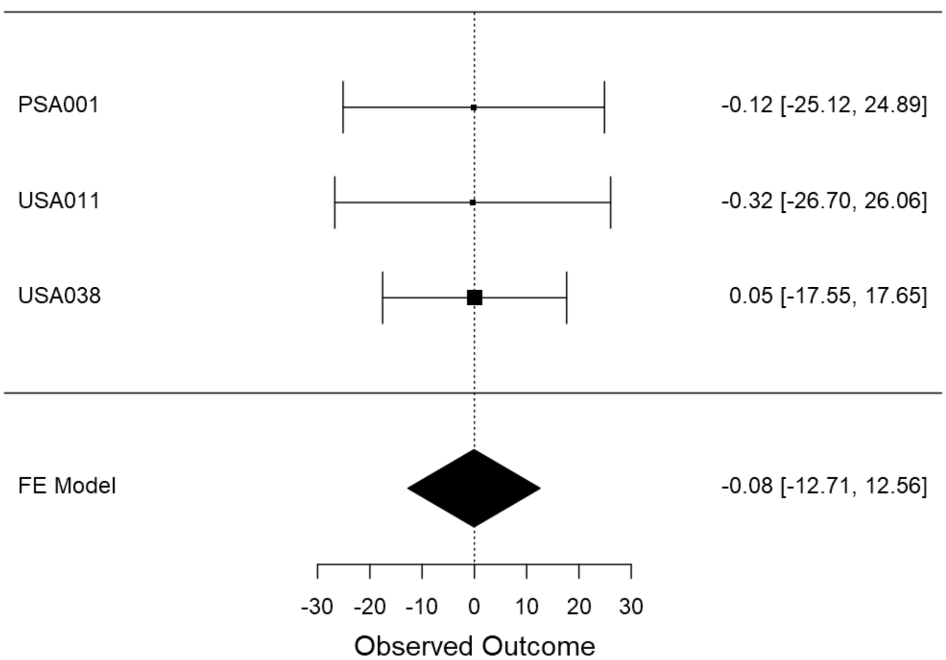

Figure 4. Forest plot of three pilot studies (see Appendix 2).

The data of sentence-picture verification across laboratories and languages will be analyzed using the meta-analytic techniques and mixed-effect model. Each analysis method will evaluate the match advantage in terms of the respective hypothesis. The meta-analysis will analyze the data obtained from the participating laboratories of a particular language. We will compute the fixed effect of each laboratory and estimate the overall effect size of match advantages. Using the results of pilot studies, we demonstrated a meta analysis to confirm the hypothesis of mental simulation (see Figure 4). The overall effect size could be far from zero if each particular language has the true effect of match advantage. In the mixed-effect model, we treat the laboratory and language as random effects and the match advantage of object orientations as a fixed effect. The random effect components will be decided by the suggested diagnostic tools (Barr, Levy, Scheepers, & Tily, 2013; Bates, Kliegl, Vasishth, & Baayen, 2015). When data is collected from more than eight languages, this mixed-effect model would show the robust estimation of orientation effect across the languages.

There are two analysis steps on the imagery scores to verify the hypotheses of mental rotation. The first step compares the imagery scores among the languages and the laboratories with the analysis of variance. The results will indicate the generalizability of mental rotation among languages. In the second step, the imagery scores and languages will be the predictor in the regression model that uses the match advantage as the dependent variable. This regression model will be the prediction model for the match advantage of object orientation in future studies.

Discussion

This project will achieve two goals after completion of the analysis of collected data. At first we will have the latest positive or negative evidence about the orientation effects being a general cognitive aspect among languages. A match advantage would support the embodied cognition studies that used the sentence-picture verification task as the paradigm in any particular language. On the other hand, a match disadvantage would serve as a warning for researchers who are planning to modify the sentence-picture verification for their studies. Our other contribution is to provide a platform to accumulate new data from the laboratories that will strictly follow this protocol to collect and analyze the data. With the continuous accumulation from more languages and laboratories, the mega empirical research framework, such as LeBel, McCarthy, Earp, Elson, and Vanpaemel (2018), will provide the robust information for who plan to measure the orientation effect in the circumstance unlike the original study.

