ARTICLE

Comparisons between processing linguistic and mathematical negations from the perspective of the practice effect and working memory

Hikari Kinjo[1] [ID] and Toshiki Saito[2,3]

[1]Faculty of Psychology, Meiji Gakuin University, Tokyo, Japan; [2]Department of Psychology, University of Fribourg, Fribourg, Switzerland and [3]Japan Society for the Promotion of Science, Tokyo, Japan
Corresponding author: Hikari Kinjo; Email: kinjo@psy.meijigakuin.ac.jp

Abstract

A negated proposition can be expressed linguistically and mathematically. The current study examined the one-step and two-step procedure accounts from the perspective of the practice effect and working memory by comparing performance in two simple linguistic and mathematical verification tasks. Two online experiments were conducted with simple verification tasks over 10 practice sessions: a figure-equation task (e.g., ● ≠ ▲) and a figure-sentence task (e.g., ● is not ▲). Although reaction times in the equation task were faster than in the sentence task, both tasks showed that reaction times in negations took longer than those in affirmations regardless of the sameness of the figures in the target propositions (i.e., TA < FN and FA < TN) in both experiments, and the trend was not changed by the practice. The similar trends across the tasks, regardless of the practice, support the two-step procedure account, in which participants first evaluate the positive argument of negation and then reverse the response in negative propositions. Furthermore, high correlations between performance in the tasks and both verbal and spatial working memory tasks suggest that verification judgments may involve not only language processing but also more general cognitive processing.

Keywords: cognitive processing; negation; practice effect; true-false verification judgment; working memory

1. Introduction

Negation has been widely recognized as a key theme that approaches not only language but also the entire cognitive system, or human intelligence (Horn, 2001). However, the cognitive process of comprehending negated propositions has not been elucidated sufficiently. The process has been often explored within the experimental paradigm of sentence-picture verification and sentence verification tasks. Both paradigms, however, have drawbacks in that they demand more cognitive processes

in addition to negation. The sentence-picture verification tasks include cognitive processes that encode both the target sentence and the picture, compare them, and judge their congruency. The sentence verification tasks require retrieving the meaning of target words or sentences and common knowledge (e.g., "not guilty" is equal to "innocent" in Mayo et al. (2004); 24 is not an odd number in Wason (1961). Recently, however, Vanek and Zhang (2023) pioneeringly used the mathematical equation task (▲≠■), which requires less cognitive processes that are not directly related to judging the propositions with negation than those in the previous tasks. However, their aim was to compare negation processes between the two language systems (i.e., a truth-based system vs. a polarity-based system), and we do not know exactly how we can extend the results in the mathematical task to those in the linguistic task under the same propositions (i.e., a figure-sentence task). In order to explore the basic mechanism of processing negations, the current study used the two simple verification tasks with combinations of three simple figures (●, ▲, and ■). Particularly, we aimed to explore how performance in the tasks can be explained by two accounts on processing negations (i.e., one-step and two-step accounts by Wang et al. (2021)), which will be discussed below. We also explored whether and how processing negations would be related to working memory and transformed through practice.

1.1 Two experimental paradigms and two hypothetical procedures on processing negations

Processing negations had been explored mainly in two experimental paradigms: sentence-picture verification (Carpenter & Just, 1975; Clark & Chase, 1972; Ferguson et al., 2008; Fischler et al., 1983; Kaup et al., 2006; Lüdtke et al., 2008; Wang et al., 2021) and sentence verification (Kinjo, Saito, Hamada, & Sakai, 2019; Mayo et al., 2004; Nieuwland & Kuperberg, 2008; Wason, 1961).

The seminal work on sentence-picture verification was conducted by Clark and Chase (1972). They asked participants to judge the congruency between a picture and a description of the picture and proposed a 'theory of sentence-picture comparison'. The theory assumes that there is a true/false flag for the relationship between the picture and the description, and that the flag keeps switching until the final judgment is made. The initial value of the true/false flag is "true", and a true affirmative (TA) can be answered without converting this flag, but in the cases of false affirmative (FA), false negative (FN), and true negative (TN), it is necessary to convert the representational relationship between the picture and the description from 'true' to 'false'. Furthermore, in the case of TN, the additional operation of converting to the 'true' flag is required. Carpenter and Just (1975) also used a picture-sentence verification task and proposed a 'constituent comparison model' that explains the difference in RTs among the true/false judgment types. In this model, the comparison process of internal representations of pictures and sentences increases incrementally in the order of FA, FN, and TN compared to TA so that both RTs and error rates increase linearly. The literature on the picture-sentence verification paradigm often found RTs in the true-false verification judgment types in the order of TA < FA < FN < TN, supporting the steplike nature of processing negations.

Another line of research in the picture-sentence verification paradigm has focused on elucidating the temporal properties of sentence processing. Kaup et al. (2006) proposed a "two-stage model" in which the processing of negation consists of two

steps. For example, in the sentence "The door is not open", participants have to mentally represent an "open door", excluding the negative in the first step. In the second step, they have to direct their attention to the representation "the door is not open (i.e., closed)," which is the negation-integrated representation, and dismiss the first representation. Other studies using event-related potentials (ERPs) also supported the two-stage model (Ferguson et al., 2008; Fischler et al., 1983; Lüdtke et al., 2008) in the experimental paradigm.

These studies of the picture-sentence verification paradigm often support an account of processing negations where the positive argument in any proposition is activated in the first place, and then incremental cognitive processes are necessary in verifying negated propositions. A similar account is also discussed in the rejection-based account by Tian et al. (2016) and the two-step procedure account by Wang et al. (2021), which will be described below. Yet other studies found different patterns from the above results, depending upon the experimental conditions manipulating various factors: timing of presenting a sentence and a picture, participants' visual abilities, and contexts suggested by probe questions (for review see Wang et al., 2021).

Another experimental paradigm to explore the cognitive process of negation is sentence verification (Kinjo et al., 2019; Mayo et al., 2004; Nieuwland & Kuperberg, 2008; Wason, 1961), in which RTs in affirmations (TA and FA) are usually faster than negations, but the patterns of verification judgments are not consistent among the literature. Wason (1961) used a single-sentence verification task in which affirmation and negation of a target sentence have a bipolar relationship (e.g., 24 is/is not an odd number) and its negation can be uniquely determined to be the only one state (e.g., odd vs. even number) and showed no difference in RTs for TN and FN (i.e., TA < FA < FN = TN). In Experiment 1 of Mayo et al. (2004), participants were presented with a character description of a person and asked to judge later whether a probe sentence showed that his/her behavior was congruent with the description. They found that faster responses (i.e., adjusted judgment latency) were facilitated in congruent probes (e.g., "Tom's clothes are folded neatly in his closet") after affirmative descriptions (e.g., "Tom is a tidy person") than in incongruent probes (e.g., "Tom forgets where he left his car keys"), whereas participants showed the opposite pattern after descriptions with negation (i.e., TA < FA < FN < TN). In their Experiment 2, types of character descriptions of the person were categorized by a bipolar condition (e.g., stupid/smart) and a unipolar condition (e.g., adventurous/not adventurous). They replicated the response pattern only for unipolar conditions but not for bipolar conditions (i.e., TA < FA < TN < FN).

