Journal of Memory and Language 148 (2026) 104749

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Action and abstraction: Motor interference changes meaning in language understanding

Pablo Solana [a,b,1,*], Omar Escámez [a,b,1], Gabriella Vigliocco [c], Daniel Casasanto [d], Julio Santiago [a,b]

[a] Mind, Brain and Behavior Research Center (CIMCYC), University of Granada, Spain
[b] Department of Experimental Psychology, University of Granada, Spain
[c] Department of Experimental Psychology, University College London, UK
[d] Department of Psychology, Cornell University, USA

ARTICLE INFO

Keywords:
Embodied cognition
Semantics
Language comprehension
Motor system
Construal level
Abstraction

ABSTRACT

Can the body shape meaning? Eight experiments (four preregistered) tested whether interfering with the motor system changes how people interpret language about actions. Participants (total $N = 880$) rhythmically moved their hands or feet while being presented with sentences describing hand or foot actions (e.g., “scoring a goal in soccer”) and asked to choose between two interpretations of their meaning: one more concrete (e.g., ”kicking a ball“) and another more abstract (e.g., ”winning a match“). Despite not all experiments showed significant results, the overall pattern revealed effector-specific effects of motor interference on meaning construction, which were further modulated by the amount of delay between the sentences and their interpretations. When the delay was short (200 ms), participants chose more concrete interpretations for described actions that involved the same effector being moved. In contrast, when the delay was long (15 s), participants who moved their feet chose more abstract interpretations for foot-related sentences. Although preliminary, these results provide the first evidence that motor action can cause qualitative changes in sentence understanding, consistent with the functional role of the motor system in lexical semantics suggested by embodiment theories.

Introduction

How does the motor system contribute to language understanding? According to 20th-century models of language, the role of the motor system was limited to producing the forms of words (Cutler & Clifton, 1999; Levelt, 1999). However, with the rise of embodied and grounded approaches to cognition, a growing body of research suggests that the motor system also plays a role in computing the meaning of words and phrases that describe actions (see Fischer & Zwaan, 2008; Mahon & Caramazza, 2008; Meteyard et al., 2012; Pulvermüller, 2013; Willems & Casasanto, 2011, for reviews offering contrasting perspectives). Here we tested whether manipulating motor activity can influence how abstractly or concretely people mentally represent actions described in language.

In neuroimaging studies, verbs like “throw” and “kick” activate hand and foot regions of motor or premotor cortices (e.g., Aziz-Zadeh et al., 2006; Hauk et al., 2004; Tettamanti et al., 2005; Willems et al., 2010a). In motor interference and neurostimulation studies, perturbing activity in hand or foot motor areas modulates performance in judgements of action verbs and sentences (e.g., Gijssels et al., 2018; Glenberg et al., 2008; Pulvermüller et al., 2005; Shebani & Pulvermüller, 2013; Willems et al., 2011). Together, these studies suggest two conclusions. First, when people understand action verbs, they create implicit simulations of the actions that the verbs name, using the same cortical motor circuits that are involved in planning and executing these actions (but see Postle et al., 2008). Second, activity in effector-specific motor areas plays some functional role in constructing the meanings of action language (but see Montero-Melis et al., 2022; Solana & Santiago, 2022, 2023).

Nonetheless, previous evidence that motor simulations functionally contribute to language understanding has focused on demonstrating *quantitative* changes in reaction times and accuracy. Could interfering with motor simulations lead to *qualitative* changes in language understanding, causing the same words or phrases to mean something different? Providing evidence for this idea could have significant

* Corresponding author at: Mind, Brain and Behavior Research Center (CIMCYC), University of Granada, Spain.
E-mail address: solana@ugr.es (P. Solana).
[1] These authors contributed equally.

https://doi.org/10.1016/j.jml.2026.104749
Received 16 June 2025; Received in revised form 9 February 2026; Accepted 16 February 2026
Available online 19 February 2026

implications for theories of embodied language comprehension. It would suggest that the motor system is not only relevant for understanding meaning efficiently but also shapes the kinds of meanings that people construe from language, thereby providing compelling evidence for a causal contribution of motor information to language semantics (Casasanto, 2023; Willems & Casasanto, 2011).

To address this question, we took advantage of the fact that actions can be construed at multiple levels of abstraction (Trope & Liberman, 2010; Vallacher & Wegner, 1987, 1989). Like many psychological constructs, the terms "abstract" and "concrete" are defined differently in different contexts (Borghi, 2022; Reilly et al., 2025). Following Vallacher and Wegner (1987) and others, here we posit that abstract interpretations of motor actions highlight the actions' goals or outcomes, and are general enough to encompass many possible realizations of the action. By contrast, relatively concrete interpretations of motor actions highlight the actions' mechanical details, and therefore specify how the action is realized. For example, an action like "casting a vote" can be understood in terms of its abstract goals (e.g., "supporting a candidate") or its concrete, mechanical details (e.g., "writing an X on a ballot"). Of crucial importance for our study, abstract and concrete action interpretations typically differ in the extent to which they are defined by the movement of particular body parts. Whereas "more abstract" action interpretations tend to be "less movement defined" (Vallacher & Wegner, 1987, p. 12), "more concrete" representations of the same actions specify "the actual movements to be undertaken" and the effectors used to perform them (ibid, p. 4).

We hypothesized that motor simulations help language users represent the motoric details of actions and, therefore, that the enhancement or suppression of motor simulations can cause qualitative changes in the interpretations that language users construe. In a previous study (Solana et al., 2024), we tested this hypothesis by adapting Vallacher and Wegner's (1989) Behavioral Identification Form (BIF), a questionnaire used to measure individuals' inclination to construe actions concretely or abstractly. For each action described, participants were asked to choose either a concrete interpretation that highlights the action's motoric details, or a more abstract interpretation that describes the action focusing on its goals or outcomes. Half the actions involved the hands, while the other half involved the feet. Before completing the task, participants were applied low-frequency repetitive transcranial magnetic stimulation (rTMS) to either the hand region of the left primary motor cortex (M1) or the vertex (as a control region). As this protocol is expected to inhibit the excitability of targeted brain regions (Chen et al., 1997), we predicted rTMS over the hand M1 to hinder the computation of the concrete details of hand sentences, biasing participants toward more abstract interpretations than for foot sentences. However, despite successfully inhibiting M1, we observed no significant changes in participants' interpretation choices.

Here we report a complementary test for the hypothesis that was conducted in parallel to the TMS study. Instead of altering participants' motor system via brain stimulation, participants in the present study tapped a complex rhythm with either their hands or their feet—a task that requires the recruitment of the neural circuits for motor planning and performance (Shebani & Pulvermüller, 2013). Interestingly, a study published during the preparation of the present work also used this manipulation and concluded in support of qualitative changes in language comprehension (Togato et al., 2021). Specifically, these authors reported that tapping a rhythm with the hands or the feet altered the likelihood of choosing hand- and foot-related meanings for homograph words. Whereas Togato and colleagues used single words and focused on the preference towards the two meanings of a homograph, here we used sentences describing actions and focused on the level of abstraction at which they can be construed. Additionally, we also explored the flexibility and time course of motor simulation during meaning construction by varying the amount of delay between the action and its interpretations across experiments: 15 s (Experiments 1, 4–7) vs. 200 ms (Experiments 2, 3 and 8). The specific design and hypotheses of each experiment are detailed in the sections below.

The manuscript is structured as follows. First, we present eight experiments designed to test our predictions and explore their modulation by delay. Next, we report a series of simulations, along with an integrative analysis of all the experiments, aiming to provide a comprehensive explanation for the pattern of results observed across experiments.

Transparency and openness

We hereby state that we report how we determined our sample sizes, all data exclusions, all manipulations, and all measures in the study. Experiments 5–8 were preregistered prior to data collection. Experiments 1–4 were not preregistered but were analyzed using the analysis pipeline preregistered for Experiments 5–8. All data, analysis code, materials, and supplementary information are available at the Open Science Framework (OSF) repository: https://osf.io/58uvr/.

Experiment 1

In the first experiment, participants were visually presented with sentences describing hand or foot actions and chose between a concrete and an abstract interpretation of each action. Between the presentation of the action and the interpretations, participants rhythmically tapped either their hands or feet for 15 s. In their seminal study, Shebani and Pulvermüller (2013) showed that moving the hands/feet for 6 s impairs memory for hand/foot verbs. Therefore, we hypothesized that performing the same rhythmic sequence for 15 s would interfere with the simulation of the motoric details of action verbs implying the same effector, thus fostering abstract interpretations. Specifically, tapping with the hands should interfere with participants' ability to simulate hand actions, leading to a greater proportion of abstract hand-action interpretations. Likewise, tapping with the feet should interfere with participants' ability to simulate foot actions, leading to a greater proportion of abstract foot-action interpretations.

Methods

Participants

Ninety students from the University of Granada (Spain) took part in the experiment ($M_{age} = 20.24$, $SD_{age} = 2.32$; 8 men; 10 left-handers). All of them were native Spanish speakers, participated voluntarily, and were compensated with course credits. The Ethics Committee on Human Research of the University of Granada approved the experiment (697/CEIH/2018). Sample size for this study was not pre-established, as it was impossible to estimate an effect size from prior literature. We set a minimum of 40 participants per group (doubling the minimum of 20 participants per cell recommended by Simmons et al., 2011) and ended up running 45 because of participant availability.

Materials

There were 24 sets of action descriptions and interpretations (see the OSF repository for the full list). Half of the items described hand actions (e.g., "writing a message on a computer"), and the other half described foot actions (e.g., "scoring a goal in soccer"). As in Vallacher and Wegner (1989), each description was followed by two interpretations, one more concrete (e.g., "typing letters/kicking a ball into a net") and the other more abstract (e.g., "getting in touch with someone/getting ahead on the scoreboard").

The 24 descriptions and interpretations used in the experiment were selected from a set of 38 hand and 38 foot descriptions, each with a concrete and an abstract interpretation, that were normed using an independent sample of 42 participants from the same population. Participants in the norming study received a leaflet containing all 76 sentences, each one paired with its two interpretations. Half the items were followed first by the concrete interpretation and the other half by the abstract interpretation. There were two leaflets with a different random order of the items. Participants marked which interpretation best described the meaning of each action description. For each description we computed the proportion of participants who chose the more abstract interpretation. To avoid ceiling/floor effects, experimental sentences were chosen from items with proportions between 0.40 and 0.60.

To ensure that the task and the selected items were valid to track changes in abstraction level, we carried out two validation studies, one in person ($N = 40$) and another online ($N = 100$).[2] In both studies, we followed the strategy developed by Mac Giolla and colleagues (2025) to validate the BIF: participants were given explicit instructions to think in concrete or abstract terms about the actions (i.e., "Focus on the physical performance of the actions" vs. "Focus on the overall purpose and meaning of the actions") and then select the interpretation they felt more adequate. Crucially, participants in the abstract groups selected significantly more abstract interpretations (in person: $M = 0.66$; online: $M = 0.63$) than participants in the concrete groups (in person: $M = 0.23$, $p < 0.001$; online: $M = 0.07$, $p < 0.001$), thus confirming the validity of the selected materials to measure changes in abstraction level. Detailed information about the preregistration, participants, procedure, analysis, and results of these validation studies is provided as supplementary materials in the OSF repository.

Procedure

Participants were randomly assigned to two groups. One group performed a rhythm by tapping with their hands on the table ($N = 45$), whereas the other group tapped with their feet on the floor ($N = 45$). The effector producing the rhythm was manipulated between groups to avoid participants guessing the objective of the study. The rhythm was the same used by Shebani and Pulvermüller (2013): RLRRLRLL (where R is the right, and L is the left hand/foot). Participants practiced the pattern until they were able to produce it fluently and accurately. Then, the experimenter explained the task to the participant and the experiment started.

Participants started the first trial by pressing the space bar, which triggered an action description to appear in the center of the screen, where it remained until the participant pressed the space bar. The sentence was then replaced by the word "RITMO" (rhythm). Participants performed the rhythm for 15 s until the word disappeared from the screen and was followed by the two interpretations of the sentence. These were presented one above the other, numbered with the digits 1 and 2. Participants read the sentences without time pressure and then pressed the key 1 or 2 to indicate which interpretation they thought best fit the action description. They were told that both interpretations were equally valid. After the response was collected, the screen was cleared and a message informed them to press the space bar, which began the next trial after a 1000 ms blank interval.

The experimental block comprised 24 trials (one per each set of action and interpretations; 12 hand actions and 12 foot actions) and was preceded by a 5-trials practice block. We did not repeat items as it seemed likely that, once the participants have chosen one interpretation for a given sentence, they would choose the same interpretation in a subsequent presentation of the sentence. There were four optional breaks, one every six trials. Sentence order was randomized for each participant, and the vertical location of the concrete and abstract interpretations for each item was counterbalanced across participants.

The experimenter stayed in the room during the whole session, outside of participant's view, monitoring rhythm production and correcting any important deviations (e.g., stopping the rhythm altogether, producing a simple alternating-hand rhythm, or failing to fall back into the sequence immediately). Occasional mis-taps were not recorded. We set out to filter out participants who systematically failed to perform the rhythm faithfully, clearly signaling that they were not performing the task as intended, but no participant met this criterion.

Data analysis

All the experiments were analyzed by means of mixed-effects binomial regressions, using the *lme4* package in R (Bates et al., 2015a). These models allowed us to treat subjects and items as simultaneous random effects, and to avoid treating categorical judgments as continuous (Jaeger, 2008).

We always started with the maximal model justified by the design, to avoid increasing Type I error (Barr et al., 2013). The fixed term of this model included the factors Verb (hand vs. foot action) and Rhythm (hand vs. foot rhythm), as well as their interaction. Factors were dummy coded as $+0.5$ and -0.5 to be treated as predictors centered in zero. The random term included random intercepts for Items and Participants, as well as the slope of the fixed factors that varied over the corresponding random factor. The dependent variable was the participant's choice: abstract (1) or concrete (0). In R notation: *glmer(data, Choice ~ Rhythm*Verb + (1 + Rhythm|Item) + (1 + Verb|Participant), family = binomial(link = logit))*.