References

Alibali, M. W., & Nathan, M. J. (2012). Embodiment in Mathematics Teaching and Learning: Evidence From Learners' and Teachers' Gestures. *Journal of the Learning Sciences*, *21*, 247–286. doi:10.1080/10508406.2011.611446

Arslan, R. C., & Tata, C. S. (2018). Rubenarslan/Formr.Org: New Ways To Expire Surveys. Zenodo. doi:10.5281/zenodo.1161181

Baayen, R. H. (2008). *Analyzing Linguistic Data: A practical introduction to statistics using R* (1st ed.). Cambridge University Press.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278. doi:10.1016/j.jml.2012.11.001

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

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

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

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

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, *1*(1), 1–20. doi:10.5334/joc.10

Chen, S.-C., Chartier, C. R., & Szabelska, A. (2018a). Object Orientation cross languages. *OSF*. https://osf.io/e428p/.

Chen, S.-C., Koning, B. de, & Zwaan, R. A. (2018b). Does Object Size Matter with Regard to the Mental Simulation of Object Orientation? *PsyArXiv*. doi:10.17605/OSF.IO/A3VWC

Cheung, I., Campbell, L., LeBel, E. P., Ackerman, R. A., Aykutoğlu, B., Bahník, š., … Yong, J. C. (2016). Registered Replication Report: Study 1 From Finkel, Rusbult, Kumashiro, & Hannon (2002). *Perspectives on Psychological Science*, *11*(5), 750–764. doi:10.1177/1745691616664694

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. doi:10.1037/a0013157

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

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

Connell, L., Lynott, D., & Dreyer, F. (2012). A Functional Role for Modality-Specific Perceptual Systems in Conceptual Representations. *PLOS ONE*, *7*(3), e33321. doi:10.1371/journal.pone.0033321

de Koning, B. B., Wassenburg, S. I., Bos, L. T., & Schoot, M. van der. (2017). Mental simulation of four visual object properties: Similarities and differences as assessed by the sentencePicture verification task. *Journal of Cognitive Psychology*, *29*(4), 420–432. doi:10.1080/20445911.2017.1281283

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

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149–1160. doi:10.3758/BRM.41.4.1149

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

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. doi:10.1111/2041-210X.12504

Hagger, M. S., Chatzisarantis, N. L. D., Alberts, H., Anggono, C. O., Batailler, C., Birt, A. R., … Zwienenberg, M. (2016). A Multilab Preregistered Replication of the Ego-Depletion Effect. *Perspectives on Psychological Science*, *11*(4), 546–573. doi:10.1177/1745691616652873

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). doi:10.22599/jesla.39

LeBel, E. P., McCarthy, R. J., Earp, B. D., Elson, M., & Vanpaemel, W. (2018). A Unified Framework to Quantify the Credibility of Scientific Findings. *Advances in Methods and Practices in Psychological Science*, *1*(3), 389–402. doi:10.1177/2515245918787489

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. doi:10.16719/j.cnki.1671-6981.20170105

Mahon, B. Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology, Paris*, *102*(1-3), 59–70. doi:10.1016/j.jphysparis.2008.03.004

Mathôt, S., Schreij, D., & Theeuwes, J. (2011). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. doi:10.3758/s13428-011-0168-7

Morey, R. D., & Rouder, J. N. (2011). Bayes factor approaches for testing interval null hypotheses. *Psychological Methods*, *16*(4), 406–419. doi:10.1037/a0024377

Morey, R. D., Rouder, J. N., Love, J., & Marwick, B. (2015). *BayesFactor: 0.9.12-2 CRAN*. Zenodo. doi:10.5281/zenodo.31202

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*. doi:10.1177/2515245918797607

Novack, M. A., & Goldin-Meadow, S. (2015). Learning from Gesture: How Our Hands Change Our Minds. *Educational Psychology Review*, *27*(3), 405–412. doi:10.1007/s10648-015-9325-3

Novack, M. A., & Goldin-Meadow, S. (2017). Gesture as representational action: A paper about function. *Psychonomic Bulletin & Review*, *24*(3), 652–665. doi:10.3758/s13423-016-1145-z

Phills, C., & Kekecs, Z. (n.d.). *Gendered Social Category Representations*.

Ping, R. M., Goldin-Meadow, S., & Beilock, S. L. (2014). Understanding gesture: Is the listener's motor system involved? *Journal of Experimental Psychology. General*, *143* (1), 195–204. doi:10.1037/a0032246

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*. doi: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*, 2218–2225. doi:10.1177/0956797613490746

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16* (2), 225–237. doi:10.3758/PBR.16.2.225

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. doi:10.1037/met0000061

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

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

Westfall, J. (2016). *PANGEA: Power ANalysis for GEneral Anova designs* (Working Paper).

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

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

Zwaan, R. A., & Pecher, D. (2012). Revisiting Mental Simulation in Language Comprehension: Six Replication Attempts. *PLoS ONE*, *7*, e51382. doi: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. doi: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. doi:10.1080/01638539309544832