Given these inconsistent results in the literature, Wang et al. (2021) tried to sort out the contradictory findings of the presence/absence of the interaction of the polarity (negative/affirmative) and the truth value (true/false) in a picture-sentence verification task. Based on the two perspectives by Tian et al. (2016), Wang et al. (2021) hypothesized two patterns of processing negations, which seem to cover most of the previous findings: one-step and two-step procedure accounts. In the two-step procedure account, they assumed that participants first evaluated the positive argument of negation [S] and then reversed the response in the negative sentence[not[S]], resulting in FN < TN. In the one-step procedure account, they assumed participants could process the states of affairs that make the negative sentence true[not[S]]and take a longer response time to falsify the state, resulting in TN < FN. In case of FN = TN, they assumed participants used both procedures. In accordance with their prediction, they found the results corresponding to each procedure in the different

contextual conditions. In line with Wangs' one-step account, Kaup et al. (2006) argued that the mental representations of negated propositions can be encapsulated as one representation: 'the door is not open' can be represented as 'the door is closed', and their results support both the one-step and two-step procedure accounts depending upon the stimuli. Similarly, the fusion model proposed by Mayo et al. (2004) accords with the one-step account in that the model assumes the core of a negated message and the negation marker are integrated into one meaningful unit. This study will discuss how negations are processed under Wangs' one-step and two-step procedure accounts.

What seems to be missing from the literature on processing negations in these experimental paradigms, however, is a perspective that these paradigms require cognitive processing other than negation. Studies on picture-sentence verification force participants to encode both a target picture and a sentence, compare them, and judge their congruency in addition to verifying negations. Similarly, studies in the sentence verification paradigm often force them to access their prerequisite mental representation, such as common knowledge, and to infer appropriate answers from given sentences in addition to verifying negated propositions. The extra cognitive load may veil the effect of processing negations.

Recently, beyond the two paradigms, novel experimental designs on negation processing have been developed (Orenes et al., 2014; Vanek et al., 2024; Vanek & Zhang, 2023). Among them, Vanek and Zhang (2023) pioneeringly used the mathematical equation task (▲≠■), which requires less additional cognitive processes that are not directly related to verifying negated propositions than those in the previous tasks. However, they did not directly examine the equation task against a linguistic task with the same proposition and we do not know whether their results in mathematical equations can be generalized to linguistic sentences.

Thus, in order to elucidate the very basic cognitive process in verifying negated propositions, this study used two tasks with simple propositional statements which require less previous knowledge, memorization, or inference to judge propositions as true or false: A figure-equation task to verify a proposition about three figures (●, ▲, ■) in a mathematical equation (e.g., ● ≠ ▲) and a figure-sentence task to verify a proposition about the figures in a single sentence (e.g., "● is not ▲"). To our knowledge, no study has directly examined processing negations in mathematical equations compared to sentences which share the same propositions with each other. Given the tasks, we wonder whether and how the results would be interpreted under the two accounts on processing negations by Wang et al. (2021).

1.2 *The practice effect on verification judgments*

Given these simple linguistic and mathematical verification tasks, we further aimed to explore the practice effect on verification judgments to see whether and how processing negations will be transformed over practice sessions. Namely, we wondered whether and how the differences in RTs among the true-false judgment types would be changed through repeated sessions. Regardless of the significance of the practice effect to test the linearity of processing negations (e.g., Carpenter & Just, 1975), only a few studies have examined the practice effects in verification judgments. Neubauer and Freudenthaler (1994) conducted an experiment where participants were given a sentence-picture matching task every 10 minutes for 9 hours. In the task,

a propositional sentence such as "The + is above the asterisk" was presented, followed by a picture with a plus mark and an asterisk, and participants were asked to judge whether the sentence and picture matched. The results showed that although RTs decreased with practice, there was no practice effect on the overall response rate. That is, RTs for correct responses took longer for negative sentences than for positive ones, more errors were found for negative sentences than for positive ones, and this trend did not change over practice. Carpenter and Just (1975) cited similar results in unpublished works (Singer, Chase, Young, and Clark (Note 1b); Young and Chase (Note 2b)), where the differences among judgment types and practice effects were not statistically examined. In both studies, RTs became slightly faster as the number of trials increased, but the response trend in the verification judgment types seemed to be consistent across days.

Thus, the second aim of this study was to investigate the practice effect among the simple propositional tasks: Participants were asked to participate in repetitions of 10 sessions over several days at their convenience so as to explore processing negations. Nieuwland and Kuperberg (2008) demonstrated difficulties associated with negation would decrease if negated propositions meet 'pragmatic licensing conditions'. We wonder whether participants can form one mental representation as a whole (e.g., ● ≠ ▲) with repeated experiences, not as representations of separate constituents of the two figures (e.g., ● and ▲) with relation to the equal or not-equal sign. If the difference in RTs between the affirmative/negative propositions would not be diminished by practice, we assume the robust linear steplike nature in verifying negations, i.e., the two-step procedure account (Tian et al., 2016; Vanek & Zhang, 2023; Wang et al., 2021). To examine the practice effect, we compared RTs at the first three sessions with those at the last three sessions.

1.3 Relationships between processing negations and working memory

Given the simple linguistic and mathematical verification tasks, the third aim of this study was to investigate whether and how working memory (WM) relates to the verification of negated propositions using verbal and nonverbal WM tasks, which has not been fully explored in literature. With regard to a verbal aspect of WM, however, early studies on WM provided an important insight into the relationship. Hitch and Baddeley (1976) suggested that the sentence verification task may involve the phonological loop and the central executive system of WM. Kyllonen and Christal (1990) conducted a study to investigate the relationship between reasoning ability and WM capacity, using multiple WM tasks selected based on Baddeley's (1968) WM model. They found high correlation coefficients between WM capacity and reasoning ability factors (rs = .80 to .90). For example, performance in the alphabet recoding task, which is one of the WM tasks designed to measure verbal WM capacity, was correlated with it in several reasoning tasks (rs = .15 to .45). These studies, however, did not examine the task with regard to processing negations. Considering that verification of negated propositions is a kind of reasoning task, it is possible to assume that verbal WM capacity would be associated with the verification. Therefore, this study used two verbal WM tasks: an N-back task and a revised version of the alphabet recording task, called a hiragana recording task.

The N-back task involves judging whether the stimulus presented N positions back in the sequence matches the current target stimulus. The task imposes a load

that gradually increases as a function of N, requiring continuous updating (i.e., encoding, maintenance, and retrieval) of target stimuli. Participants must compare and match sequentially presented stimuli, making it a WM task involving both storage and processing demands. Awh et al. (1996) reported that brain activities during the N-back task using letters indicated the involvement of articulatory rehearsal and phonological storage, indicating activity in the phonological loop. The difference in correlation coefficients between the types of verification judgements in the figure-equation and figure-sentence tasks and the N-back task will indicate which conditions in the tasks require more cognitive load than others.

The hiragana recoding task is a newly devised WM task based on the alphabet recoding task developed by Woltz (1988). In the original task, for example, a three-letter alphabetic stimulus (e.g., "G N B") is presented first on the screen, followed by an operator and a number on the next screen (e.g., "+2"), and then the participant types the letter shifted alphabetically by the number indicated by the operator. The correct answer in this example is "I P D". In the present study, we modified the task to convert alphabet letters to Japanese hiragana characters and to make it easier than the N-back task by showing the letter panel on the screen. Namely, participants can succeed in the task as long as they remember a sequence of hiragana characters and a numeric operator along with the on-screen letterboard cue, resulting in the task requires less mental operation than the N-back task, which requiring continuous updating (i.e., encoding, maintenance, and retrieval) of target stimuli on WM.