We then took a top-down approach and searched for the simplest model that retained the maximum degree of fit to the data (Bates et al., 2015b). The full model selection process for all experiments is detailed in the analysis scripts available in the OSF repository. *P*-values were obtained from type II Wald chi-square tests, using the *car* package (Fox & Weisberg, 2019). Effect sizes were calculated as odds ratios (*OR*) by exponentiating the estimates of the models, and 95% CIs around them were computed using bootstrapping methods.

Results and discussion

As predicted, motor interference affected meaning construction from action sentences in an effector-specific way, as indexed by a significant interaction between Rhythm and Verb ($\chi^2(1) = 23.03$, $p < 0.001$, $OR = 2.67$, *95% CI* = [1.78, 4.01]; Fig. 1). Participants in the hand group chose more abstract interpretations for hand sentences ($M = 0.52$) than for foot sentences ($M = 0.44$; $\chi^2(1) = 4.86$, $p = 0.03$, $OR = 1.52$, *95% CI* = [1.03, 2.27]), while participants in the foot group selected more abstract interpretations for foot sentences ($M = 0.34$) than for hand sentences ($M = 0.25$; $\chi^2(1) = 4.20$, $p = 0.04$, $OR = 1.75$, *95% CI* = [1.01, 3.13]). Experiment 1 also revealed a main effect of Rhythm: participants who tapped with the hands chose more abstract interpretations than participants who tapped with the feet ($\chi^2(1) = 14.35$, $p < 0.001$, $OR = 1.70$, *95% CI* = [0.98, 2.99]). The effect of Verb was not significant ($\chi^2(1) = 0.01$, $p = 0.94$, $OR = 0.58$, *95% CI* = [0.36, 0.92]).

These results provide a clear-cut case for a causal role of motor processes in construing meaning from language: satiating motor areas with a secondary repetitive motor task during language comprehension qualitatively changes the meaning arrived at.

According to prominent proposals, simulations require time to unfold (e.g., Barsalou et al., 2008). Consequently, altering simulations at different points of their development may differentially affect how meaning is construed. In line with this idea, previous findings support that motor-language interactions change over the temporal course of language comprehension. For instance, in a kinematic study, Boulenger et al. (2006) showed that processing action verbs interferes with a concurrent reaching movement, whereas the same words facilitate responses when presented before movement onset. Another example is provided by studies showing that the sign of the Action-Sentence Compatibility Effect (ACE) depends on the SOA between sentence and response: at short SOAs, there is an interference on reaction times, whereas facilitation is found with longer SOAs (de Vega et al., 2013; Winter et al., 2022). In the present experiment, we included a 15 s delay between sentence and interpretations. Would it be possible that a

[2] These validation studies, as well as the one reported in Experiment 6, were carried out once the experimental series was completed, but they are reported here to increase the readability of the manuscript.

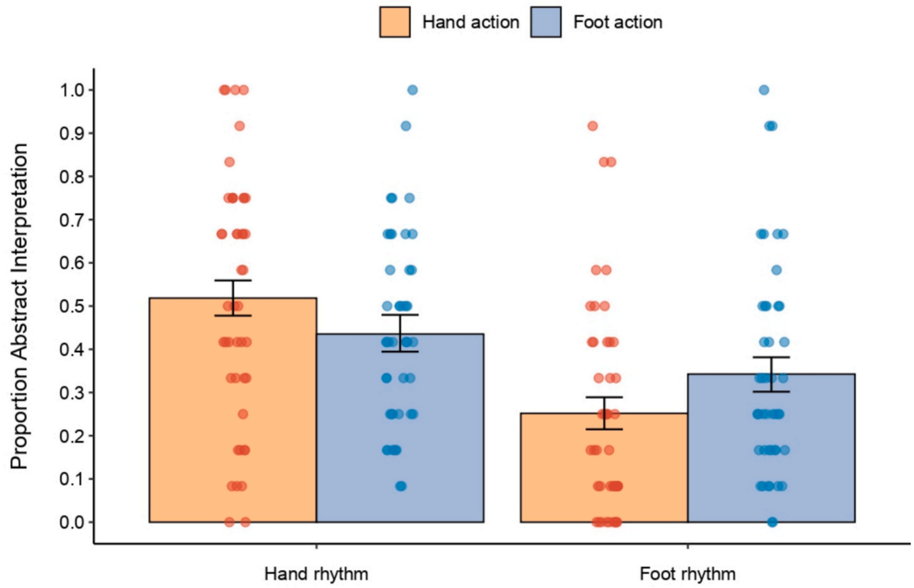

Fig. 1. Proportion of abstract interpretations in Experiment 1 as a function of Rhythm and Verb. Jittered points represent the individual mean of each participant. Error bars represent the 95% CI around the mean.

shorter delay changes the observed interaction? To test for this possibility, in the next experiment, we used a shorter delay of 200 ms.

Experiment 2

Experiment 2 was devoted to exploring the generalizability, flexibility and time course of the involvement of motor simulation in language comprehension. To this end, the delay between sentence and interpretations was very short (200 ms), and participants produced the rhythm in a continuous fashion throughout the session. The sentences were presented auditorily and participants gave their response verbally. Drawing on previous evidence supporting that motor-language interactions change over the temporal course of language processing (see above), we predicted the shape of the interaction in Experiment 1 to change, although we did not have clear expectations for the exact form of the change.

The changes in the modality of presentation and the simultaneity between rhythm and sentence processing were implemented to examine the generalizability of the effect to other plausible contexts, reflecting how language is typically experienced: mainly orally and often accompanied by gestures. Accordingly, we expected these changes to have no effect—although we controlled them in later experiments. The modality of response was changed from manual to verbal in order to avoid participants using their hands to respond, which may influence motor system activation (e.g., Hauk et al., 2004). We also expected this change to have no specific effect.

Method

Participants

We set the same sample size as in Experiment 1. Ninety new native Spanish speakers from the same population as in Experiment 1 took part in the experiment ($M_{age} = 21.2$, $SD_{age} = 2.76$; 18 men; 6 left-handers).

Materials and procedure

Everything was similar to Experiment 1, with the following exceptions. At the beginning, participants were instructed to perform the rhythm during the whole session. The experiment started when the researcher pressed the space bar and the word “RITMO” appeared on the screen. After performing the rhythm for 10 s, an action description was auditorily presented through headphones. We used the same 24 sentences as in Experiment 1, recorded by a female native Spanish speaker. 200 ms after the end of the sentence, the word “RITMO” was replaced by the two interpretations of the sentence. Participants gave their response aloud and the experimenter inputted the selected number using a keyboard, out of participants’ sight. We used a random inter-trial interval between 200 and 1000 ms.

Results and discussion

In line with our prediction, the results showed again an interaction between Rhythm and Verb ($\chi^2(1) = 123.81, p < 0.001, OR = 11.11$, *95% CI* = [7.14, 16.67]; Fig. 2), but with a complete reversal in its shape: participants in the hand group chose less abstract interpretations for hand sentences ($M = 0.16$) than for foot sentences ($M = 0.39$; $\chi^2(1) = 31.13, p < 0.001, OR = 3.85$, *95% CI* = [2.33, 6.25]), while participants in the foot group selected less abstract interpretations for foot sentences ($M = 0.26$) than for hand sentences ($M = 0.45$; $\chi^2(1) = 10.11, p < 0.001, OR = 3.03$, *95% CI* = [1.50, 6.31]). Additionally, the present experiment also found a main effect of Rhythm ($\chi^2(1) = 4.74, p = 0.03, OR = 2.03$, *95% CI* = [1.36, 3.03]), with participants who tapped with the feet choosing more abstract interpretations than participants who tapped with the hands. The main effect of Verb was not significant ($\chi^2(1) = 0.001, p = 0.92, OR = 2.83$, *95% CI* = [1.62, 5.03]).

This result stands in sharp contrast to the findings of Experiment 1 and suggest a dissociation between the processes underlying meaning computation in the two experiments. Yet, the cause of this dissociation is so far unclear, given that several aspects changed between Experiment 1 and 2: besides delay duration, the reversal of the interaction could also be explained by the use of the auditory-vocal modality for sentence presentation and response (vs. the visual-manual modality of Experiment 1), or because, contrary to the previous experiment, the rhythm was being performed at the same time as the sentences were being listened to. Experiments 3–5 were aimed to disentangle these potential causes.

Experiment 3

In Experiment 3, we eliminated the simultaneity between performing the rhythm and processing the sentences, while maintaining the short

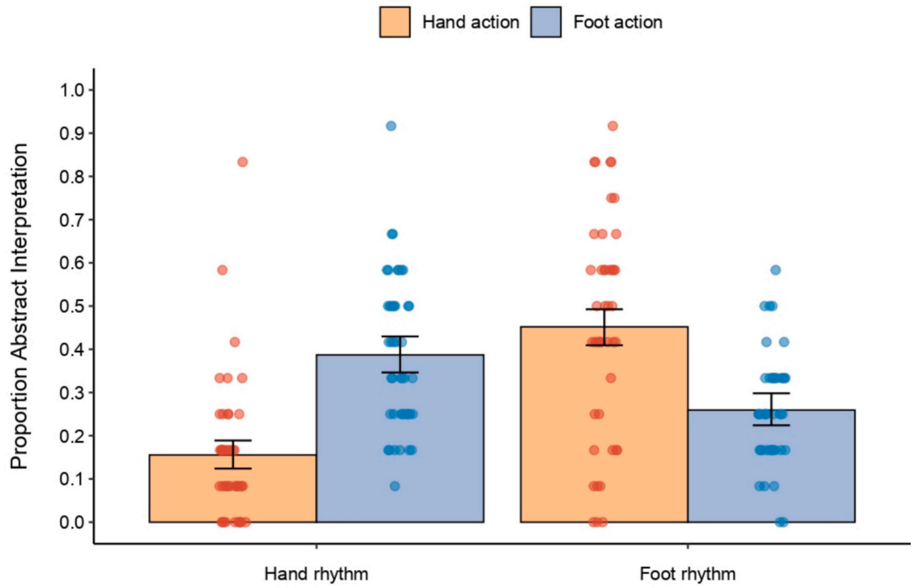

Fig. 2. Proportion of abstract interpretations in Experiment 2 as a function of Rhythm and Verb. Jittered points represent the individual mean of each participant. Error bars represent the 95% CI around the mean.

delay and the auditory-vocal modality. As we expected the delay to be the crucial factor mediating the shape of the interaction, we predicted the same pattern of results as in Experiment 2.

Methods

Participants

We set the same sample size than in the previous experiments. Ninety new native Spanish speakers from the same population as in Experiment 1 took part in the experiment (M_{age} = 20.02, SD_{age} = 2.45; 19 men; 3 left-handers).

Materials and procedure

Everything was kept exactly as in Experiment 2, with the following exceptions. At the beginning of each trial, the word "RITMO" appeared on the screen for 10 s and the participants performed the rhythm during this time. Then, the word "RITMO" was replaced by the word "STOP", indicating the participants to stop performing the rhythm. After a pause of 1000 ms, a sentence was presented through headphones, and 200 ms later, the word "STOP" was replaced by the two interpretations.

Results and discussion

The analysis revealed no main effect of Verb ($\chi^2(1) = 1.32$, $p = 0.25$, *OR* = 2.13, *95% CI* = [1.31, 3.52]) nor of Rhythm ($\chi^2(1) = 0.09$, $p = 0.76$, *OR* = 2.60, *95% CI* = [1.68, 4.06]). However, the interaction between Rhythm and Verb was significant ($\chi^2(1) = 107.23$, $p < 0.001$, *OR* = 8.33, *95% CI* = [5.56, 12.50]; Fig. 3). Participants in the hand group chose fewer abstract interpretations for hand sentences ($M = 0.22$) than for foot sentences ($M = 0.48$; $\chi^2(1) = 35.96$, $p < 0.001$, *OR* = 0.26, *95%*

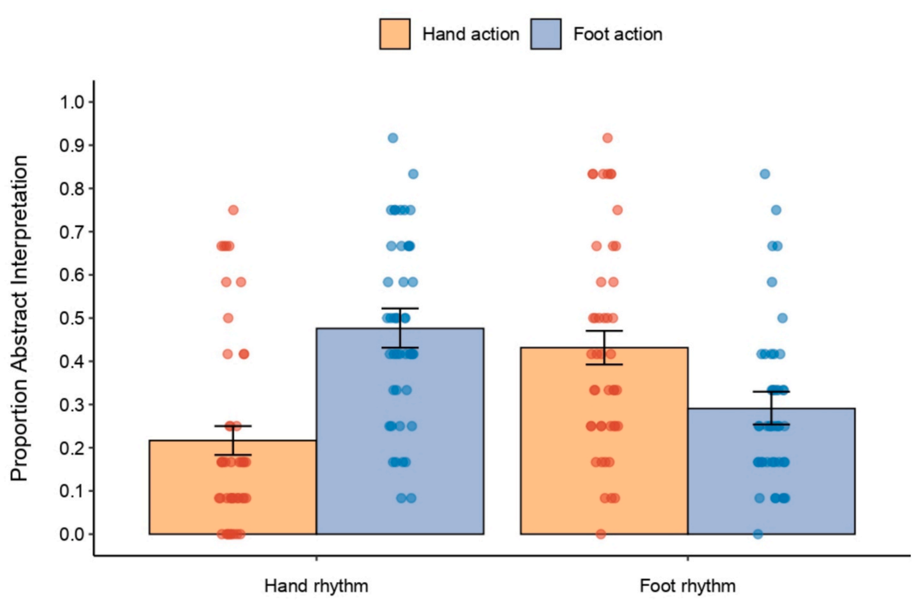

Fig. 3. Proportion of abstract interpretations in Experiment 3 as a function of Rhythm and Verb. Jittered points represent the individual mean of each participant. Error bars represent the 95% CI around the mean.

