



No support for a causal role of primary motor cortex in construing meaning from language: An rTMS study

Pablo Solana^{a,b,*}, Omar Escámez^{a,b}, Daniel Casasanto^c, Ana B. Chica^{a,b}, 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 Psychology, Cornell University, USA

ARTICLE INFO

Keywords:

Embodied cognition
Language comprehension
Motor cortex
Construal level
rTMS

ABSTRACT

Embodied cognition theories predict a functional involvement of sensorimotor processes in language understanding. In a preregistered experiment, we tested this idea by investigating whether interfering with primary motor cortex (M1) activation can change how people construe meaning from action language. Participants were presented with sentences describing actions (e.g., "turning off the light") and asked to choose between two interpretations of their meaning, one more concrete (e.g., "flipping a switch") and another more abstract (e.g., "going to sleep"). Prior to this task, participants' M1 was disrupted using repetitive transcranial magnetic stimulation (rTMS). The results yielded strong evidence against the idea that M1-rTMS affects meaning construction ($BF_{01} > 30$). Additional analyses and control experiments suggest that the absence of effect cannot be accounted for by failure to inhibit M1, lack of construct validity of the task, or lack of power to detect a small effect. In sum, these results do not support a causal role for primary motor cortex in building meaning from action language.

1. Introduction

How is language processed such that we can understand its content? And which neural systems underlie this complex operation? Traditional cognitive models of language comprehension assume that meaning computation is an amodal process that relies on abstract and arbitrary symbols (e.g., Fodor, 1975). Nonetheless, there is more and more evidence supporting the view that language is embodied. Under this view, meaning is grounded by bodily (perceptual and motor) experiences, and its comprehension stands on detailed mental simulations carried out by brain regions that support perception and motor action (for reviews, see Barsalou, 2008; Binder and Desai, 2011; Kiefer and Pulvermüller, 2012; Meteyard et al., 2012).

Important evidence for this view comes from functional neuroimaging studies. As an example, in their seminal study, Hauk et al. (2004) found that verbs related to arm, leg, and face actions (e.g., "pick", "kick", or "lick") activate respectively the hand, foot, and mouth portions of the motor cortex in an effector-specific way. Relatedly, Willems et al. (2010a) showed that, when processing manual verbs, right-handed people recruited left premotor hand areas, while left-handers preferentially activated the right premotor cortex (but see

Postle et al., 2008; Watson et al., 2013). Moreover, according to neurophysiological evidence from EEG and MEG studies, this motor activation is fast enough to be considered part of meaning access (within the first 250 ms after word presentation; e.g., Hauk and Pulvermüller, 2004; Shtyrov et al., 2014; but see Papeo et al., 2009).

However, these correlational results do not speak to the functional relevance of the motor cortex in meaning representation (e.g., Mahon and Caramazza, 2008). In response to the need to move beyond correlational data, several studies have tested for a causal contribution of motor areas to language processing by interfering with motor system activation during action language comprehension (for a review, see Ostarek and Bottini, 2021). For instance, Shebani and Pulvermüller (2013) found that interfering with motor cortex activity by moving repeatedly the hands and the feet reduced the retrieval of hand and foot verbs, respectively. Relatedly, neuropsychological studies on patients with motor impairment such as Parkinson's disease (e.g., Boulenger et al., 2008; Fernandino et al., 2013; García et al., 2018), as well as lesion symptom mapping studies (e.g., Dreyer et al., 2020; Kemmerer et al., 2012), suggest that alterations to the motor system negatively affect the processing of action-related language. Most importantly for the present work, disrupting the motor and premotor cortex with

* Corresponding author. Department of Experimental Psychology, University of Granada, 18071, Granada, Spain.

E-mail address: solana@ugr.es (P. Solana).

non-invasive brain stimulation techniques such as Transcranial Magnetic Stimulation (TMS) or Transcranial Direct Current Stimulation (tDCS) has also been shown to alter performance in multiple tasks involving action language (e.g., Birba et al., 2020; Gijssels et al., 2018; Lo Gerfo et al., 2008; Pulvermüller et al., 2005; Repetto et al., 2013; Tremblay et al., 2012; Vitale et al., 2021; Vukovic et al., 2017; Willems et al., 2011).

Such results have long been considered to be strong evidence for a causal role of motor brain areas in action language understanding. Nonetheless, a growing set of studies has started to question the credibility of this literature. For instance, a preregistered and well-powered study (Montero-Melis et al., 2022) failed in replicating the above-mentioned Shebani and Pulvermüller's study (see also Postle et al., 2013; Strozyk et al., 2019). Similarly, by evaluating a sample of patients with lesions to the sensorimotor cortices, Argiris et al. (2020) reported that this condition does not have a negative impact on the comprehension of action-related verbs, and comparable results have been found in Parkinson's disease patients (e.g., Aiello et al., 2022; Humphries et al., 2019; Kemmerer et al., 2013) and lesion symptom mapping studies (e.g., Arévalo et al., 2012; Reilly et al., 2014). As a final example, we recently meta-analyzed 43 TMS and tDCS studies of motor cortex stimulation during action language comprehension by means of *p*-curve analyses and tests for excess significance (Solana and Santiago, 2022), as well as *z*-curve analyses (Solana and Santiago, 2023). Our results suggested that (1) we cannot conclude that these studies explore real effects, (2) their estimated underlying power is quite low (i.e., the majority of them may not replicate if repeated identically), and (3) they contain clear signs of a large publication bias.