In addition to these two verbal WM tasks with different amounts of cognitive load, this study also included a mental hand rotation task as a nonverbal WM task. In linguistics, the affinity between negation and spatial cognition has been discussed. Arimitsu (2020) reviewed examples of "cognitive schemes" of various negation propositions in mental space and argued that affirmation and negation could be considered as a figure-ground relation. According to Arimitsu, the ability to understand the meaning of negation is related to the accumulation of concrete experiences of spatial cognition in daily life. Yet, little empirical evidence has been found to reveal the relationship between processing verifying negated propositions and spatial cognition (cf., Orenes et al., 2014; Vanek & Zhang, 2023). One study by Vanek and Zhang (2023) reported the possible relationship between processing negations and not spatial but visual sensory cognition under the embodied cognition theory (Barsalou, 1999). With the non-linguistic figure-equation task, their results suggested the two-step account: To understand ▲ ≠ ■, individuals build an iconic mental model of the corresponding affirmative ▲ = ■ and then integrate the negation symbol ≠. If true, this may suggest that processing negations is related to visual cognition. No study has reported the relationship between processing negations and spatial cognition.

Spatial cognition is not a single ability but is composed of at least three different abilities, including metal rotation, visualization, and spatial perception (Kaufman, 2007). Among these abilities, mental rotation is considered to be more strongly related to WM. For example, activation of ventral prefrontal cortex regions was related to WM during mental rotation tasks (Schendan & Stern, 2007). Based on a finding of the relationship between mental hand rotation tasks and WM (Tanaka & Yoshida, 2015), this study chose a mental hand rotation task as a nonverbal WM task. The task involves mental operations of holding the pictures of the hand presented on the screen and matching them with the actual image of a right or left hand. Because the mental hand rotation task and the N-back task involve mental operations of

holding and matching the target with the stimuli presented on the screen one after another, we assume they need more cognitive load than the hiragana recording task. Due to the limited duration of experiments, this study included only one nonverbal task.

If processing negations in either verification task or both tasks were related to only the verbal cognitive process, then we would observe positive correlations between the performance of the verification judgments and verbal WM tasks. If they were related to not only verbal but also nonverbal cognitive processing in either task or both tasks, then we would observe positive correlations between the performance in the verification judgments and these WM tasks. The details of each WM task will be described in Method.

1.4 Hypotheses of the current study

The current study examined the one-step and two-step procedure accounts from the perspective of the practice effect and WM by comparing performance in the two simple verification tasks (i.e., the figure-equation and figure-sentence tasks). Our first inquiry was whether processing negations in mathematical equations is operated similarly to that in linguistic sentences, which share the same propositions with each other; that has not been directly explored. If so, then the trends in the differences in RTs in the true/false verification judgments would be similar between the tasks. Based on the results in Experiment 1 in Vanek and Zhang (2023), where participants were asked to judge to agree or disagree with the figure-equations as fast and accurately as possible, we predicted the two-step process in the task where negations have longer RTs than affirmations (TA < FN (e.g., ▲ = ▲ < ▲ ≠ ▲) and FA < TN (e.g., ▲ = ■ < ▲ ≠ ■)) (Hypothesis 1). This is because the two-step procedure assumes that to understand negated equations (e.g., ▲≠■), participants first build a mental representation of affirmative ▲ = ■ and then integrate the negation symbol ≠. Given the similarity of processing negations between the two tasks, the same logic is held in the hypothesis of the figure-sentence task. In order to understand results straightforwardly in the mixed-effect regression models as in Vanek and Zhang (2023), the four judgment types were categorized based on two factors, polarity (affirmative/negative) and sameness of two shapes in the task (same/different): TA is an affirmative-same judgment. FA is an affirmative-different condition. FN is a negative-same judgment. TN is a negative-different judgment. Under this categorization, Hypothesis 1 predicted the significant fixed effect on polarity where RTs in negations take longer than those in affirmations regardless of the sameness conditions or tasks.

We also wondered whether negation processing will be transformed from the two-step to the one-step process through practice, namely whether a mental representation as a whole, including the negation symbol (e.g., ▲≠▲ or ▲≠■), could be built through practice in either task or both tasks. If not, the tendency that negations have longer RTs than affirmations (i.e., TA < FN (e.g., ▲ = ▲ < ▲≠▲) and FA < TN (e.g., ▲ = ■< ▲ ≠ ■)) will be unchanged in both tasks. In other words, there is a significant difference in polarity not only at the first sessions but also in the late sessions (Hypothesis 2). If processing negations requires high cognitive load in verbal WM, we predict significant correlations between performance in the verification tasks and the verbal WM tasks. If the correlation coefficients were also

significant with the spatial WM task, then we assume processing negations would be related to a cognitive process underlying both verbal and non-verbal WM (Hypothesis 3).

2. Methods

Two experiments were conducted to investigate the above three hypotheses. The procedures of these experiments were almost the same except for their instruction and feedback. In Experiment 1, both speed and accuracy were emphasized at the beginning of each task, and feedback of correct/incorrect and RT (ms) for a participant's response was presented at the end of each trial. Experiment 2 was conducted to confirm the extent to which performance was improved by changing the instruction, where only accuracy was emphasized at the beginning of each task and only feedback of correct/incorrect was presented at the end of each trial. The change was necessary because one may argue that the results of Experiment 1 may have been affected by a trade-off between speed and accuracy. For example, Baddeley and Hitch (1974) reported that participants' accuracy improved with the cost of RT when they were instructed to focus only on accuracy. In fact, a correlation test between accuracies and RTs of all trials for Experiments 1 and 2 were rs = .110 ($p < .001$) and .014 ($p = .15$), suggesting there was indeed some speed-accuracy trade-off in Experiment 1 and the effect was washed out by the instruction to weight more accuracy over RT. Except for these changes, the procedures of the two experiments were almost same and they are described collectively below.

2.1 Participants

Twenty undergraduate and graduate students completed Experiment 1 of 24 participants. Data from one participant with poor accuracy (i.e., less than 2.5 SD from the general mean accuracy for all participants on both tasks) were excluded, resulting in 19 participants (15 females, age range = 19-26, mean age = 21.05 (SD = 1.64)). Another 22 undergraduate and graduate students completed Experiment 2 of 24 participants. Data from one participant with poor performance (i.e., less than 2.5 SD from the general mean accuracy for all participants in the figure-equation task) were excluded, resulting in 21 participants (17 females, age range = 19-25, mean age = 21.57 (SD = 1.62)). All participants were monolingual Japanese. They were asked to perform Short Form Edinburgh Handedness Inventory (Veale, 2013), which resulted in 17 righties and 2 lefties in Experiment 1 and 18 righties, 1 lefty, and 2 ambidextrous in Experiment 2. Because the handedness did not change the conclusions, all of their data were included.

2.2 Materials

2.2.1 Two verification tasks with figures
As described above, this study used the two simple verification tasks with three figures (●, ▲, ■); a figure-equation task and a figure-sentence task. Each task consisted of four types of true-false judgments. For example, for a figure-sentence task, "● is ●" (TA), "● is not ▲" (TN), "● is ■""(FA), and "● is not ●" (FN). Each task consisted of four practice trials and 24 main trials consisting of six trials in each type

of judgment: one of nine possible pairs of the three figures was inserted into the left and right positions in a sentence/mathematical formula with and without negation, resulting in $3 \times 3 \times 2 = 18$ trials. To make the number of true/false judgments even, three more pairs of each of the same figures with and without negation were added for FN and TA judgments, resulting in six more trials for TA and FN judgment types. The accuracy was calculated for each judgment type in each task.