CI = [0.16, 0.41]), while participants in the foot group selected fewer abstract interpretations for foot sentences ($M = 0.29$) than for hand sentences ($M = 0.43$; $\chi^2(1) = 6.62$, $p = 0.01$, $OR = 2.19$, *95% CI* = [1.18, 4.14]).

These results replicated Experiment 2 when the rhythm was stopped before the sentence was presented (instead of being simultaneous). This suggests that the dimension of simultaneity-sequentiality is not responsible of the reversal in the shape of the interaction that we witnessed from Experiment 1 to Experiment 2. In the next experiment, we put to direct test the hypothesis that the length of the delay interval is the crucial factor in determining the shape of the interaction.

Experiment 4

In Experiment 4, we kept everything identical to Experiment 3 (including the auditory-vocal modality), but exchanged the positions of the beginning of the rhythm and the presentation of the sentence, such that the delay between them became as long as in Experiment 1. We hypothesized that the length of the delay is the crucial factor determining the shape of the interaction, so we expected to observe a pattern of results as in Experiment 1.

Methods

Participants

We set the same sample size than in the previous experiments. Ninety new native Spanish speakers from the same population as in Experiment 1 took part in the experiment ($M_{age} = 20.43$, $SD_{age} = 2.51$; 22 men; 9 left-handers).

Materials and procedure

The procedure was as in Experiment 3, with the following changes. At the beginning of the trial, a sentence was presented auditorily. 200 ms later, the word "RITMO" appeared on the screen, indicating the participants to perform the rhythm, which was maintained for 15 s. After that, the word disappeared and was replaced by the two interpretations of the sentence. Then, the participants stopped performing the rhythm and made their decision.

Results and discussion

The experiment showed a significant interaction between Rhythm and Verb ($\chi^2(1) = 5.28$, $p = 0.02$, $OR = 1.60$, *95% CI* = [1.07, 2.41]; Fig. 4), which exhibited the predicted qualitative pattern of results also found in Experiment 1: participants in the hand group chose more abstract interpretations for hand sentences ($M = 0.39$) than for foot sentences ($M = 0.37$; $\chi 2(1) = 0.49$, $p = 0.48$, $OR = 1.18$, *95% CI* = [0.74, 1.91]), whilst participants in the foot group selected more abstract interpretations for foot sentences ($M = 0.34$) than for hand sentences ($M = 0.29$; $\chi 2(1) = 1.39$, $p = 0.24$, $OR = 0.73$, *95% CI* = [0.43, 1.23]), although these pair-wise comparisons were not significant. Neither the main effect of Rhythm ($\chi^2(1) = 2.38$, $p = 0.12$, $OR = 1.19$, *95% CI* = [0.68, 2.07]) nor the main effect of Verb ($\chi^2(1) = 0.02$, $p = 0.88$, $OR = 1.33$, *95% CI* = [0.79, 2.27]) was significant.

These results support the notion that the key factor which controls the shape of the interaction is the length of the delay between sentence and interpretation: with a short delay, a secondary motor task biases participants toward more concrete interpretations, while with a long delay, it biases participants toward more abstract interpretations—always in an effector-specific way. Yet, the size of the interaction was smaller in Experiment 4, and the pairwise contrasts of Verb in each motor interference condition did not reach significance. To ensure the robustness of present findings, we carried out two additional experiments. First, another long-delay experiment, but incorporating again the simultaneity between rhythm and sentence, which was expected to have no influence on the effect (Experiment 5). Second, a replication in English of Experiment 4 (Experiment 6). Moreover, for the first time in this experiment series, we based the sample sizes on power calculations and preregistered the studies before data collection.

Experiment 5

In Experiment 5, we maintained the long delay and the auditory-vocal modality, but participants performed the rhythm during the whole experimental session while they processed the sentences. Since we believed that the simultaneity between performing the motor task and processing the actions was not relevant to our results, we expected to observe an interaction with the same shape as in Experiments 1 and 4. Hypotheses, sample size, design, and analysis plan for this experiment were preregistered prior to data collection: https://osf.io/fcjxp/files/e43p9.

Methods

Participants

Experiment 4 found that the interaction between Rhythm and Verb had an effect size of $OR = 1.60$, which corresponds with an estimate of $\beta = 0.47$. By means of Monte Carlo simulations using the R package *simr* (Green & MacLeod, 2016), we computed the sample size required to observe an effect size of $\beta = 0.5$ with 80% power and an alpha of 5%. The computations resulted in a minimum of 60 participants per group (total $N = 120$). However, following the principles of sequential analysis (Wald, 2004; see also Lakens, 2014), we established an additional constraint: we would divide alpha between two analysis points. When we reached $N = 120$, we would test the effect of interest at alpha = 0.03. If not significant, we would continue data collection until reaching 65 participants per group (total $N = 130$) and test the effect of interest at alpha = 0.02. By establishing this conditional stopping rule, we keep Type I error at the desired level of 5%.

At last, 120 new native Spanish speakers from the same population as in Experiment 1 took part in the experiment ($M_{age} = 21.38$, $SD_{age} = 3.66$; 14 men; 12 left-handers). We did not reach $N = 130$ because the results with 120 participants clearly suggested the absence of the expected interaction, and we believed that adding five more participants per group would not make the desired effect arise.

Materials and procedure

This experiment was the same as Experiment 4, with the only difference that the participants performed the motor task in a continuous fashion during the whole session (with optional breaks every six trials). Each trial started when the experimenter pressed the space bar, which made the word "RITMO" appear on the screen. Two seconds later, the action description was presented auditorily. After 15 s, the word RITMO disappeared and participants were presented with the two interpretations.

Results and discussion

There was not a main effect of Verb ($\chi^2(1) = 0.73$, $p = 0.39$, $OR = 0.82$, *95% CI* = [0.49, 1.35]) nor Rhythm ($\chi^2(1) = 0.03$, $p = 0.86$, $OR = 1.03$, *95% CI* = [0.69, 1.51]). Contrary to our expectations, the interaction between Rhythm and Verb was also non-significant ($\chi^2(1) = 0.003$, $p = 0.96$, $OR = 1.01$, *95% CI* = [0.72, 1.42]; Fig. 5). Participants who moved the hands did not differ in their proportion of abstract choices for hand ($M = 0.30$) and foot sentences ($M = 0.34$; $\chi^2(1) = 0.63$, $p = 0.42$, $OR = 0.82$, *95% CI* = [0.49, 1.37]). Participants who moved the feet also did not differ in their proportion of abstract choices for hand ($M = 0.29$) and foot sentences ($M = 0.34$; $\chi^2(1) = 0.61$, $p = 0.43$, $OR = 0.81$, *95% CI* = [0.48, 1.40]). However, it might be argued that the lack of interaction was due to the procedural changes implemented in the present experiment. This concern was mitigated by the following experiment, a closer replication of Experiment 4.

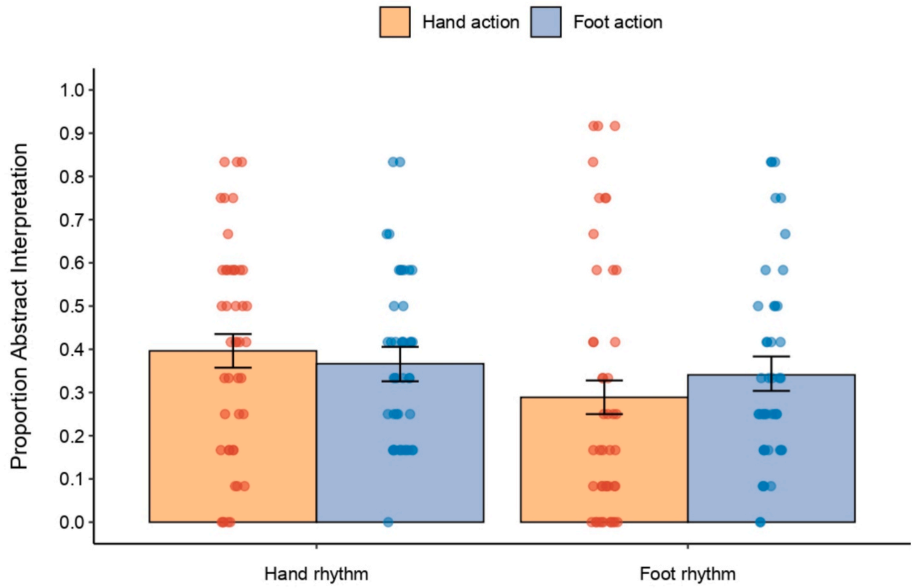

Fig. 4. Proportion of abstract interpretations in Experiment 4 as a function of Rhythm and Verb. Jittered points represent the individual mean of each participant. Error bars represent the 95% CI around the mean.

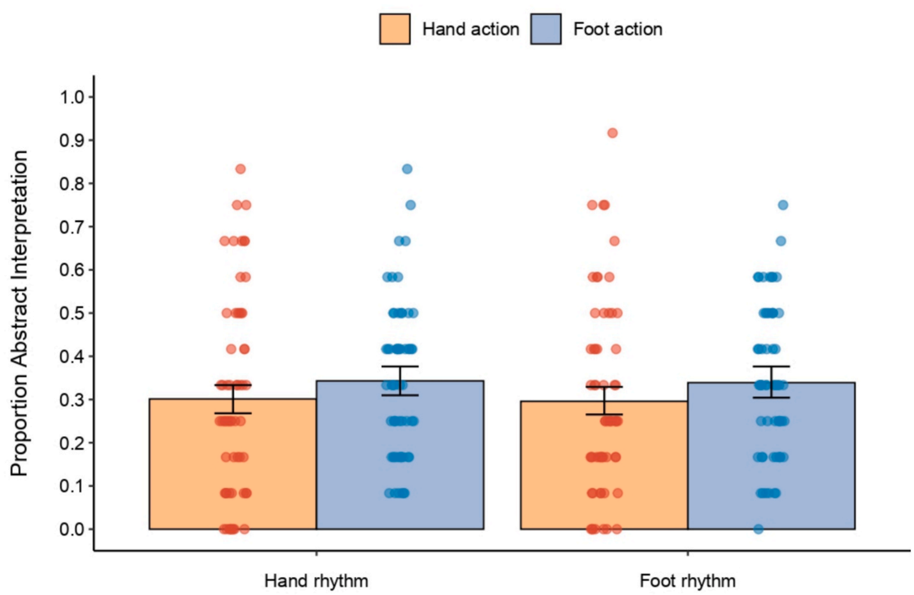

Fig. 5. Proportion of abstract interpretations in Experiment 5 as a function of Rhythm and Verb. Jittered points represent the individual mean of each participant. Error bars represent the 95% CI around the mean.

Experiment 6

Experiment 6 was a direct replication of Experiment 4. However, because this study was conducted while the Spanish lab was running Experiment 5, it took place in the United States, using English materials and American participants. We did not expect linguistic or cultural differences to play any role, so we predicted to find the same pattern of results than in Experiment 4 (and 1). The study was preregistered prior to data collection: https://aspredicted.org/mrv5-27qx.pdf.

Methods

Participants

We run the same power calculations described in detail in Experiment 5. However, as the change in the language of the materials gave us the opportunity to create new sentences, we also increased their number and used 28 sentences instead of 24. This caused the estimated sample size for securing 80% power to come down to 50 participants per group (total $N = 100$). Yet, we again divided the alpha level between two analysis points: $N = 100$ at $\alpha = 0.03$, and $N = 120$ at $\alpha = 0.02$.

Overall, 235 students from Cornell University (USA) took part in the experiment. Following our preregistration, we discarded 96 participants who were not native English speakers, and 19 who did not perform the tapping task correctly. Thus, as preregistered, only the data from 120 people ($M_{age} = 19.58$, $SD_{age} = 1.96$; 9 men; 11 left-handers) were taken into account for the analyses. All the participants took part in the study voluntarily and were paid with money. The Institutional Review Board for Human Participant Research of Cornell University approved the study (#1709007462).

Materials and procedure

Following the same logic presented in detail in Experiment 1, we created a new set of 28 sentences in English and their corresponding interpretations (14 hand actions and 14 foot actions), whose proportion of choice of the abstract interpretation in a norming study ($N = 100$) was between 0.40 and 0.60. These materials were validated in an online experiment ($N = 100$) with the same procedure as previously detailed for the Spanish materials: participants in the abstract group selected more abstract interpretations ($M = 0.83$) than those in the concrete group ($M = 0.17$; $p < 0.001$; detailed information about this validation study can be found as supplementary material at the OSF repository). Besides the materials, the procedure of the experiment was exactly the same as in Experiment 4.

Results and discussion

Data from the first analysis point ($N = 100$, $\alpha = 0.03$) revealed a non-significant interaction between Rhythm and Verb. For simplicity, we here just report the results from the second analysis point ($N = 120$, $\alpha = 0.02$; the script at the OSF repository allows reproducing the analysis for the first stopping point in detail). The main effect of Rhythm was significant ($\chi^2(1) = 5.69$, $p = 0.017$, $OR = 1.45$, *95% CI* = [0.98, 2.15]), with participants in the hand moving group choosing more abstract interpretations. The main effect of Verb was not significant ($\chi^2(1) = 0.28$, $p = 0.56$, $OR = 1.03$, *95% CI* = [0.67, 1.58]), as well as the key interaction between Rhythm and Verb ($\chi^2(1) = 0.84$, $p = 0.36$, $OR = 1.15$, *95% CI* = [0.85, 1.57]; Fig. 6). Participants who moved the hands did not differ in their proportion of abstract choices for hand ($M = 0.45$) and foot sentences ($M = 0.42$; $\chi^2(1) = 0.76$, $p = 0.38$, $OR = 1.18$, *95% CI* = [0.80, 1.76]). Participants who moved the feet also did not differ in their proportion of abstract choices for hand ($M = 0.35$) and foot sentences ($M = 0.34$; $\chi^2(1) = 0.02$, $p = 0.88$, $OR = 1.04$, *95% CI* = [0.65, 1.66]).

Present data failed in replicating the findings of Experiment 4. Since motor-language interactions are expected to occur in any language (e.g., Gianelli et al., 2020), there is no reason to think that conducting the experiment in English rather than in Spanish could account for the non-replication of Experiment 4. Yet, to definitely rule out this possibility, the next experiment was devoted to replicate Experiment 4 in a Spanish sample and using the original stimuli.