Notice that this study was conducted in Japanese and used a negation 'ない' in the sentence task. The position of the negation 'ない' comes at the end of a sentence in Japanese, but its function and meaning are the same as 'not' in English. For example, the mathematical equation, ● = ▲, is equal to the sentence '●は▲である', where 'は' indicates that ● is the topic of the sentence and 'である' is a formal and declarative copula that asserts the equivalence of ● as ▲. The mathematical equation, ● ≠ ▲, is equal to the sentence '●は▲ではない', where 'ではない' is the negation of the copula 'である', which negates the state or equivalence expressed by ▲.

2.2.2 Three WM tasks

This study used the two verbal WM tasks (i.e., the N-back task and the hiragana recoding task) and one non-verbal WM task (i.e., the mental hand rotation task). The N-back task with Japanese hiragana characters as stimuli asked participants to judge whether the stimulus presented at N positions back matches the current stimulus, where one of four hiragana characters with the same vowel sound ("あ", "か", "さ", and "た") was presented one character at a time. The task comprised three levels: 0-back, 1-back, and 2-back, each consisting of 25 trials. The order of presentation of the hiragana characters was predetermined to ensure the equal frequency of appearance for each task, resulting in an equal number of matching judgments. The accuracy rate was calculated by combining scores of the correct trials in all of the N-back trials.

The hiragana recoding task is a newly devised WM task based on the alphabet recoding task developed by Woltz (1988). In this task, stimuli were a sequential hiragana character set ranging from "あ" to "と", resulting in 20 characters. To reduce the processing load and prevent input errors, a hiragana character panel was displayed on the screen, organized in 4 rows × 5 columns (columns representing each hiragana vowel-consonant groups from left to right, "あ", "か", "さ", and "た"). Participants were required to click on the character panel to provide their answers. Arithmetic operations ranged from −2 to +2, and the set size was raised incrementally from 3 to 4 to 5. When the operator was +0, it became a simple sequence reproduction task. As the absolute value of the arithmetic number increased, the positions to shift the characters also increased, resulting in a higher processing load. However, participants could refer to the character panel while answering if they memorized the initially presented hiragana string and the operator, mitigating the processing load. At each trial, the hiragana string was first displayed for 1500 ms. Subsequently, a number representing the amount to shift (−2 to +2) was displayed for 1000 ms. Participants had to shift the presented hiragana characters forward or backward by the indicated number. For example, if "あ, き, つ" were followed by "+2," the correct answer would be "う, け, と". The presented character strings were created so that they did not form meaningful words, each hiragana's frequency was balanced, and character shifts were designed

not to extend the range of the character set. In each character set size, there were three trials for each of the five movement distances (i.e., from −2 to +2), totaling 45 trials. Scores for correct answers were weighted according to the set sizes. The accuracy rate was calculated by aggregating the weighted correct scores. Our preliminary experiments showed that the accuracy rate in this task was higher than that in the N-back task.

The mental hand rotation task involves judging whether the presented image of a hand is the left or right hand (Hida & Sekiyama, 2013). The images of hands were colored illustrations of the palm and back of the left and right hands (272 × 400 pixels). In this task, images were presented rotated at 0°, 90°, 180°, and 270°. Participants were instructed to press the buttons by judging the image right or left as accurately and quickly as possible. The task consisted of 8 practice trials and 16 main trials. Scoring involved calculating the proportion of correct responses (p(c)), taking into account reaction bias (p (c) = 1/2 (HIT rate + (1 − FA rate))) as suggested by Hautus et al. (2021). Here, the HIT rate refers to the percentage of right-hand trials correctly identified as right, and the FA rate represents the percentage of left-hand trials incorrectly identified as right.

In addition to these tasks, this study included the baseline tasks aimed to examine the stability of participants' responses to stimuli in online experiments: a perception task, an identification task, and a matching task. In the perception task, participants responded when the figure ● appeared in the center of the screen. In the identification task, participants responded when the figure ● appeared in the center of the screen among three figures: ▲, ●, and ■. In the matching task, participants judged whether the figures to the left and right of the screen were the same or different. Because no issues were identified regarding the stability of responses, the results will not be discussed in this paper.

2.3 Procedure

Two experiments were conducted online, and each experiment had 10 sessions. Each session had a pair of test batteries, Set A and Set B, considering the participants' cognitive load. Set A consisted of three baseline tasks and the two verification tasks. Set B consisted of the three WM tasks and another verification task (i.e., the general-knowledge sentence task), which will not be discussed in this paper due to space limitations. Participants took part in the experiment at their convenience, with a maximum of three sessions per day within the span of two weeks in Experiments 1 and 2: Median = 5.0 and 5.0 days; Mean = 5.35 and 5.95 days; Range = 4–10 and 4–15 days. Each test battery lasted about 10–15 minutes. Breaks were allowed between sets and sessions. The tasks were created using lab.js (Henninger, Shevchenko, Mertens, Kieslich, & Hilbig, 2020). Participants who completed the experiment were rewarded with an Amazon gift card worth 3000 Japanese yen as compensation. This study was approved by the ethics review board of the author's affiliated institution. Because this study intended to examine how verification judgments would be improved through repetition, there was no filler items.

The test battery Set A consisted of five tasks. The first three tasks were designed to measure the basic response baseline in online experiments. They were followed by two verification tasks: a figure-equation task and a figure-sentence task. Participants were instructed to respond to the true/false questions using the F key for "true" and

the J key for "false" as accurately and quickly as possible (Experiment 1) and as accurately as possible (Experiment 2). The order of all trials within the tasks was randomized, and the sequence of the verification tasks was counterbalanced. Because no issues were identified regarding the stability of responses, the results of the baseline tasks will not be discussed in this paper.

The test battery Set B consisted of four tasks: the N-back task, the hiragana recoding task, the mental hand rotation task, and the general-knowledge-sentence task. The order of the tasks was randomized, and the trials in the general-knowledge-sentence task and hand mental rotation tasks were randomized. Considering the easiness of performing the tasks, within the N-back task, the sequence of subtasks was consistent in the order of 0-back, 1-back, and 2-back. Similarly, within the hiragana recording task, the sequence of subtasks was consistent in the order of the set size 3, 4, and 5 characters. In the 1-back and 2-back of the N-back task and the hiragana recording task, the order of presentation could not be randomized due to the nature of the tasks. To compensate for the restriction, multiple patterns of orders in stimulus presentation were predetermined to ensure that the same pattern did not repeat consecutively across sessions.

Experiment 1 emphasized speed and accuracy along with feedback of correct or incorrect and RT (ms) for the participant's response at the end of each trial. In Experiment 2, the focus was placed on accuracy, and only feedback of correct or incorrect was given.

3. Results

For analyzing mixed-effect regression models, we used the lme4 package (Bates et al., 2015) in R (Version 4.1.1; R Development Core Team, 2021) based on the open access code in R by Vanek and Zhang (2023). For other analyses, we used SPSS 29.0 for our convenience. The Bonferroni method was used for multiple comparisons. The significance level for all statistical analyses was set at .01. The codes from two experiments in R are available at https://osf.io/jvp3w.