Experiment 7

Experiment 7 was an exact replication of Experiment 4, conducted with Spanish participants and using the original materials from Experiment 4. The study was preregistered prior to data collection: https://aspredicted.org/78pj-cbjq.pdf.

Methods

As described in detail in Experiment 5, the estimated sample size to replicate Experiment 4 with 80% power was $N = 120$. We again implemented a sequential testing approach with two stopping points: $N = 120$ at $\alpha = 0.03$, and $N = 140$ at $\alpha = 0.02$. Overall, 140 new native Spanish speakers from the same population as in Experiment 1 took part in the experiment ($M_{age} = 21$, $SD_{age} = 3.33$; 26 men; 7 left-handers). The materials and procedure of the experiment were exactly as in Experiment 4.

Results and discussion

Data from the first analysis point ($N = 120$, $\alpha = 0.03$) revealed a non-significant interaction, so here we just report the results from the second point ($N = 140$, $\alpha = 0.02$). The analysis did not yield a significant effect of Rhythm ($\chi^2(1) = 4.18$, $p = 0.04$, $OR = 0.69$, *95% CI* = [0.49, 0.69]) nor Verb ($\chi^2(1) = 5.03$, $p = 0.025$, $OR = 0.48$, *95% CI* = [0.27, 1.78]).[3] More importantly, the Rhythm x Verb interaction was also not significant ($\chi^2(1) = 0.37$, $p = 0.54$, $OR = 1.15$, *95% CI* = [0.74, 1.77]). However, the pattern of results was not completely contrary to predictions: although participants who moved their hands did not significantly differ in their proportion of choosing abstract interpretations for foot ($M = 0.28$) and hand sentences ($M = 0.22$; $\chi^2(1) = 3.39$, $p = 0.07$, $OR = 1.99$, *95% CI* = [0.96, 4.17]), participants who moved their feet exhibited a greater proportion of abstract choices for foot sentences ($M = 0.34$) than for hand sentences ($M = 0.25$; $\chi^2(1) = 5.97$, $p = 0.01$, $OR = 2.04$, *95% CI* = [1.15, 3.70]; Fig. 7).

Thus, when we engaged in the exact same procedure and materials than Experiment 4, we found partial support for our predictions, although the key interaction was not statistically replicated. As a final step in this experiment series, the next study attempted to replicate the pattern found in the short-delay experiments (Experiments 2 and 3).

Experiment 8

Experiment 8 was a direct replication of Experiment 3. We decided to replicate Experiment 3, instead of Experiment 2, to avoid introducing again the issue of the simultaneity between rhythm and sentence processing. The study was preregistered prior to data collection: https://aspredicted.org/2f5t-rxhg.pdf.

Methods

We run a power analysis of the same kind as described in detail in Experiment 5. The interaction between Rhythm and Verb in Experiment 3 had an effect size of $OR = 8.33$, which corresponds with an estimate of β = -2.12. As this effect size is quite large (Chen et al., 2010), and probably unrealistic, we decided to run our simulations considering a smaller effect size of β = -1 ($OR = 2.72$). The analysis indicated that using 45 participants per group secures more than 90% of power. Yet, as in the previous experiments, we divided alpha between two analysis points: $N = 90$ at $\alpha = 0.03$, and $N = 140$ at $\alpha = 0.02$.

Finally, 141 new native Spanish speakers from the same population as in Experiment 1 took part in the experiment. Data from one participant who knew the objective of the study were excluded. This left us with the preregistered sample size of $N = 140$ ($M_{age} = 20.74$, $SD_{age} = 3.15$; 31 men; 8 left-handers). The materials and procedure of the experiment were exactly as in Experiment 3.

Results and discussion

Data from the first stopping point ($N = 90$, $\alpha = 0.03$) revealed a non-significant interaction, so here we just report the results from the second point ($N = 140$, $\alpha = 0.02$). The interaction between Rhythm and Verb was not significant ($\chi^2(1) = 0.19$, $p = 0.67$, $OR = 0.93$, *95% CI* = [0.65, 1.31]; Fig. 8). Participants who moved the hands did not differ in their proportion of abstract choices for hand ($M = 0.23$) and foot sentences ($M = 0.29$; $\chi^2(1) = 1.51$, $p = 0.22$, $OR = 0.68$, *95% CI* = [0.36, 1.30]). Participants who moved the feet also did not differ in their proportion of abstract choices for hand ($M = 0.22$) and foot sentences ($M = 0.28$; $\chi^2(1) = 0.92$, $p = 0.34$, $OR = 0.74$, *95% CI* = [0.39, 1.42]). Neither the main

[3] Both main effects may appear significant under the conventional alpha level of 0.05, but they were not significant under the alpha level used in the analysis: 0.02.

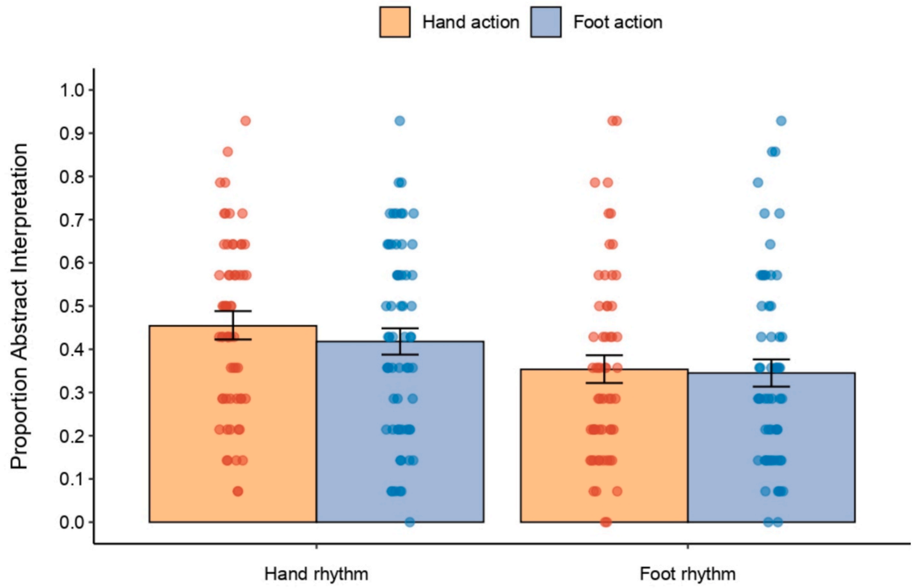

Fig. 6. Proportion of abstract interpretations in Experiment 6 as a function of Rhythm and Verb. Jittered points represent the individual mean of each participant. Error bars represent the 95% CI around the mean.

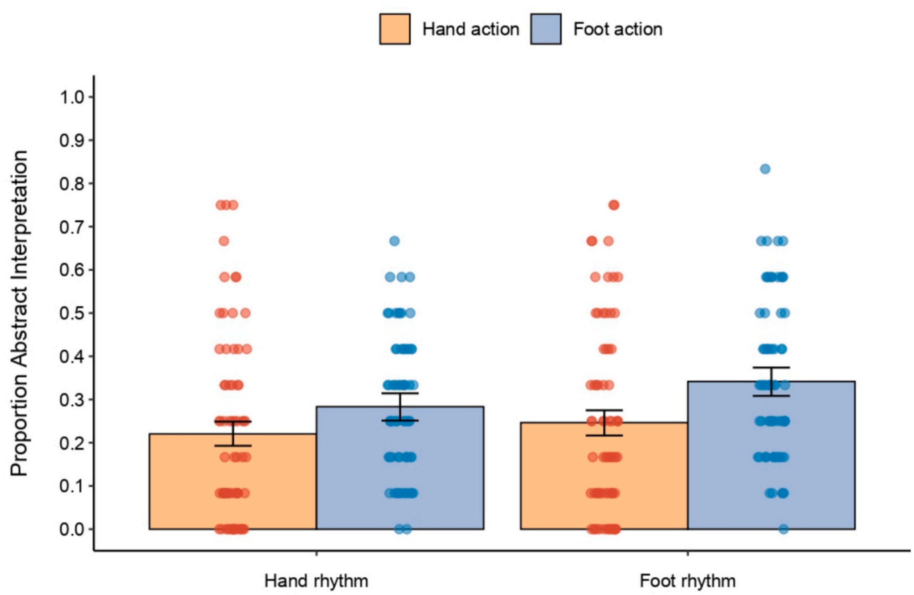

Fig. 7. Proportion of abstract interpretations in Experiment 7 as a function of Rhythm and Verb. Jittered points represent the individual mean of each participant. Error bars represent the 95% CI around the mean.

effect of Rhythm ($\chi^2(1) = 0.14$, $p = 0.71$, $OR = 1.12$, *95% CI* = [0.71, 1.77]) nor Verb ($\chi^2(1) = 1.20$, $p = 0.27$, $OR = 0.74$, *95% CI* = [0.39, 1.39]) were significant. These results failed to replicate the pattern found in the short-delay experiments.

Interim summary

Altogether, the results of this experiment series provide mixed support for the idea that motor interference changes meaning from action language, with some studies rendering significant effects and others failing to find them (see Table 1 for a summary of the design and results of the experiments). Therefore, no clear conclusion can be drawn at this point. The next sections aim to understand the reasons behind this mixed pattern of results and provide a comprehensive interpretation of them.

Towards an explanation for the instability of the effect

Inflated false-positive rate?

A first possibility that comes to mind when an effect does not replicate is that the original reports of the effect are false positives. False positives generally arise from questionable research practices such as the optional stopping of data collection, dropping outliers, or trying several ways to analyze the data (John et al., 2012; Simmons et al., 2011). We hereby state that none of these practices were present in Experiments 1–4. Sample size for Experiment 1 was initially determined based on a rule of thumb, and increased because of participant availability. Experiments 2–4 used the same sample size as Experiment 1. No interim analyses were ever carried out before completing data collection. No outliers were dropped, and no variables other than those

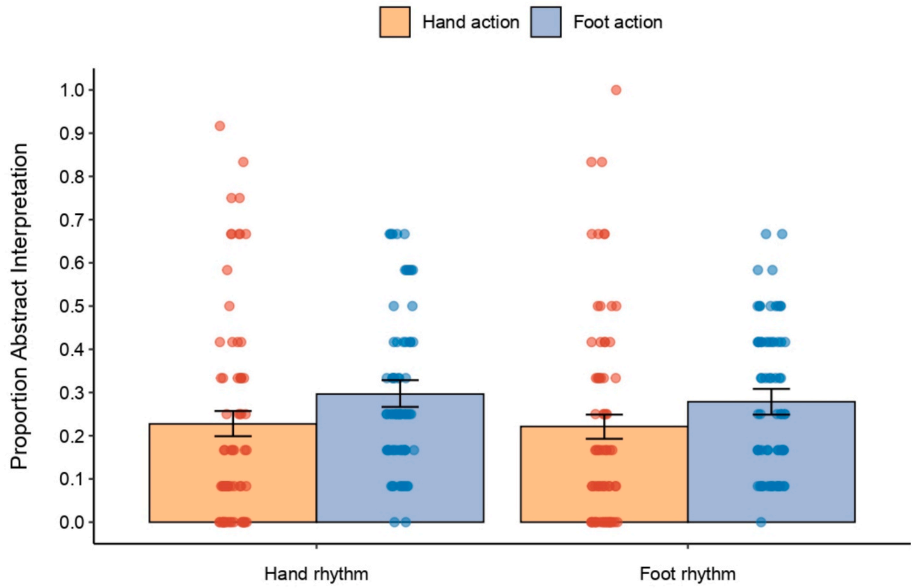

Fig. 8. Proportion of abstract interpretations in Experiment 8 as a function of Rhythm and Verb. Jittered points represent the individual mean of each participant. Error bars represent the 95% CI around the mean.

Table 1
Summary of the experimental series.

Experiment	Sample size	Language	Delay	Presentation modality	Response modality	Rhythm-sentence simultaneity	Rhythm x Verb interaction
Exp 1	$N = 90$	Spanish	Long	Visual	Manual	No	Significant
Exp 2	$N = 90$	Spanish	Short	Auditory	Vocal	Yes	Significant
Exp 3	$N = 90$	Spanish	Short	Auditory	Vocal	No	Significant
Exp 4	$N = 90$	Spanish	Long	Auditory	Vocal	No	Significant
Exp 5	$N = 120$	Spanish	Long	Auditory	Vocal	Yes	Not significant
Exp 6	$N = 120$	English	Long	Auditory	Vocal	No	Not significant
Exp 7	$N = 140$	Spanish	Long	Auditory	Vocal	No	Not significant
Exp 8	$N = 140$	Spanish	Short	Auditory	Vocal	No	Not significant

reported were measured or manipulated. Most importantly, data were analyzed exactly, and only, as reported. In fact, despite Experiments 1–4 not being preregistered, we employed the same analytic pipeline preregistered for Experiments 5–8.

Still, when working with mixed-effects models, the choice of the final model for analysis can affect the rate of false positives, as a suboptimal choice can inflate Type I error (Bates et al., 2015b). Did our model selection increase Type I error, thereby yielding false positives in Experiments 1–4? To test for this possibility, we took advantage of the statistical principle that, in absence of effect, the rate of significant results must coincide with the alpha level assumed by the researcher (generally 5%; Cumming, 2008). Following this logic, we used the R package *simr* (Green & MacLeod, 2016) to run 1000 simulations per experiment, using their corresponding data and analysis model, but setting the size of the key interaction between Rhythm and Verb to zero ($OR = 1$; $\beta = 0$), and computed the proportion of times that there was a significant result in that interaction. The alpha level for each model was adjusted to the level used in the final analysis point of each experiment: $\alpha = 0.05$ for Exps.1–4, $\alpha = 0.03$ for Exp. 5, and $\alpha = 0.02$ for Exps. 6–8. Finding a number of significant results greater than alpha would indicate that the selected model inflates the likelihood of false positives.

The results of the simulations are detailed in Table 2. As it can be seen, the proportion of false positives in each experiment was very close to the intended alpha level, which was consistently captured within the 95% CI around the estimate. These results confirm that model selection was done properly, and that the probability of finding a significant interaction in Experiments 1–4 just by chance was not larger than alpha. There still remains the possibility that Experiments 1–4 were false positives (i.e., that they all fell into the 0.05 probability of rejecting the null hypothesis when true), but the likelihood of that scenario is exceedingly unlikely (exactly $0.05^4 = 0.00000625$). Even if we take into account the whole experiment series, the probability of obtaining four false positives out of eight experiments when there is no underlying effect is just 0.00036.

Table 2
Results of the simulations of Experiments 1–8 assuming a null effect size ($OR = 1$; $\beta = 0$).

Experiment	Assumed false-positive rate	Estimated false-positive rate	95% CI for the false-positive rate estimation
Experiment 1	5%	4.00%	2.87––5.41%
Experiment 2	5%	4.10%	2.96 – 5.52%
Experiment 3	5%	5.20%	3.91 – 6.76%
Experiment 4	5%	5.00%	3.73 – 6.54%
Experiment 5	3%	2.60%	1.71 – 3.79%
Experiment 6	2%	1.30%	0.69 – 2.21%
Experiment 7	2%	2.30%	1.46 – 3.43%
Experiment 8	2%	1.50%	0.84 – 2.46%

Insufficient power to detect a small effect?

Another possibility behind the instability of the effect could be that the present experiments were not correctly powered. Although we carried out power analyses to secure at least 80% power, it is well known that seminal studies tend to overestimate real effect sizes (Klein et al., 2018; Open Science Collaboration, 2015). Thus, if the real size of the pursued effect is smaller than those found in our significant experiments, then the sample sizes estimated by the subsequent power analyses might not be sufficient to reliably detect the effect. Moreover, although the sequential sampling strategy that we followed facilitates collecting data efficiently and without increasing the false-positive rate, this strategy is also known to reduce power, as it reduces the alpha level at each testing point, which therefore reduces the chances of finding a significant effect (see Lakens, 2014).

To test for this possibility, we ran a new series of simulations, but assuming an effect size of $d = 0.2$ (corresponding to an $OR = 1.437$)[4] and adjusting the alpha level according to the final alpha used in each experiment. We decided to use a $d = 0.2$ because of two reasons. First of all, if moving a body part has any influence on meaning construction, we reasoned that the effect should be small, consistent with previous findings in the embodiment literature (e.g., see the *meta*-analysis by Winter et al., 2022: average $d = 0.15$). An effect size of $d = 0.2$ is considered a small effect accordingly to the most extended benchmarks for interpreting effect sizes (Cohen, 1988). Second, some large-scale studies have estimated that the average effect size in psychology is around $d = 0.2$ (Klein et al., 2018; Schäfer & Schwarz, 2019; but see Brysbaert, 2019), so it is reasonable that the pursued effect has a similar size.

The results are presented in detail in Table 3. As it can be observed, under these specifications, current experiments are far from the recommended 80% power (Button et al., 2013). Specifically, their mean estimated power is just 41.79%: 41.55% for the experiments with significant interactions (Exps. 1–4) and 42.03% for the null experiments (Exps. 5–8). This means that, if the effect is as small as $d = 0.2$, the probability of observing a significant result in Experiments 5–8 was even smaller than obtaining heads when flipping a coin. Indeed, with this power, the likelihood of observing a significant finding in all the eight experiments is just 0.00091 (the product of multiplying the eight power estimations). Also worth noticing, Experiments 5–8 were not better-powered than Experiments 1–4, despite having larger sample sizes.

Table 3
Results of the simulations of Experiments 1–8 assuming a small effect size of $d = 0.2$ ($OR = 1.437$).

Experiment	Sample size	Alpha level	Estimated power	95% CI around the power estimate
Experiment 1	90	0.05	40.40%	37.34 – 43.52%
Experiment 2	90	0.05	41.80%	38.72 – 44.93%
Experiment 3	90	0.05	42.70%	39.61 – 45.83%
Experiment 4	90	0.05	41.30%	38.23 – 44.42%
Experiment 5	120	0.03	44.30%	41.19 – 47.44%
Experiment 6	120	0.02	47.40%	44.27 – 50.55%
Experiment 7	140	0.02	40.70%	37.64 – 43.82%
Experiment 8	140	0.02	35.70%	32.73 – 38.76%

This confirms that the implementation of a sequential testing strategy, though efficient in terms of data collection, was detrimental to power. Accordingly, these results provide a plausible explanation for the current mixed pattern of results: the present experiments were not properly powered to detect a small (but likely) effect size.

How likely is the present scenario?

What are the odds of encountering such a puzzling scenario as the present one? This question can be answered with a Likelihood Ratio Test, which can be easily performed using the following app: https://shiny.ieis.tue.nl/mixed_results_likelihood/ (for details, see Lakens & Etz, 2017). Conducting this analysis under the same specifications detailed above (4/8 significant studies, average alpha of 0.033, and average power of 41.79% for an effect size of $d = 0.2$) revealed that, if there is a true effect in the data, finding four out of eight significant findings occurs in 24.65% of the cases. Despite this probability may seem not too high, finding four out of eight significant findings if there is no effect is much less likely: it only occurs in 0.00005% of the cases. In fact, the likelihood ratio between those probabilities is 4910.63: the presence of a real effect is 4910.63 times more likely than the absence of effect (ratios above 32 are considered strong evidence for H_1; Lakens & Etz, 2017).

All in all, these additional analyses converge on the existence of a true effect in our experiment series. Importantly, it is unlikely that we are overinterpreting false positives. In turn, mixed findings may be related to the small size of the effect and to having found, by chance, larger effect sizes in our first experiments, which led to insufficient power to detect it in subsequent experiments.

Integrative analysis of experiments 1–8

If the former conclusion is true and the absence of significant results in some of the studies relates to the low power to detect a small effect, we hypothesized that increasing sample size, and therefore power (e.g., Button et al., 2013), should reveal the expected effect. Thus, we performed an integrative data analysis by pooling together all our eight experiments (e.g., Curran & Hussong, 2009). Not only this secures much more power than the individual studies, but will also help to summarize our mixed findings and obtain a clearer conclusion from them (Curran & Hussong, 2009). We decided to carried out a pooled analysis, instead of a conventional *meta*-analysis, since working with the raw data provides more reliable and less biased estimates than just averaging point estimates (Blettner et al., 1999; Stewart & Parmar, 1993).

Following the pattern of results observed in Experiments 1–4, we predicted a significant interaction between the effector used to perform the rhythm and the effector implied by the actions described in the sentences, which should be further modulated by the amount of delay: in short-delay conditions, moving the effector related to the sentences should lead to more concrete interpretations, whereas in long-delay conditions, moving the effector related to the sentences should lead to more abstract interpretations.

Data analysis

We combined the data from the eight interference experiments, leading to a total sample size of 880 participants ($M_{age} = 20.58$, $SD_{age} = 2.92$; 147 men; 66 left-handers). Table 4 summarizes the data aggregated for the analysis. Then, we ran a mixed-effects model analysis following the same specifications described in detail in Experiment 1, but incorporating some changes to the maximal model. First, because we combined data from several experiments, which also used sentences in two different languages, we assigned individual identifiers to each unique Participant and Item for each Experiment and Language. Second, as we expected the interaction to change depending on the delay between sentence and interpretations, we included a new fixed factor:

[4] This conversion was performed using the formula developed by Sánchez-Meca et al. (2003): $d = \log(OR)\ x$.

Table 4
Mean proportions of abstract choices and standard deviations (in brackets) for each experiment.

Delay	Experiment	Sample size	Hand rhythm Hand sentences	Foot rhythm Foot sentences	Hand sentences	Foot sentences	Congruency effect
Long delay	Exp 1	$N = 90$	0.52 (0.50)	0.44 (0.50)	0.25 (0.43)	0.34 (0.48)	0.09 (0.48)
	Exp 4	$N = 90$	0.40 (0.49)	0.37 (0.48)	0.29 (0.45)	0.34 (0.47)	0.04 (0.46)
	Exp 5	$N = 120$	0.30 (0.46)	0.34 (0.48)	0.30 (0.46)	0.34 (0.47)	0.001 (0.47)
	Exp 6	$N = 120$	0.45 (0.50)	0.42 (0.49)	0.35 (0.48)	0.35 (0.48)	0.01 (0.49)
	Exp 7	$N = 140$	0.22 (0.41)	0.28 (0.41)	0.25 (0.43)	0.34 (0.43)	0.02 (0.42)
	Total	$N = 560$	0.37 (0.47)	0.36 (0.47)	0.29 (0.45)	0.34 (0.47)	0.03 (0.46)
Short delay	Exp 2	$N = 90$	0.16 (0.36)	0.39 (0.49)	0.45 (0.50)	0.26 (0.44)	−0.21 (0.45)
	Exp 3	$N = 90$	0.22 (0.41)	0.48 (0.50)	0.43 (0.50)	0.29 (0.45)	−0.20 (0.47)
	Exp 8	$N = 140$	0.23 (0.42)	0.30 (0.47)	0.22 (0.42)	0.28 (0.45)	−0.06 (0.44)
	Total	$N = 320$	0.20 (0.40)	0.37 (0.49)	0.35 (0.47)	0.28 (0.45)	−0.16 (0.45)

Note: Congruency effect = match minus mismatch conditions (i.e., raw effect size of the key interaction). Match and mismatch conditions refer to those where the effector moved for the rhythm and the effector implied by the sentences were the same (hand-hand/foot-foot) or not (hand-foot/foot-hand), respectively.

Delay (short vs. long), together with all its possible interactions with Rhythm and Verb. In R notation, the maximal model was the following: *glmer(data, Choice ~ Rhythm*Verb*Delay + (1 + Rhythm|Item_Language) + (1 + Verb|Participant_Experiment), family = binomial(link = logit))*.

Results and discussion

The detailed results of the integrative analysis can be reproduced from the R script at the OSF repository. Of crucial importance, the analysis revealed a significant three-way interaction between Rhythm, Verb and Delay ($\chi^2(1) = 111.76$, $p < 0.001$, $OR = 6.38$, *95% CI* = [4.50, 8.93]). The constituent two-way interactions were also significant and showed a pattern of results consistent with our predictions (Fig. 9). When the delay was short, participants who moved the hands chose fewer abstract interpretations for hand ($M = 0.24$) than for foot sentences ($M = 0.37$; $p < 0.0001$), whereas participants who tapped with the feet chose fewer abstract interpretations for foot ($M = 0.28$) than for hand sentences ($M = 0.35$; $p < 0.0001$; interaction: $\chi^2(1) = 52.73$, $p < 0.001$, $OR = 4.62$, *95% CI* = [3.06, 6.98]). However, when the delay was long, the pattern of results partially reversed: moving the feet led to more abstract choices for foot ($M = 0.34$) than hand sentences ($M = 0.29$), whereas no differences between hand ($M = 0.37$) and foot sentences ($M = 0.36$) were found in the hand-tapping group ($p = 0.82$; interaction: $\chi^2(1) = 12.64$, $p < 0.001$, $OR = 1.41$, *95% CI* = [1.17, 1.71]).

Additional Bayesian analyses provided converging support for these key interactions—following the interpretation criteria proposed by van Doorn et al. (2021). The Bayes factor (BF) for the higher-order interaction between Rhythm, Verb and Delay was $BF_{10} = 2.89 \times 10^{20}$, indicating extreme support for its existence. The BF for the Rhythm x Verb interaction in the short-delay condition also provided strong evidence ($BF_{10} = 167126.6$). Finally, the Rhythm x Verb interaction in the long-delay condition was supported as well, although the evidence was moderate ($BF_{10} = 3.62$). The details of these analyses are provided in a supplementary file available in the OSF repository.

Altogether, these results support our initial hypothesis that motor interference can induce effector-specific changes in how people construe meaning from action sentences, and indicate that the direction of these effects is further modulated by the delay between the sentence and its interpretations. Moreover, these findings also support the notion that low power is a reasonable cause behind the instability of the effect, as pooling data from more than 800 participants rendered the effect observable. Notably, as argued by Wegener and colleagues (2022; see also Vosgerau et al., 2019), confidence in integrative analyses depends critically on the absence of selective reporting. In this regard, we state that all experiments conducted within this research line are fully

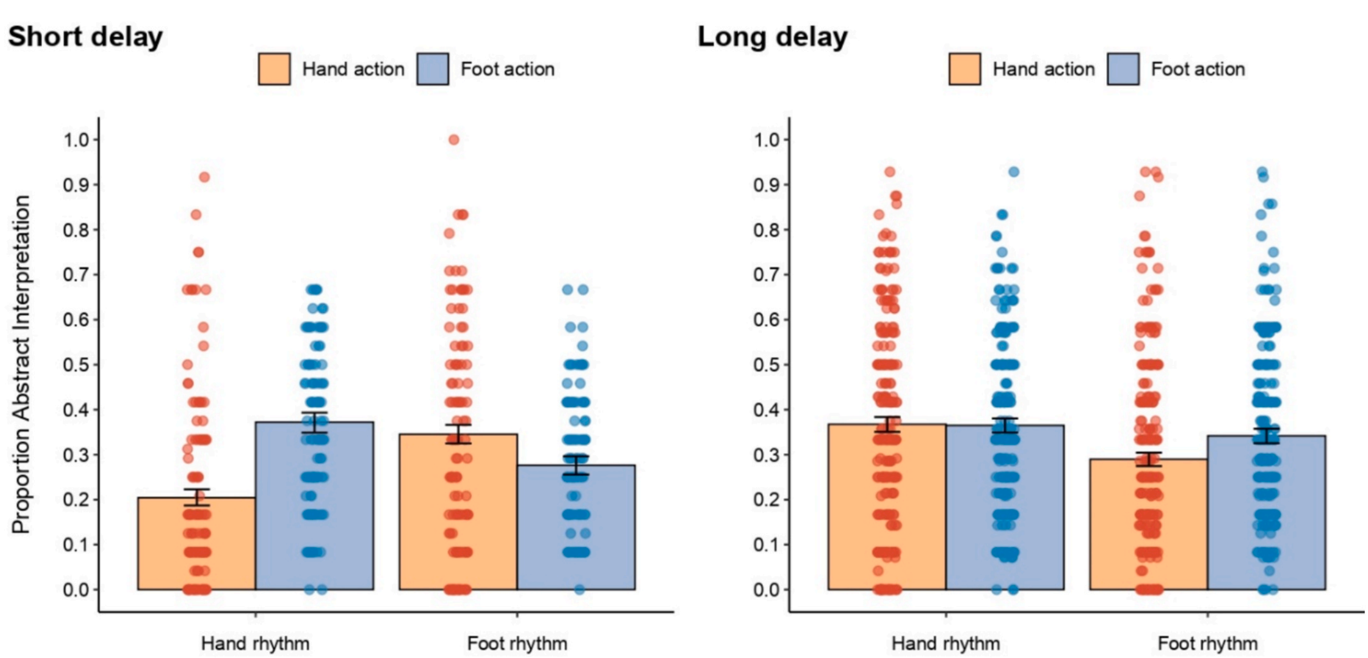

Fig. 9. Proportion of abstract interpretations in the integrative analysis of Experiments 1–8 as a function of Rhythm, Verb, and Delay. Jittered points represent the individual mean of each participant. Error bars represent the 95% CI around the mean.

reported in the present paper and included in the analysis. Accordingly, we are confident in the results yielded by the present pooled analysis. Nevertheless, since this analysis was not preregistered, it should be regarded as exploratory, its conclusions considered preliminary.