The means and *SD* of error rates per participant were .09 (.06) in the figure-equation task and .11 (.05) in the figure-sentence task in Experiment 1 and .05 (.04) and .04 (.04) in Experiment 2. RTs of correct responses were used for the analysis. Outliers of RTs exceeding the upper limit value of mean + 2.5 *SD* of the RTs for each condition for each participant were replaced by the upper limit value. As a results, 1.7% and 1.8% of trials in the figure-equation task and 0.5% and 4.3% of trials in the figure-sentence task were replaced to the value in Experiments 1 and 2, respectively.

Figure 1 plots the trend of RTs across sessions per verification judgment type in Experiments 1 and 2. The figure shows the decrease in RTs gets closer to the asymptote at the last sessions in both experiments. First, to examine the trend of RTs across sessions, a three-factor within-subject design of analysis of variances was conducted for RTs for each experiment. The three factors were the number of practices (10 sessions), the verification tasks (figure-equation and figure-sentence tasks), and the four judgments types (TA, FA, TN, and FN). The results in Experiment 1 showed the main effect of the number of practice sessions ($F(4.01, 68.11) = 18.21, p < .001$, partial $\eta^2 = .52$) with the interaction with the judgment types ($F(27, 459) = 3.10, p < .001$, partial $\eta^2 = .15$). A simple main effect test of the number of practices showed RTs in the 10th session were
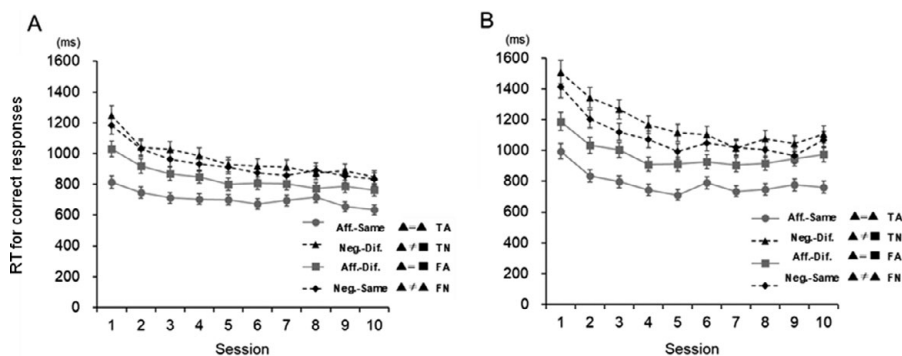
Figure 1. Reaction times in the four verification judgments across 10 sessions aggregated across the two verification tasks in Experiments 1 (A) & 2 (B).
Note: Error bars indicate 95%CI.

faster than RTs in the first three sessions for FA, TN, and FN, and faster than RTs in the first two sessions for TA. The results in Experiment 2 showed a similar trend: the main effect of the number of practice sessions ($F(9, 180) = 14.42$, $p < .001$, partial $\eta^2 = .42$) with the multiple comparisons showing that RTs were slower in the first and second sessions than in the other sessions, but there were no differences in RTs thereafter. The interaction between the number of practices and tasks and the three-way interaction were not significant ($F(4.99, 99.78) = 1.67$, $p = .04$, partial $\eta^2 = .08$; $F(54, 1080) = 1.06$, $p = .36$, partial $\eta^2 = .05$). Given these findings, in order to have the following analyses simpler and more evident, we grouped the first and last three sessions together, which is a common technique to examine the practice effect (Simon et al., 2016). Figure 2 shows RTs in the four judgment types aggregated in the first and last three sessions in the two verification tasks in Experiments 1 and 2.

To examine Hypotheses 1 and 2 together, we built mixed-effect regression models using the lme4 package (Bates et al., 2015) in R (Version 4.1.1; R Development Core Team, 2021) based on the open access code in R by Vanek and Zhang (2023). Our fixed effect factors were Polarity (affirmative/negative), Sameness (same/different), Task (equation/sentence), and Practice (first three sessions/last three sessions), and the random effect factors were Participant and Stimulus items.

3.1 Trends in the difference in RTs in the judgment types between the tasks and their practice effects

Table 1 showed the mixed effect models with all possible random effects in Experiments 1 and 2. The two experiments demonstrated quite similar patterns. In both experiments, all fixed effect factors were significant: RTs were longer in negations than in affirmations, in the judgment types with different figures than in those with the same figures, in the first three sessions than in the last three sessions, and in the figure-sentence task than in the figure-equation task. The polarity and sameness factors showed interactions in that the difference in RTs between the negative and affirmative conditions were larger in the same figure condition than those in the different figure condition, although negations always took longer than affirmations in both experiments. Namely, mean RT and SD in the negations was
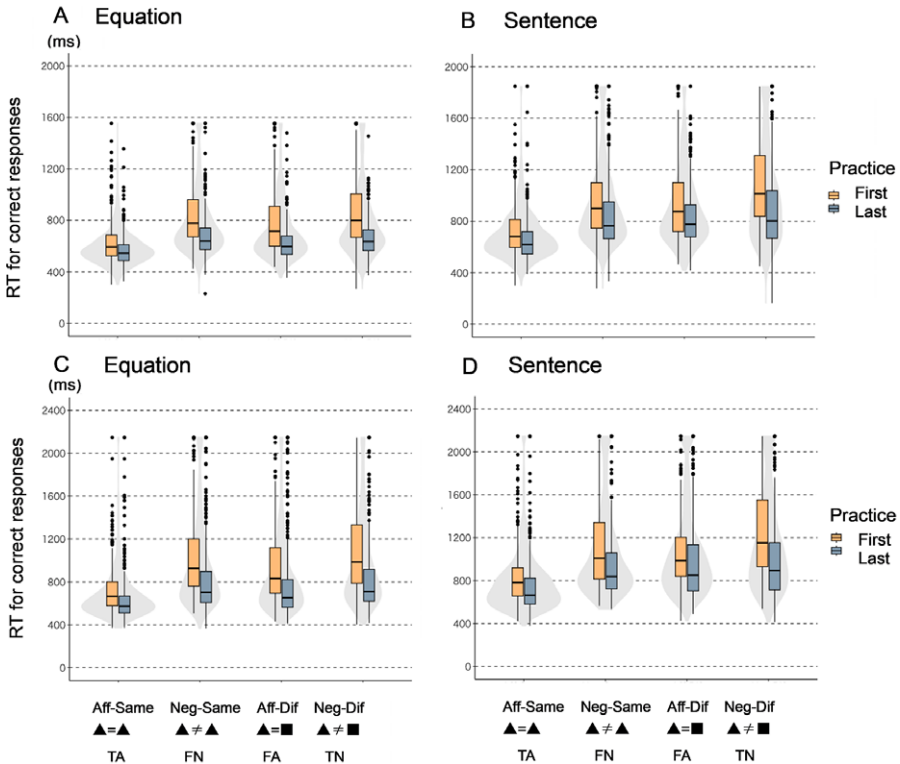
Figure 2. Reaction times in the four verification judgments of the two verification tasks aggregated in the first and last three sessions in Experiments 1 (A, B) & 2 (C, D).
Note: For comparison, the horizontal labels for the two tasks are the same.