Importantly, we believe that the absence of a significant effect of moving the hands in the long-delay condition does not invalidate the effector-specificity of the present results. The significant Rhythm x Verb interaction crucially implies that moving the hands and the feet differentially affects the comprehension of hand and foot sentences. This is evidenced by the significant difference between hand and foot sentences in the foot-tapping group, which supports the notion of effector-specificity, albeit in a weaker form than would be expected from a fully crossed interaction (see Shebani & Pulvermüller, 2018). This asymmetric pattern of results may be related to people's relative inexperience with foot actions (e.g., Liu et al., 2022): while the hands are routinely used to perform complex activities such as writing or cooking, the feet are largely restricted to locomotion. Accordingly, tapping with the feet may have been demanding enough to interfere with the simulation of foot actions at long delays, whereas moving the hands might not have imposed sufficient motor demands to disrupt the simulation of hand actions. Still, the visual inspection of the results may lead some readers to argue that the long-delay interaction is better explained by a reduction in abstract interpretations of hand sentences in the foot-movement condition, thus not reflecting an effector-specific embodied effect. Although we cannot entirely rule out this possibility, we consider it unlikely. Specifically, it is unclear how or why moving one body part would selectively affect the processing of sentences referring to a different body part, in contrast to prior evidence showing that the involvement of the motor system in action language comprehension is effector-specific (Hauk et al., 2004; Pulvermüller et al., 2005; Shebani & Pulvermüller, 2013).

General discussion

In eight experiments, we tested whether manipulating motor activity can qualitatively change how people interpret language about actions. The results of the individual experiments were mixed, with some yielding significant results, and other yielding non-significant results. However, subsequent simulations suggested that significant findings were unlikely to be false positives. Instead, it is likely that the null results were the result of insufficient statistical power to detect a small effect. Supporting this idea, pooling the data from the eight experiments showed effector-specific effects of motor action on meaning construction. These effects were further modulated by the amount of delay between the sentences and their interpretations. When the delay was short (200 ms), participants who tapped a complex rhythm with their hands chose more concrete interpretations of hand actions (compared to foot actions), whereas participants assigned to tap with their feet chose more concrete interpretations of foot actions (compared to hand actions). In contrast, when the delay was long (15 s), this pattern partially reversed, with participants who tapped with their feet choosing more abstract interpretations for foot than for hand sentences.

Although these findings should be taken with caution in light of the inconsistency across experiments, we have provided converging reasons to believe that there is a true effect of motor action on meaning construction. The effect, however, is probably small and requires large samples to be detected. In any case, the conclusions presented below should be treated as preliminary and would benefit from future extensions and replications.

Integrating previous findings

The present study provides the first evidence that modulating motor activity can produce qualitative changes in sentence understanding, leading to more concrete or abstract conceptualizations of the same actions. Previous studies using neurostimulation or motor interference interventions have suggested that neural systems for planning and performing actions are causally involved in determining how quickly or accurately people can respond to action words and sentences, or hold them in memory (e.g., Bidet-Ildei et al., 2017; Gijssels et al., 2018; Glenberg et al., 2008; Pulvermüller et al., 2005; Shebani & Pulvermüller, 2013; Willems et al., 2011). Here we show that, beyond affecting speed or accuracy, interventions on the motor system can influence the kinds of meanings that language users construe, thus providing additional support for the causality of motor activity in language understanding.

These results are compatible with, and complementary to, those of Togato and colleagues. Whereas Togato et al. (2021) focused on the preference towards the two meanings of single-word homographs, here we focused on the level of abstraction at which action sentences can be construed. But how can they be reconciled with our previous TMS study (Solana et al., 2024), in which we found that inhibiting primary motor cortex (M1) did not change meaning construction? This dissociation may offer important insights into the neural bases of the effect. Some studies have argued that motor simulations are primarily sustained by premotor circuits, rather than primary motor regions (Gijssels et al., 2018; Willems et al., 2010b, 2011). Other action-related regions beyond the primary motor system—such as the supplementary motor area (SMA), the cerebellum, or the inferior frontal gyrus (IFG)—have also been implicated in action language comprehension (Aziz-Zadeh et al., 2006; García et al., 2017; Moreno et al., 2015). As tapping a complex rhythm likely engages the motor system in a stronger and broader way than just stimulating a part of it (e.g., MacRae & Matheson, 2025), present findings may reflect the workings of regions other than M1, or the coupling between M1 and other parts of the motor system.

Our results may also be relevant to assist the interpretation of previous behavioral findings. While several studies have reported significant and effector-specific effects of secondary motor task (e.g., tapping) on action language processing (e.g., Glenberg et al., 2008; Shebani & Pulvermüller, 2013; Togato et al., 2021), many others have not (e.g., Montero-Melis et al., 2022; Postle et al., 2013; te Rietmolen et al., 2025). Present results suggest a possible reconciliation between these conflicting results. As indicated by our simulations and integrative analyses, the effects of motor action on language comprehension may indeed exist, but are likely to be small, thus requiring large samples and cumulative evidence beyond individual experiments to be reliably detected. However, this conclusion raises a fundamental question: if the effects are so small and hard to replicate, how relevant is the influence of the motor system on semantics? The answer will benefit from expert consensus regarding the range of effect sizes that should be considered theoretically and practically meaningful (Anvari et al., 2023; Götz et al., 2022).

Mechanisms of the effect

We believe that the current results are consistent with our hypothesis that neural circuits for performing motor actions help language users to mentally represent the concrete, motoric details of actions described in language. Accordingly, altering the activity of these circuits with a secondary motor task affects how abstractly/concretely meaning is construed. Nevertheless, this study should be regarded as a proof of concept: our goal was just to establish the existence of the effect. The specific neurocognitive mechanisms underlying the phenomenon, as well as those mediating its flexibility and direction, remain open questions for future research. Here we present some possibilities, although they should be only treated as speculative.

One possibility is that motor simulations gradually activate the motor neural circuits over time (e.g., Barsalou et al., 2008). In the initial tens or hundreds of milliseconds, language-driven activity in the motor system may still be ramping up. If so, when the delay is short (200 ms), the activation produced by moving a relevant effector boosts the activity of the motor system, thereby helping the computation of low-level details and yielding more concrete representations of the actions. However, as times goes by, the motor system becomes more and more involved in the simulation. Accordingly, when the delay is long (15 s), the activation generated by moving a relevant effector competes for the same neural resources that sustain the simulation. As a result, the computation of concrete information is hampered, which biases participants towards less motoric interpretations of the actions.

This interpretation is consistent with proposals and findings supporting that simulation is a constructive process that requires time to develop (Barsalou et al., 2008; MacRae & Matheson, 2025; Moreno et al., 2015). Importantly, this should not be considered against findings showing that motor simulations emerge very quickly, around 200 ms after word onset (e.g., Hauk & Pulvermüller, 2004; Klepp et al., 2014): simulations may arise very fast but continue to develop at later stages, as suggested by recent EEG studies (Harpaintner et al., 2022; MacRae & Matheson, 2025). This consideration is particularly relevant for our study, which used sentences rather than isolated words: sentences likely require further semantic elaboration and integration, potentially resulting in prolonged motor simulations (e.g., Moreno et al., 2015).

Another possibility concerns the fact that varying the delay between sentence and interpretations also affected the onset of the motor task. In Experiments 1 and 4 (long delay), the rhythm began after sentence presentation; in Experiments 2 and 3 (short delay), it began before. This raises an alternative explanation for the observed results: initiating movement before the sentence might preactivate the motor system and facilitate the simulation of motor details, whereas initiating it afterward—while the simulation is taking place—could hinder it (akin to Boulenger et al., 2006). However, Experiments 2 and 3 already compared concurrent vs. non-concurrent rhythm and sentence processing, keeping delay constant, and yielded consistent results. Thus, although we cannot completely discard that possibility, we believe that differences in the shape of the interaction are more likely due to delay duration than rhythm onset.

Uncovering the mechanisms that underpin the present findings might benefit from the addition of a baseline condition, which would allow disentangling whether we are observing a match effect, a mismatch effect, or both. However, it is difficult to design a truly neutral baseline in the present case. Moving an alternative effector (e.g., the lips) does not differ from moving the hands when understanding a foot-action sentence, or vice versa (i.e., both are mismatch conditions). Not moving any effector does not activate the motor system yet fails to replicate the non-motor components of the rhythm task (e.g., increased cognitive load or divided attention), which may in turn affect the overall preference for concrete vs. abstract interpretations, confounding comparisons to the baseline. Another avenue for future research may be the inclusion of physiological measures of motor activation (e.g., Nazir et al., 2017). This would serve as a manipulation check for the effectiveness of the motor system manipulation and would assist the interpretation of the changes in the direction of the effect. Brain measures with great temporal resolution, such as M/EEG, will be also crucial to understand the temporal dynamics of the effect.

Still, regardless of the exact mechanisms behind the effect, we believe that present findings align with the "embodied" hypotheses that (1) there are shared substrates for language semantics and motor action; (2) the motor system plays a causal role in language understanding; and (3) motor-language interactions are flexible and change along the temporal course of language comprehension (e.g., Barsalou, 2008; García & Ibáñez, 2016; Pulvermüller, 2005; Willems & Casasanto, 2011). However, before accepting these conclusions, we should consider whether there could be "disembodied" explanations of the present results.

A first alternative is priming: Performing actions with the hands/feet could direct attention to the relevant part of the body, causing *amodal* information about hand/foot actions to be more accessible in memory, and thus yielding more concrete interpretations of hand/foot sentences. Yet, although this account can provide an explanation for the results of the short-delay experiments, it hardly accounts for the findings from the long-delay experiments, in which moving an effector enhanced abstract interpretations for actions that involve that effector. Therefore, although we cannot discard that priming plays a role, current findings fit better with the "embodied" idea of a common neural basis for meaning computation and motor action. Moreover, it should be noted that some studies have found that priming also involve the recruitment of motor circuits (e.g., Grisoni et al., 2016), so even if priming was the cause of the observed effects, it would not necessarily imply disembodied processes.

Another possibility is that the apparent transparency between the manipulation (moving the hands/feet) and what was measured (interpretation of hand/foot sentences) made participants aware of the objective of the study and encouraged them to respond in a strategic manner (e.g., Mahon & Caramazza, 2008). In anticipation of this issue, we manipulated the effector of the rhythm between groups, making it unlikely that participants could infer the relationship between the rhythm and the sentences. Debriefings after the experiments supported this idea, as most participants reported believing that performing the rhythm was merely a distractor task. Furthermore, this explanation cannot account for the observed changes in the effect as a function of delay.

Participants may also have voluntarily imagined the actions, rather than implicitly simulating them, which is known to recruit the motor system as well (e.g., Willems et al., 2010b). We consider this explanation unlikely since participants were required to tap a difficult rhythm simultaneously or immediately after being presented with the sentences, so their attentional resources were arguably mostly focused on performing the rhythm. According to Shebani and Pulvermüller (2013, p. 4): "this exercise is difficult, even for musicians not experienced with drumming".

Finally, any account that fails in considering the effector-specificity of the effect—e.g., an overall effect of delay on construal level (e.g., Trope & Liberman, 2000) or a general effect of tapping in language comprehension (e.g., te Rietmolen et al., 2025)—cannot provide a full explanation for the present findings either.

Implications for theories of semantic representation

A central debate within cognitive science concerns the format of semantic representations. On the one hand, classic cognitive theories conceive semantic representations as abstract and amodal symbols detached from bodily experience (e.g., Fodor, 1975; Pylyshyn, 1984). On the other hand, embodied accounts propose that semantics are intimately linked to bodily experience, and that meaning is achieved through experiential simulations implemented in modality-specific brain circuits (e.g., Barsalou, 2008; Pulvermüller, 2005). Previous efforts to move beyond this dichotomy have focused on the time course and neural circuits over which representations develop. Several proposals converge on the suggestion that a comprehender's earliest response to words consists in some sort of shallow understanding that arises "in the linguistic system" (Barsalou, 2008, p. 622), and that "deep conceptual processing" (ibid), which relies on the simulation system, arises subsequently (see also Connell, 2019; Louwerse, 2011). This view faces several challenges, however, including results suggesting that simulations may not play any detectable role in conceptualization until several hundreds or thousands of milliseconds after stimulus presentation (Simmons et al., 2008; Vignali et al., 2023), far slower than words' meanings are typically processed (e.g., Kutas & Hillyard, 1980; MacGregor et al., 2012).