993 ms (46) and those in the affirmations was 734 ms (24) at the same figure condition, and 1061 ms (46) and 955 ms (42) at the different figure condition in Experiment 1. A similar pattern was found in Experiment 2: 842 ms (23) and 645 ms (17) at the same figure condition, and 883 ms (29) and 797 ms (20) at the different figure condition. The practice factor interacted with the polarity factor in both experiments: In Experiment 1, the difference in mean RTs between the affirmations and negations at the last three sessions was smaller than that at the first three sessions (-176 and -107 ms), but the trend that RTs in negations took longer than in affirmations stayed significant even after practice. A similar significant trend of the differences was found in Experiment 2 (-248 and -118 ms). No second and third-order interactions between the polarity, sameness, practice, and task factors suggest that the trend between the polarity and sameness factors did not change regardless of the practice and task factors. Thus, these results clearly suggest both tasks support the two-step account because RTs in negations took longer than those in affirmations regardless of the sameness (i.e., TA < FN and FA < TN) in both experiments, and the trend was not changed by the practice.

Table 1. Coefficients for a mixed effects model fitted to the RTs in the verification judgments in Experiments 1 & 2

Fixed effects	Experiment 1				Experiment 2			
	Estimate	SE	t value	p	Estimate	SE	t value	p
Intercept	629.71	21.57	29.19	<0.001***	718.45	28.87	24.89	<0.001***
Polarity (neg)	216.52	20.51	10.56	<0.001***	330.84	37.60	8.80	<0.001***
Sameness (dif)	156.39	24.34	6.43	<0.001***	221.02	34.93	6.33	<0.001***
Practice (last)	−67.85	17.93	−3.79	<0.001***	−81.49	23.42	−3.48	<0.001***
Task (sentence)	95.74	24.31	3.94	<0.001***	110.78	26.50	4.18	<0.001***
Polarity (neg) × Sameness (dif)	−124.89	29.01	−4.31	<0.001***	−165.28	44.39	−3.72	<0.001***
Polarity (neg) × Practice (last)	−97.63	25.67	−3.80	<0.001***	−133.87	33.40	−4.01	<0.001***
Sameness (dif) × Practice (last)	−93.66	25.37	−3.69	<0.001***	−80.22	33.23	−2.42	0.02*
Polarity (neg) × Task (sentence)	33.51	25.81	1.30	0.20	−33.64	34.25	−0.98	0.33
Sameness (dif) × Task (sentence)	70.97	32.48	2.19	0.03*	34.51	36.54	0.94	0.35
Practice (last) × Task (sentence)	4.21	25.42	0.17	0.87	1.97	33.26	0.06	0.96
Polarity (neg) × Sameness (dif) × Practice (last)	45.00	36.81	1.22	0.22	21.63	47.59	0.45	0.65
Polarity (neg) × Sameness (dif) × Task (sentence)	18.96	37.13	0.51	0.61	64.84	48.50	1.34	0.18
Polarity (neg) × Practice (last) × Task (sentence)	49.70	36.43	1.36	0.17	50.17	47.48	1.06	0.29
Sameness (dif) × Practice (last) × Task (sentence)	29.12	36.26	0.80	0.42	93.11	47.18	1.97	0.005*
Polarity (neg) × Sameness (same) × Practice (last) × Task (sentence)	−73.96	52.49	−1.41	0.18	−126.05	67.50	−1.87	0.06

Random effects	Variance	SD		Variance	SD			
Participants	3235.5	56.88		10165.2	100.82			
Item	412.21	20.30		230.57	15.18			

*p < .05,
***p < .001.

3.2 *Correlations between performance of the two verification tasks and the three WM tasks*

To examine Hypothesis 3, we conducted correlation analyses between accuracy for each judgment type in the two verification tasks in Experiments 1 and 2 and the percentage of correct responses in the WM tasks (Table 2).

Before conducting the analysis, the practice effect was examined for performance in the three WM tasks by performing a one-factor ANOVA for performance of each of the three WM tasks in each experiment as a function of the number of practice sessions. Results showed that there was no practice effect: in Experiment 1, the N-back task ($F(9,171) = 1.35$, $p = .22$, partial $\eta^2 = .07$), the mental hand rotation task ($F(9,171) = .42$, $p = .93$, partial $\eta^2 = .02$)); in Experiment 2, the hiragana recording task ($F(2.76, 81.36) = 2.76$, $p = .04$, partial $\eta^2 = .12$), N-back task ($F(9,189) = .79$, $p = .62$, partial $\eta^2 = .04$), and the mental hand rotation task ($F(2.67, 55.98) = 2.09$, $p = .12$, partial $\eta^2 = .09$). A main effect was found only for the hiragana recording task in Experiment 1 ($F(4.09, 73.58) = 7.72$, $p < .001$, partial $\eta^2 = .30$), with the multiple comparisons revealing that performance was lower on the first session than on the sixth and subsequent sessions, with no significant difference in performance on the fourth and subsequent sessions. Because there was almost no significant practice effect for each WM task, and data with a large number of observation points "…can more generally and more accurately determine the presence or absence of experimental effects than data with a small number of observations (Usami & Soujima, 2015, p. 4)", we calculated the mean of each WM task performance across the 10 sessions for each experiment.

Given the same logic, because the main effect of the number of practice sessions was not significant in accuracy in the verification tasks in both experiments in the analyses of variance (i.e., $F(4.64, 74.19) = 344$, $p = .87$, partial $\eta^2 = .02$; $F(9, 180) = 1.40$, $p = .19$, partial $\eta^2 = .07$), we aggregated the mean of accuracies for each judgment types for each task for each experiment.

Table 2 shows the results of the correlation coefficients between accuracy for each WM task and each of the four jument types in the two verification tasks in both experiments. As a whole, the correlation coefficients for the hiragana recording task were lower than those for the other two cognitively demanding WM tasks, and those for TA were lower than for the other judgment types in both verification tasks in both experiments. The correlation coefficients in the FN judgments were higher than in the TA judgments in the cognitively demanding WM tasks in both verification tasks in both experiments. Although those in the TN judgments were higher than in the FA judgments in Experiment 1, those in the FA judgments increased in Experiment 2 in both verification tasks. We confirmed with scatter plots that the high correlation coefficients were not due to outliers.

4. Discussion

4.1 *Processing linguistic and mathematical negations and their practice effect*

This study used the two simple verification tasks, the figure-sentence and figure-equation tasks, which require perceptual information in the very target statements and mathematical equations without reference to prior knowledge or extra memorization. The results showed that RTs were faster in the figure-equation task than in the figure-sentence task in both experiments. The differences in the tasks suggest

Table 2. Correlation coefficients between accuracy of the verification judgment types in the two verification tasks and the three working memory tasks aggregated across the 10 practice sessions

		Equation				Sentence				Hiragana recording	N-back	Mental rotation of hands
		Affirmative-same ▲ = ▲ TA	Negatice-same ▲ ≠ ▲ FN	Affirmative-different ▲ = ■ FA	Negative-different ▲ ≠ ■ TN	Affirmative-same ▲ = ▲ TA	Negatice-same ▲ ≠ ▲ FN	Affirmative-different ▲ = ■ FA	Negative-different ▲ ≠ ■ TN			
Exp1	Hiragana recording	.092	−.191	−.191	.336	−.041	.018	.321	.283	−	.361	.319
	N-back	−.080	.360	.447	.766***	−.141	.545*	.074	.705***		−	.511*
	Mental hand rotation	.498*	.489*	.318	.600**	−.005	.474*	.150	.711***			−
Exp2	Hiragana recording	−.183	.378	.285	.366	−.039	.091	.431	.446*	−	.456*	.303
	N-back	.017	.446*	.706***	.580**	.134	.466*	.468*	.584**		−	.720**
	Mental hand rotation	.180	.408	.658**	.613**	.400	.432*	.585**	.543**			−

$*p < .05,$
$**p < .01,$
$***p < .001.$

the true/false judgments of mathematical equations were processed faster than those of linguistic statements even though their propositional logic was supposed to be the same. Yet, the patterns in the verification judgments were quite similar between the tasks.