Present results may suggest a potential reconciliation between "embodied" and "disembodied" representations that does not rely on timing or on dividing between neurocognitive systems for language and simulation, but rather on the experiential context in which verbal cues are processed. Context has been called "a sleeping giant in the discussion on embodiment… likely to be a major factor in mediating the relative impact of [embodied] and [disembodied] representations" (Zwaan, 2014, p. 230). The present experiments rely on the fact that people's bodies, and the actions they perform with them, are a part of the context in which they use their minds (Casasanto, 2011). If we consider that more concrete interpretations are more embodied than abstract ones (i.e., more reliant on sensorimotor information; Vallacher & Wegner, 1987), then our results suggest that semantic representations shift between more embodied and more disembodied forms—even when timing is held constant within experiments. What determines their final form is, in our view, the context in which the sentences are processed: when the context included actions performed by the same effectors that would carry out the described actions, then the construed meanings are qualitatively different to those that emerge when the performed and described actions involved different effectors.

In this interplay between context and semantics, attentional mechanisms may well play a key role. For example, when an action description is presented, all its features—whether concrete (sensorimotor details) or abstract (goals, consequences)—may become active, generating a rich semantic space that can support multiple possible interpretations of the sentence. The context created by participants' bodies may then increase the salience of some features over others, directing attention toward them. Ultimately, this would make these features more likely to be incorporated into the final interpretation made by language comprehenders.

Importantly, this proposal does not dismiss that timing plays a relevant role, as suggested by the differences observed between short- and long-delay experiments. However, present results crucially showed evidence for motor simulation in both early ($\approx$ 200 ms) and (very) late ($\approx$ 15 s) processing stages. Rather than supporting a time-dependent distinction between embodied and disembodied processes (e.g., Barsalou et al., 2008), we believe that current results better align with the idea that motor-language interactions are flexible, so that they can vary in form and relevance depending on the constraints of the context (Ibáñez et al., 2023; Shebani & Pulvermüller, 2018; Willems & Casasanto, 2011). In this line, it is noteworthy that other contextual factors beyond delay (e.g., language modality) do not appear to exert strong influences on the effect. Still, this possibility should be more exhaustively tested in future studies, which might include manipulations not addressed here, such as varying the depth of the task (e.g., Vukovic et al., 2017) or presenting the sentences in a second language (e.g., Monaco et al., 2021).

Another longstanding challenge to theories of embodied cognition is abstraction (Chatterjee, 2010; Casasanto & Gijssels, 2015). If thoughts about actions are implemented in motor simulations, how can people conceptualize actions abstractly enough to generalize across diverse instances, or to understand actions in terms of their goals and outcomes rather than their mechanics? The suggestion that simulations are schematic, lacking fine-grained details (Barsalou, 2008), does not fully address this question. If we conceptualize voting as "supporting a candidate" or "influencing an election," these actions do not correspond to schematic hand movements. Arguably, they do not correspond to any hand movements at all. Present results suggest that the context created by one's own motor activity can lead to abstraction: Different conceptions of the same action vary in their level of abstraction because they vary in the extent to which they rely on effector-specific motor simulations.

These results raise a further question of critical importance to theories of embodied semantics. When people conceptualize actions as more abstract and less motorically-grounded, does this mean they are activating mental representations that are relatively disembodied? This possibility is consistent with proposals that embrace graded levels of embodiment (Chatterjee, 2010; Zwaan, 2014): Perhaps more abstract means less embodied (i.e., less reliant on simulations). Alternatively, abstract action representations could continue to rely on simulations in other neural systems. Which other systems? Besides perceptual and motor systems, theories of embodied cognition also posit that concepts can be grounded in systems for interoception, emotion, and social cognition (Borghi et al., 2025; Kousta et al., 2011; Pexman et al., 2023; Vigliocco et al, 2014). The physical concrete act of "making an X on a ballot" may be primarily motoric, but the abstract notion goal of "influencing an election" is likely to be affectively charged, and to have important social implications. Perhaps motor interference does not affect simulations in the motor system, but rather shifts them to neural systems that enable us to socialize and experience emotions and other internal states.

Concluding remarks

Science is a cumulative process; even if not all of the studies in a literature show significant results, a true effect may exist (Cumming, 2014). In fact, "sets of studies that contain only significant findings can be too good to be true" (Lakens & Etz, 2017, p. 880). However, in the idealized world of scientific publications, significant results are the norm, whereas null results rarely see the light (Bakker et al., 2012). The present experimental series provides a realistic illustration of how science typically works, and underscores the importance of publishing and integrating all research outcomes within a series of experiments to achieve a more comprehensive understanding of the topic under investigation.

In sum, although present results should be taken with caution, they provide novel evidence that meaningless motor activity can cause qualitative changes in action language comprehension: Performing different actions can lead to different understandings of the same sentences. This provides additional evidence for the idea that the motor system plays a causal role in language comprehension. Yet, further research is needed to clarify the reliability of present findings, as well as their cognitive and neural mechanisms. Further studies are also needed to test whether interfering with simulations in other neural systems (e.g., visual, auditory, or interoceptive) can also change meaning, and to test whether people with different bodies and sensorimotor experiences (e.g., patients with motor impairment) interpret language differently.

Declaration of AI use in the writing process

AI was used exclusively to improve the readability and language of the manuscript.

CRediT authorship contribution statement

Pablo Solana: Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Omar Escámez:** Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Gabriella Vigliocco:** Writing – review & editing, Conceptualization. **Daniel Casasanto:** Writing – review & editing, Supervision, Methodology, Investigation, Funding acquisition, Conceptualization. **Julio Santiago:** Writing – review & editing, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Funding

This research was funded by grants PSI2015-67531-P to JS and DC and PID2022-142583NB-I00 to JS, DC and PS, a Leverhulme Visiting Professorship (VP1-2012–032) to JS, a James S. McDonnell Foundation Scholar Award (220020236) to DC, and a FPU grant (FPU20/01946) to PS.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We are grateful to all the members of the Grounded Cognition Lab and the Experience and Cognition Lab who contributed to this work, especially to Celia Barnés-Castaño for recording the audio files used in the study, and to Jason Hamm for creating the English materials and leading the data collection of the US experiment. We also thank Vahid Nassiri and Francisco Garre-Frutos for their statistical advice. This work is part of the doctoral dissertation of PS in the Psychology Program of the University of Granada under the supervision of JS.

Data availability

All data, analysis code, materials, and supplementary information are available in the Open Science Framework (OSF) repository: https://osf.io/58uvr/

References

Anvari, F., Kievit, R., Lakens, D., Pennington, C. R., Przybylski, A. K., Tiokhin, L., & Orben, A. (2023). Not all effects are indispensable: Psychological science requires verifiable lines of reasoning for whether an effect matters. *Perspectives on Psychological Science, 18*(2), 503–507. https://doi.org/10.1177/17456916221091565

Aziz-Zadeh, L., Wilson, S. M., Rizzolatti, G., & Iacoboni, M. (2006). Congruent embodied representations for visually presented actions and linguistic phrases describing actions. *Current Biology, 16*(18), 1818–1823. https://doi.org/10.1016/j.cub.2006.07.060

Bakker, M., Van Dijk, A., & Wicherts, J. M. (2012). The rules of the game called psychological science. *Perspectives on Psychological Science, 7*(6), 543–554. https://doi.org/10.1177/1745691612459060

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

Barsalou, L. W., Santos, A., Simmons, W. K., & Wilson, C. D. (2008). Language and simulation in conceptual processing. In M. de Vega, A. Glenberg, A. Graesser, & Arthur (Eds.), *Symbols and Embodiment: Debates on meaning and cognition* (pp. 245–284). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199217274.003.0013.

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*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015b). Parsimonious mixed models. *arXiv*, arXiv:1506. https://doi.org/arXiv:1506.04967.

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

Bidet-Ildei, C., Meugnot, A., Beauprez, S.-A., Gimenes, M., & Toussaint, L. (2017). Short-term upper limb immobilization affects action-word understanding. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(7), 1129–1139. https://doi.org/10.1037/xlm0000373

Blettner, M., Sauerbrei, W., Schlehofer, B., Scheuchenpflug, T., & Friedenreich, C. (1999). Traditional reviews, meta-analyses and pooled analyses in epidemiology. *International Journal of Epidemiology, 28*(1), 1–9. https://doi.org/10.1093/ije/28.1.1

Borghi, A. M. (2022). Concepts for which we need others more: The case of abstract concepts. *Current Directions in Psychological Science, 31*(3), 238–246. https://doi.org/10.1177/09637214221079625

Borghi, A. M., Mazzuca, C., & Tummolini, L. (2025). The role of social interaction in the formation and use of abstract concepts. *Nature Reviews Psychology, 4*, 470–483. https://doi.org/10.1038/s44159-025-00451-z

Boulenger, V., Roy, A. C., Paulignan, Y., Deprez, V., Jeannerod, M., & Nazir, T. A. (2006). Cross-talk between language processes and overt motor behavior in the first 200 msec of processing. *Journal of Cognitive Neuroscience, 18*(10), 1607–1615. https://doi.org/10.1162/jocn.2006.18.10.1607

Brysbaert, M. (2019). How many participants do we have to include in properly powered experiments? a tutorial of power analysis with reference tables. *Journal of Cognition, 2*(1), 16. https://doi.org/10.5334/joc.72

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14*(5), 365–376. https://doi.org/10.1038/nrn3475

Casasanto, D. (2011). Different bodies, different minds: The body specificity of language and thought. *Current Directions in Psychological Science, 20*(6), 378–383. https://doi.org/10.1177/0963721411422058

Casasanto, D. (2023). *Embodied Semantics*. In F. T. Li (Ed.), *Handbook of Cognitive Semantics*. Brill: Leiden.

Casasanto, D., & Gijssels, T. (2015). What makes a metaphor an embodied metaphor? *Linguistic Vanguard, 1*(1), 327–337. https://doi.org/10.1515/lingvan-2014-1015

Chatterjee, A. (2010). Disembodying cognition. *Language and Cognition, 2*(1), 79–116. https://doi.org/10.1515/LANGCOG.2010.004

Chen, R. M. M. F., Classen, J., Gerloff, C., Celnik, P., Wassermann, E. M., Hallett, M., & Cohen, L. G. (1997). Depression of motor cortex excitability by low-frequency transcranial magnetic stimulation. *Neurology, 48*(5), 1398–1403. https://doi.org/10.1212/WNL.48.5.1398

Chen, H., Cohen, P., & Chen, S. (2010). How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. *Communications in Statistics - Simulation and Computation, 39*(4), 860–864. https://doi.org/10.1080/03610911003650383

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences. L. Erlbaum Associates.*

Connell, L. (2019). What have labels ever done for us? the linguistic shortcut in conceptual processing. *Language, Cognition and Neuroscience, 34*(10), 1308–1318. https://doi.org/10.1080/23273798.2018.1471512

Cumming, G. (2008). Replication and *p* intervals: *P* values predict the future only vaguely, but confidence intervals do much better. *Perspectives on Psychological Science, 3*(4), 286–300. https://doi.org/10.1111/j.1745-6924.2008.00079.x

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science, 25*(1), 7–29. https://doi.org/10.1177/0956797613504966

Curran, P. J., & Hussong, A. M. (2009). Integrative data analysis: The simultaneous analysis of multiple data sets. *Psychological Methods, 14*(2), 81.

Cutler, A., & Clifton, C., Jr (1999). Comprehending spoken language: A blueprint of the listener. In C. M. Brown, & P. Hagoort (Eds.), *The Neurocognition of Language* (pp. 123–166). Oxford University Press.

de Vega, M., Moreno, V., & Castillo, D. (2013). The comprehension of action-related sentences may cause interference rather than facilitation on matching actions. *Psychological Research Psychologische Forschung, 77*(1), 20–30. https://doi.org/10.1007/s00426-011-0356-1

Fischer, M. H., & Zwaan, R. A. (2008). Embodied language: A review of the role of the motor system in language comprehension. *Quarterly Journal of Experimental Psychology, 61*(6), 825–850. https://doi.org/10.1080/17470210701623605

Fodor, J. A. (1975). *The language of thought*. Harvard University Press.

Fox, J., & Weisberg, S. (2019). *An R Companion to Applied Regression. SAGE*.

García, A. M., Abrevaya, S., Kozono, G., Cordero, I. G., Córdoba, M., Kauffman, M. A., & Ibáñez, A. (2017). The cerebellum and embodied semantics: Evidence from a case of genetic ataxia due to STUB1 mutations. *Journal of Medical Genetics, 54*(2), 114–124. https://doi.org/10.1136/jmedgenet-2016-104148

García, A. M., & Ibáñez, A. (2016). A touch with words: Dynamic synergies between manual actions and language. *Neuroscience & Biobehavioral Reviews, 68*, 59–95. https://doi.org/10.1016/j.neubiorev.2016.04.022

Gianelli, C., Kühne, K., Presti, S. L., Mencaraglia, S., & Dalla Volta, R. (2020). Action processing in the motor system: Transcranial magnetic Stimulation (TMS) evidence of shared mechanisms in the visual and linguistic modalities. *Brain and Cognition, 139*, Article 105510. https://doi.org/10.1016/j.bandc.2019.105510

Gijssels, T., Ivry, R. B., & Casasanto, D. (2018). tDCS to premotor cortex changes action verb understanding: Complementary effects of inhibitory and excitatory stimulation. *Scientific Reports., 8*(1), 11452. https://doi.org/10.1038/s41598-018-29600-6

Glenberg, A. M., Sato, M., & Cattaneo, L. (2008). Use-induced motor plasticity affects the processing of abstract and concrete language. *Current Biology, 18*(7), R290–R291. https://doi.org/10.1016/j.cub.2008.02.036

Götz, F. M., Gosling, S. D., & Rentfrow, P. J. (2022). Small effects: The indispensable foundation for a cumulative psychological science. *Perspectives on Psychological Science, 17*(1), 205–215. https://doi.org/10.1177/1745691620984483

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

Grisoni, L., Dreyer, F. R., & Pulvermüller, F. (2016). Somatotopic semantic priming and prediction in the motor system. *Cerebral Cortex, 26*(5), 2353–2366. https://doi.org/10.1093/cercor/bhw026

Harpaintner, M., Trumpp, N. M., & Kiefer, M. (2022). Time course of brain activity during the processing of motor-and vision-related abstract concepts: Flexibility and task dependency. *Psychological Research Psychologische Forschung, 86*(8), 2560–2582. https://doi.org/10.1007/s00426-020-01374-5

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron, 41*(2), 301–307. https://doi.org/10.1016/S0896-6273(03)00838-9

Hauk, O., & Pulvermüller, F. (2004). Neurophysiological distinction of action words in the fronto-central cortex. *Human Brain Mapping, 21*(3), 191–201. https://doi.org/10.1002/hbm.10157

Ibáñez, A., Kühne, K., Miklashevsky, A., Monaco, E., Muraki, E., Ranzini, E., & Tuena, C. (2023). Ecological meanings. a consensus paper on individual differences and contextual influences in embodied language. *Journal of Cognition, 6*(1), 59. https://doi.org/10.5334/joc.228

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*(4), 434–446. https://doi.org/10.1016/j.jml.2007.11.007

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science, 23*(5), 524–532. https://doi.org/10.1177/0956797611430953

Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Jr, Alper, S., & Sowden, W. (2018). Many Labs 2: Investigating variation in replicability across samples and settings. *Advances in Methods and Practices in Psychological Science, 1*(4), 443–490. https://doi.org/10.1177/2515245918810225

Klepp, A., Weissler, H., Niccolai, V., Terhalle, A., Geisler, H., Schnitzler, A., & Biermann-Ruben, K. (2014). Neuromagnetic hand and foot motor sources recruited during action verb processing. *Brain and Language, 128*(1), 41–52. https://doi.org/10.1016/j.bandl.2013.12.001

Kousta, S. T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: Why emotion matters. *Journal of Experimental Psychology: General, 140*(1), 14. https://doi.org/10.1037/a0021446

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*(4427), 203–205.

Lakens, D. (2014). Performing high-powered studies efficiently with sequential analyses. *European Journal of Social Psychology, 44*(7), 701–710. https://doi.org/10.1002/ejsp.2023

Lakens, D., & Etz, A. J. (2017). Too true to be bad: When sets of studies with significant and nonsignificant findings are probably true. *Social Psychological and Personality Science, 8*(8), 875–881. https://doi.org/10.1177/1948550617693058

Levelt, W. J. (1999). Producing spoken language: A blueprint of the speaker. In C. M. Brown, & P. Hagoort (Eds.), *The Neurocognition of Language* (pp. 123–166). Oxford University Press.

Liu, Y., Caracoglia, J., Sen, S., Freud, E., & Striem-Amit, E. (2022). Are reaching and grasping effector-independent? Similarities and differences in reaching and grasping kinematics between the hand and foot. *Experimental Brain Research, 240*, 1833–1848. https://doi.org/10.1007/s00221-022-06359-x

Louwerse, M. M. (2011). Symbol interdependency in symbolic and embodied cognition. *Topics in Cognitive Science, 3*(2), 273–302. https://doi.org/10.1111/j.1756-8765.2010.01106.x

Mac Giolla, E., Luke, T. J., Calderon, S., & Ask, K. (2025). *Validating measures of Mental Abstraction. PysArXiv.*. https://doi.org/10.31234/osf.io/v6xt4

MacGregor, L. J., Pulvermüller, F., Van Casteren, M., & Shtyrov, Y. (2012). Ultra-rapid access to words in the brain. *Nature Communications, 3*(1), 711. https://doi.org/10.1038/ncomms1715

MacRae, S., & Matheson, H. E. (2025). Juggling with rubber hands, leaping with rubber feet: Sensorimotor reuse during verb comprehension. *Brain and Language, 271*, Article 105639. https://doi.org/10.1016/j.bandl.2025.105639

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.

Meteyard, L., Rodríguez-Cuadrado, S., Bahrami, B., & Vigliocco, G. (2012). Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex, 48*(7), 788–804. https://doi.org/10.1016/j.cortex.2010.11.002

Monaco, E., Jost, L. B., Lancheros, M., Harquel, S., Schmidlin, E., & Annoni, J. M. (2021). First and Second Language at Hand: A Chronometric Transcranial-magnetic Stimulation Study on Semantic and Motor Resonance. *Journal of Cognitive Neuroscience, 33*(8), 1563–1580. https://doi.org/10.1162/jocn_a_01736

Montero-Melis, G., Van Paridon, J., Ostarek, M., & Bylund, E. (2022). No evidence for embodiment: The motor system is not needed to keep action verbs in working memory. *Cortex, 150*, 108–125. https://doi.org/10.1016/j.cortex.2022.02.006

Moreno, I., De Vega, M., León, I., Bastiaansen, M., Lewis, A. G., & Magyari, L. (2015). Brain dynamics in the comprehension of action-related language. a time-frequency analysis of mu rhythms. *NeuroImage, 109*, 50–62. https://doi.org/10.1016/j.neuroimage.2015.01.018

Nazir, T. A., Hrycyk, L., Moreau, Q., Frak, V., Cheylus, A., Ott, L., & Delevoye-Turrell, Y. (2017). A simple technique to study embodied language processes: The grip force sensor. *Behavior Research Methods, 49*, 61–73. https://doi.org/10.3758/s13428-015-0696-7

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251). https://doi.org/10.1126/science.aac4716.

Pexman, P. M., Diveica, V., & Binney, R. J. (2023). Social semantics: The organization and grounding of abstract concepts. *Philosophical Transactions of the Royal Society B, 378*(1870), Article 20210363. https://doi.org/10.1098/rstb.2021.0363

Postle, N., Ashton, R., McFarland, K., & De Zubicaray, G. I. (2013). No specific role for the manual motor system in processing the meanings of words related to the hand. *Frontiers in Human Neuroscience, 7*, 11. https://doi.org/10.3389/fnhum.2013.00011

Postle, N., McMahon, K. L., Ashton, R., Meredith, M., & de Zubicaray, G. I. (2008). Action word meaning representations in cytoarchitectonically defined primary and premotor cortices. *NeuroImage, 43*(3), 634–644. https://doi.org/10.1016/j.neuroimage.2008.08.006

Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience, 6*(7), 576–582. https://doi.org/10.1038/nrn1706

Pulvermüller, F. (2013). How neurons make meaning: Brain mechanisms for embodied and abstract-symbolic semantics. *Trends in Cognitive Sciences, 17*(9), 458–470. https://doi.org/10.1016/j.tics.2013.06.004

Pulvermüller, F., Hauk, O., Nikulin, V. V., & Ilmoniemi, R. J. (2005). Functional links between motor and language systems. *European Journal of Neuroscience, 21*(3), 793–797. https://doi.org/10.1111/j.1460-9568.2005.03900.x

Pylyshyn, Z. W. (1984). *Computation and cognition*. Cambridge, MA: Bradford Books, MIT Press.

Reilly, J., Shain, C., Borghesani, V., Kuhnke, P., Vigliocco, G., Peelle, J. E., & Vinson, D. (2025). What we mean when we say semantic: Toward a multidisciplinary semantic glossary. *Psychonomic Bulletin & Review, 32*(1), 243–280. https://doi.org/10.3758/s13423-024-02556-7

Sánchez-Meca, J., Marín-Martínez, F., & Chacón-Moscoso, S. (2003). Effect-size indices for dichotomized outcomes in meta-analysis. *Psychological Methods, 8*(4), 448. https://doi.org/10.1037/1082-989X.8.4.448

Schäfer, T., & Schwarz, M. A. (2019). The meaningfulness of effect sizes in psychological research: Differences between sub-disciplines and the impact of potential biases. *Frontiers in Psychology, 10*, 813. https://doi.org/10.3389/fpsyg.2019.00813

Shebani, Z., & Pulvermüller, F. (2013). Moving the hands and feet specifically impairs working memory for arm- and leg-related action words. *Cortex, 49*, 222–231.

Shebani, Z., & Pulvermüller, F. (2018). Flexibility in language action interaction: The influence of movement type. *Frontiers in Human Neuroscience, 12*, 252. https://doi.org/10.1016/j.brainres.2012.10.004

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Simmons, W. K., Hamann, S. B., Harenski, C. L., Hu, X. P., & Barsalou, L. W. (2008). fMRI evidence for word association and situated simulation in conceptual processing. *Journal of Physiology-Paris, 102*(1), 106–119.

Solana, P., Escámez, O., Casasanto, D., Chica, A. B., & Santiago, J. (2024). No evidence for a causal role of primary motor cortex in construing meaning from language: An rTMS study. *Neuropsychologia, 196*, Article 108832. https://doi.org/10.1016/j.neuropsychologia.2024.108832

Solana, P., & Santiago, J. (2022). Does the involvement of motor cortex in embodied language comprehension stand on solid ground? a *p*-curve analysis and test for excess significance of the TMS and tDCS evidence. *Neuroscience and Biobehavioral Reviews, 141*, Article 104834. https://doi.org/10.1016/j.neubiorev.2022.104834

Solana, P., & Santiago, J. (2023). Worse than expected: A z-curve reanalysis of motor cortex stimulation studies of embodied language comprehension. *Psicológica, 44*(2), Article e15661. https://doi.org/10.20350/digitalCSIC/15661

Stewart, L. A., & Parmar, M. K. (1993). Meta-analysis of the literature or of individual patient data: Is there a difference? *The Lancet, 341*(8842), 418–422. https://doi.org/10.1016/0140-6736(93)93004-K

te Rietmolen, N., Strijkers, K., & Morillon, B. (2025). Moving rhythmically can facilitate naturalistic speech perception in a noisy environment. *Proceedings of the Royal Society B: Biological Sciences, 292*(2044), Article 20250354. https://doi.org/10.1098/rspb.2025.0354

Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., Scifo, P., Fazio, F., Rizzolatti, G., Cappa, S. F., & Perani, D. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *Journal of Cognitive Neuroscience, 17*(2), 273–281.

Togato, G., Andras, F., Miralles, E., & Macizo, P. (2021). Motor processing modulates word comprehension. *British Journal of Psychology, 112*(4), 1028–1052. https://doi.org/10.1111/bjop.12507

Trope, Y., & Liberman, N. (2000). Temporal construal and time-dependent changes in preference. *Journal of Personality and Social Psychology, 79*(6), 876–889. https://doi.org/10.1037/0022-3514.79.6.876

Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review, 117*(2), 440–463. https://doi.org/10.1037/a0018963

Vallacher, R. R., & Wegner, D. M. (1987). What do people think they're doing? Action identification and human behavior. *Psychological Review, 94*(1), 3–15.

Vallacher, R. R., & Wegner, D. M. (1989). Levels of personal agency: Individual variation in action identification. *Journal of Personality and Social Psychology, 57*, 660–671.

van Doorn, J., van den Bergh, D., Böhm, U., Dablander, F., Derks, K., Draws, T., & Wagenmakers, E. J. (2021). The JASP guidelines for conducting and reporting a Bayesian analysis. *Psychonomic Bulletin & Review, 28*, 813–826. https://doi.org/10.3758/s13423-020-01798-5

Vigliocco, G., Kousta, S. T., Della Rosa, P. A., Vinson, D. P., Tettamanti, M., Devlin, J. T., & Cappa, S. F. (2014). The neural representation of abstract words: The role of emotion. *Cerebral Cortex, 24*(7), 1767–1777. https://doi.org/10.1093/cercor/bht025

Vignali, L., Xu, Y., Turini, J., Collignon, O., Crepaldi, D., & Bottini, R. (2023). Spatiotemporal dynamics of abstract and concrete semantic representations. *Brain and Language, 243*, Article 105298. https://doi.org/10.1016/j.bandl.2023.105298

Vosgerau, J., Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2019). 99% impossible: A valid, or falsifiable, internal meta-analysis. *Journal of Experimental Psychology: General, 148*(9), 1628. https://doi.org/10.1037/xge0000663

Vukovic, N., Feurra, M., Shpektor, A., Myachykov, A., & Shtyrov, Y. (2017). Primary motor cortex functionally contributes to language comprehension: An online rTMS study. *Neuropsychologia, 96*, 222–229. https://doi.org/10.1016/j.neuropsychologia.2017.01.025

Wald, A. (2004). *Sequential analysis. Courier Corporation.*

Wegener, D. T., Fabrigar, L. R., Pek, J., & Hoisington-Shaw, K. (2022). Evaluating research in personality and social psychology: Considerations of statistical power and concerns about false findings. *Personality and Social Psychology Bulletin, 48*(7), 1105–1117. https://doi.org/10.1177/01461672211030811

Willems, R. M., & Casasanto, D. (2011). Flexibility in embodied language understanding. *Frontiers in Psychology, 2*(16), 1–11. https://doi.org/10.3389/fpsyg.2011.00116

Willems, R. M., Hagoort, P., & Casasanto, D. (2010a). Body-specific representation of action verbs: Neural evidence from right- and left- handers. *Psychological Science, 21*(1), 67–74. https://doi.org/10.1177/0956797609354072

Willems, R. M., Labruna, L., D'Esposito, M., Ivry, R., & Casasanto, D. (2011). A functional role for the motor system in language understanding: Evidence from theta-burst transcranial magnetic stimulation. *Psychological Science, 22*, 849–854. https://doi.org/10.1177/0956797611412387

Willems, R. M., Toni, I., Hagoort, P., & Casasanto, D. (2010b). Neural dissociations between action verb understanding and motor imagery. *Journal of Cognitive Neuroscience, 22*(10), 2387–2400. https://doi.org/10.1162/jocn.2009.21386

Winter, A., Dudschig, C., Miller, J., Ulrich, R., & Kaup, B. (2022). The action-sentence compatibility effect (ACE): Meta-analysis of a benchmark finding for embodiment. *Acta Psychologica, 230*, Article 103712. https://doi.org/10.1016/j.actpsy.2022.103712

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