Firstly, we wonder why the mathematical equations were processed faster than the linguistic statements. We speculate that letter counts in the sentence task were larger than those in the equation task, which may hinder processing efficiency in the sentence task compared to the equation task. Building on it, we may tend to perceive mathematical symbols as cohesive units rather than as isolated operators (e.g., ≠ vs. not equal to). To test the speculation, we can compare ● ≠ ▲ to "● unequal ▲", which is a logical step to investigate processing linguistic and mathematical negations for a future study in the current experimental paradigm.

The current two experiments demonstrated no practice effect on RTs in both verification tasks: Although speed in RTs accelerated to a certain level in the two tasks over 10 sessions, the trends of the difference in RTs between affirmations and negations among the verification judgments showed the same pattern in both tasks of both experiments at both the first and last sessions: TA < FN and FA < TN. Thus, the results support the two-step process in both tasks where negations have longer RTs than affirmations (TA < FN (e.g., ▲ = ▲ < ▲ ≠ ▲) and FA < TN (e.g., ▲ = ■ < ▲ ≠ ■) regardless of the difference in the form of representation (i.e., equations or sentences), the amount of practice, and the instructions. Therefore, Hypotheses 1 and 2 were supported. The current results replicated the findings with the mathematical equation (Vanek & Zhang, 2023) and further demonstrated a similar process in negations would be operated in the linguistic task with the same propositions.

4.2 Relationship between verification judgments and WM

In both verification tasks, significant correlation coefficients were consistently found in the mental hand rotation and N-back tasks but not in the hiragana recoding task across both experiments. One of the key points is the difference in cognitive load between the hiragana recording task and the other two WM tasks. Our preliminary experiment showed that the cognitive load for the hiragana recoding task was not as heavy as that for the other tasks. This is because in the hiragana recoding task, participants could use the hiragana character panel on the screen as a cue to solve problems as long as they remembered the probe strings of hiragana characters and the probe digit of the operator to shift the strings. The N-back task, on the other hand, involves the mental operation of holding and matching the target with stimuli presented on the screen one after another. Similarly, the hand rotation task involves the mental operation of matching a target picture of the hand presented on the screen with the actual image of a right or left hand. For the latter two WM tasks, the mentally holding and matching process must be required at each trial. The high correlation coefficients between the two tasks support this speculation (rs = .511 and .720). In both tasks of both experiments, negations (FN and TN) showed higher correlations than affirmations (TA and FA) with the two cognitively demanding tasks, suggesting negations involve holding and matching processing of mental representations even in the simple linguistic and mathematical propositional tasks. The correlation

coefficients in the FA judgments in Experiment 2 became higher than those in Experiment 1, indicating changing instructions to weigh more in accuracy may forth the judgement more cognitively demanding. In short, Hypothesis 3 was supported in that processing negations would be related to a cognitive process underlying both verbal and non-verbal WM, and seems to be more cognitively demanding.

Arimitsu (2020) argued the possibility that the ability to understand the meaning of negation is related to the accumulation of concrete experiences of spatial cognition in daily life. However, the relations between processing negations and spatial WM have been barely explored. This study becomes one of the first clear evidence for the relations, and further research on them will shed light to the mechanism of processing negations.

Tian et al. (2016) and Wang et al. (2021) argue that natural language is so pragmatically dynamic that participants accommodate a particular contextual situation and optimize their performance. Similar arguments are discussed in literature (Nieuwland & Kuperberg, 2008; Wason, 1965; Zhang et al., 2022). We wondered whether an ease of processing with practice, or reduction in the 'cost of computing the actual state of affairs (Wang et al., 2021)' due to learning experience, would transfer to processing negations. Yet, at least 10 sessions did not transfer processing negations from a two-step to a one-step procedure.

In summary, although ● ≠ ▲ seems to be a simple proposition that elementary school students can understand, the RTs for judgments of negated propositions were longer than for affirmations (i.e., TA < FN and FA < TN) and did not reach the level of TA judgment even after 10 practice sessions even by the university students, meaning that mental representations of negative propositions could not be encapsulated into one mental representation as assumed by the one-step procedure account.

Our findings of the relationship between verification judgments and WM provide new insights into processing negations, in that verifying negations may be associated with cognitive processing beyond linguistics. Especially, the high correlation coefficients between the linguistic and mathematical verification tasks and the N-back and mental hand rotation tasks indicate that holding and matching mental representations may be a shared cognitive process that is not limited to a modal process, either verbal or nonverbal. Thus, these findings demonstrate that exploring negation can be an effective approach not only to language but also to a broader cognitive system, or human intelligence (Horn, 2001), by bridging cognitive processing between linguistic and mathematical verification in relation to verbal and spatial WM.

4.3 Limitations of this study and future issues

This study demonstrated that processing negations in a single sentence and mathematical equation might share the same steplike cognitive process, the two-step procedure, and that holding and matching mental representations in accord with negation/affirmation propositions would be performed each time of a verification judgment even after the practice. However, whether or not the 10 practice sessions would have been sufficient to reject the one-step procedure account awaits further study.

Although the overall analyses clearly suggested the two-step process account in the simple linguistic and mathematical tasks in both experiments, we noticed some individual differences in processing negations in the case of FA and TN. All participants showed the trend (TA < FN) in both tasks of both experiments even after the practice, indicating the two-step process that participants first build a mental representation affirmative ▲ = ▲ and then integrate the negation symbol ≠. However, in the case of FA and TN, some participants showed the trend (FA < TN) but not others (FA = TN or FA > TN), although the overall analyses showed the significant difference (i.e., FA < TN). The findings in the case of FA and TN may suggest individual differences in the cognitive demand between these judgment types. The issue could be another factor to explain the previous inconsistent patterns of results in addition to the experimental conditions (e.g., timing of presenting a sentence and a picture, participants' visual abilities, and contexts suggested by probe questions) suggested by Wang et al. (2021). Future studies should explore the speculation.

Lastly, we did not empirically examine how mental representations are actually represented in the brain and how holding and matching them is performed in this present task. Considering the fact that negative expressions are used in daily life, further research is necessary to examine processing negations from children to older adults with simple verification tasks, such as those used in this study, in order to examine the developmental changes.

Data availability statement. The data from two experiments used in the analyses is available at https://osf.io/jvp3w.

References

Arimitsu, N. (2020). Negation. In Y. Ikegami, & M. Yamanashi (Eds.), *The innovation and inheritance of language studies (Volume 5) Cognitive linguistics II* (pp. 389–417). Hitsuzi Shobo.

Awh, E., Jonides, J., Smith, E. E., Schumacher, E. H., Koeppe, R. A., & Katz, S. (1996). Dissociation of storage and rehearsal in verbal working memory: Evidence from positron emission tomography. *Psychological Science*, 7(1), 25–32. https://doi.org/10.1111/j.1467-9280.1996.tb00662.x

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral Brain Sciences*, 22, 577–660. https://doi.org/10.1017/S0140525X99532147

Baddeley, A. D., & Hitch, G. (1974). Working memory. *The psychology of learning and motivation*, 8, 47–90.

Baddeley, A. D. (1968). A 3 min reasoning test based on grammatical transformation. *Psychonomic science*, 10(10), 341–342.

Bates, D, Mächler, M, Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. https://doi.org/10.18637/jss.v067.i01.

Carpenter, P. A., & Just, M. A. (1975). Sentence comprehension: A psycholinguistic processing model of verification. *Psychological Review*, 82(1), 45–73. https://doi.org/10.1037/h0076248

Clark, H. H., & Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology*, 3(3), 472–517. https://doi.org/10.1016/0010-0285(72)90019-9

Ferguson, H. J., Sanford, A. J., & Leuthold, H. (2008). Eye-movements and ERPs reveal the time course of processing negation and remitting counterfactual worlds. *Brain Research*, 1236, 113–125. https://doi.org/10.1016/j.brainres.2008.07.099

Fischler, I., Bloom, P. A., Childers, D. G., Roucos, S. E., & Perry Jr, N. W. (1983). Brain potentials related to stages of sentence verification. *Psychophysiology*, 20(4), 400–409. https://doi.org/10.1111/j.1469-8986.1983.tb00920.x

Hautus, M., Macmillan, N. A., & Creelman, C. D. (2021). *Detection theory: A user's guide* (3rd ed.). Routledge. https://doi.org/10.4324/9781003203636

Henninger, F., Shevchenko, Y., Mertens, U. K., Kieslich, P. J., & Hilbig, B. E. (2020). *Lab.js: A free, open, online study builder.* https://doi.org/10.3758/s13428-019-01283-5

Hida, K., & Sekiyama, K. (2013). A study of body schema of adults with autistic spectrum disorders: The influence of posture on mental rotation task of hands. *Kumamoto University Studies in Social and Cultural Sciences*, 11, 171–184.

Hitch, G. J., & Baddeley, A. D. (1976). Verbal reasoning and working memory. *The Quarterly Journal of Experimental Psychology*, 28(4), 603–621. https://doi.org/10.1080/14640747608400587

Horn, L. R. (2001). *A natural history of negation*. CSLI Publications, Leland Stanford Junior University.

Kaufman, S. B. (2007). Sex differences in mental rotation and spatial visualization ability: Can they be accounted for by differences in working memory capacity? *Intelligence*, 35(3), 211–223. https://doi.org/10.1016/j.intell.2006.07.009

Kaup, B., Lüdtke, J., & Zwaan, R. A. (2006). Processing negated sentences with contradictory predicates: Is a door that is not open mentally closed? *Journal of Pragmatics*, 38(7), 1033–1050. https://doi.org/10.1016/j.pragma.2005.09.012

Kinjo, H., Saito, T., Hamada, A., & Sakai, A. (2019). Comparing ability of comprehending negative sentences about general knowledge and "myself " between young and older adults in a questionnaire situation. *Bulletin of Psychology of Meiji Gakuin University*, 29, 1–17.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity ?! *Intelligence*, 14(4), 389–433. https://doi.org/10.1016/S0160-2896(05)80012-1

Lüdtke, J., Friedrich, C. K., Filippis, D., & Kaup, B. (2008). Event-related potential correlates of negation in a sentence – picture verification paradigm. *Journal of Cognitive Neuroscience*, 20(8), 1355–1370. https://doi.org/10.1162/jocn.2008.20093

Mayo, R., Schul, Y., & Burnstein, E. (2004). "I am not guilty" vs "I am innocent": Successful negation may depend on the schema used for its encoding. *Journal of Experimental Social Psychology*, 40, 433–449. https://doi.org/10.1016/j.jesp.2003.07.008

Neubauer, A. C., & Freudenthaler, H. H. (1994). Reaction times in a sentence-picture verification test and intelligence: Individual strategies and effects of extended practice. *Intelligence*, 19(2), 193–218. https://doi.org/10.1016/0160-2896(94)90013-2

Nieuwland, M. S., & Kuperberg, G. R. (2008). When the truth is not too hard to handle: An event-related potential study on the pragmatics of negation. *Psychological Science*, 19(12), 1213–1218. https://doi.org/10.1111/j.1467-9280.2008.02226.x

Orenes, I., Beltrán, D., & Santamaría, C. (2014). How negation is understood: Evidence from the visual world paradigm. *Journal of Memory and Language*, 74, 36–45. https://doi.org/10.1016/j.jml.2014.04.001

R Development Core Team. (2021). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. http://www.R-project.org/

Schendan, H. E., & Stern, C. E. (2007). Mental rotation and object categorization share a common network of prefrontal and dorsal and ventral regions of posterior cortex. *Neuroimage*, 35(3), 1264–1277. https://doi.org/10.1016/j.neuroimage.2007.01.012

Simon, D., Boot, W., Charness, N., Gathercole, S., Chabris, C., Hambrick, D., & Stine-Morrow, E. (2016). Do "Brain-Training" programs work? *Psychological Science in the Public Interest*, 17, 103–186. https://doi.org/10.1177/1529100616661983

Tanaka, M., & Yoshida, K. (2015). Relationship of mental rotation task with motor imagery vividness and working memory. *Journal of Kyushu University of Health and Welfare*, 16, 103–110. https://doi.org/10.15069/00000957

Tian, Y., Ferguson, H., & Breheny, R. (2016). Processing negation without context – why and when we represent the positive argument. *Language, Cognition and Neuroscience*, 31(5), 683–698. https://doi.org/10.1080/23273798.2016.1140214

Usami, K., & Soujima, K. (2015). Hattatsu shinrigaku no tameno toukeigaku: Jyuudann data no bunnseki. In *Statistics for developmental psychology: Analysis for longitudinal studies*. Seishin-Shobo.

Vanek, N., Škorić, A., Košutar, S., Matějka, Š., & Stone, K. (2024). Mental simulation of the factual and the illusory in negation processing: evidence from anticipatory eye movements on a blank screen. *Scientific Reports*, 14, 2844. https://doi.org/10.1038/s41598-024-53353-0

Vanek, N. & Zhang, H. (2023). On truth and polarity in negation processing: Language-specific effects in non-linguistic context. *Frontiers in Psychology*, 14, 1244249. https://doi.org/10.3389/fpsyg.2023.1244249

Veale, J. F. (2013). Edinburgh handedness inventory—short form: A revised based on confirmatory factor analysis. *Laterality: Asymmetries of Body, Brain and Cognition*, 19(2), 164–177.

Wang, S., Sun, C., Tian, Y., & Breheny, R. (2021). Verifying negative sentences. *Journal of Psycholinguistic Research*, 50(6), 1511–1534. https://doi.org/10.1007/s10936-021-09798-9

Wason, P. C. (1961). Response to affirmative and negative binary statements. *British Journal of Psychology*, 52(2), 133–142. https://doi.org/10.1111/j.2044-8295.1961.tb00775.x

Wason, P. C. (1965). The contexts of plausible denial. *Journal of Verbal Learning and Verbal Behavior*, 4, 7–11. https://doi.org/10.1016/S0022-5371(65)80060-3

Woltz, D.J. (1988). An investigation of the role of working memory in procedural skill acquisition. *Journal of Experimental Psychology: General*, 117(3), 319–331. https://doi.org/10.1037/0096-3445.117.3.319

Zhang, H., Wang, Y., & Vanek, N. (2022). Negation processing in Chinese–English bilinguals: insights from the Stroop paradigm and an orientation task. *International Journal of Bilingual Education and Bilingualism*, 25(10), 3711–3728. https://doi.org/10.1080/13670050.2022.2074274