Another important caveat regarding this body of research, which has been mostly overlooked, relates to the validity of its results to infer a causal contribution of the motor system in computing meaning from language. Researchers have focused on demonstrating that disturbing motor activation can quantitatively alter language processing. That is, affecting reaction time (e.g., Repetto et al., 2013; Vukovic et al., 2017), accuracy (e.g., Birba et al., 2020; Gijssels et al., 2018), or even recall rate (e.g., Shebani and Pulvermüller, 2013; Vitale et al., 2021). Yet, do subtle changes in this kind of measure really reflect a relevant change in language comprehension? (Chatterjee, 2010). Such measures are indexes of how difficult it is to access, compute, and retrieve the semantic information of language. However, they might not tackle directly the central meaning construction processes that confer meaning its final form. Accordingly, a more powerful proof in favor of a causal role of the motor system in language comprehension would be demonstrating that interfering with the activation of motor brain areas can qualitatively change meaning itself (see Casasanto, 2023). At the time we conceptualized the present study, no published study had addressed this issue (but see Togato et al., 2021, for a recent behavioral study assessing a similar question). The present study aimed to test the causal involvement of the primary motor cortex (M1) in action language comprehension, given the inconclusive results in the literature. And specifically, to assess whether interfering with M1 activation can qualitatively change the meaning that people construe from language.

How can meaning change in response to the interference? According to several theories, such as the Construal Level Theory (Trope and Liberman, 2010) or the Action Identification Theory (Vallacher and Wegner, 1987), actions (e.g., "voting") can be construed in either a more concrete or a more abstract way. Concrete representations are focused on the lower-level, sensorimotor details of the action (e.g., "marking a ballot"), while abstract ones rely on higher-level aspects such as the goals and the consequences of the act (e.g., "influencing an election"). From some influential versions of embodiment theory, it follows that a possible role for the cortical motor system in language comprehension is to represent and simulate the motoric, experience-related, and thus more concrete aspects of the meaning of language (e.g., Barsalou, 1999, 2008; Pulvermüller, 2005). Returning to our example, these theories would predict that while processing the meaning of "voting", people

engage in a simulation of the movements required for "marking a ballot". That way, the same neural tissue used to perform those manual movements (i.e., the hand areas of the brain) would also be involved in shaping meaning (Bergen, 2015). If so, we believe that interfering with motor system activation during action language comprehension should alter the level of abstraction at which meaning is construed.

To test this idea, participants were presented with sentences containing foot actions (e.g., "braking a car") or hand actions (e.g., "turning off the light") and chose between two different interpretations of the same action: one interpretation was more concrete (e.g., "flipping a switch") and the other one was more abstract (e.g., "going to sleep"). To interfere with motor system activation, before the task, low-frequency repetitive TMS (rTMS) was applied over the hand portion of their left primary motor cortex (M1). A control group received active stimulation to the vertex. This stimulation protocol has been shown to inhibit the proper functioning of a neural population (including M1; Chen et al., 1997), making it a well-established method to test the causal contribution of a certain brain region to a particular cognitive process (for a review, see Polanía et al., 2018). Indeed, several previous rTMS studies supporting embodiment have employed this protocol (e.g., Lo Gerfo et al., 2008; Repetto et al., 2013; Tremblay et al., 2012). Left M1 was selected as target since it is the region most studied and supported by previous rTMS studies of embodied language processing (e.g., Lo Gerfo et al., 2008; Repetto et al., 2013; Vukovic et al., 2017; for review, see Solana and Santiago, 2022).

If primary motor cortex is causally involved in computing the more concrete features of the meaning of action language (e.g., see Pulvermüller, 2005), we hypothesized that inhibiting the hand motor area of the brain would prevent participants from simulating the concrete details of the meanings of sentences describing hand actions. Therefore, we predicted a greater proportion of abstract interpretations for hand sentences in the M1-rTMS condition, as compared to the vertex control condition. Since the effect was expected to be effector-specific (e.g., Pulvermüller et al., 2005; Shebani and Pulvermüller, 2013) we predicted no changes in foot sentences.

In addition, as a manipulation check for motor cortex inhibition, we recorded motor-evoked potentials (MEPs) from the first dorsal interosseous muscle of the hand (FDI) before the stimulation and after completing the task (Chen et al., 1997; Huang et al., 2005). Proper inhibition of motor cortex should decrease MEP amplitudes at the end of the task (compared to before the stimulation) for the M1 stimulation group, but not for the vertex group. To the best of our knowledge, this is the first rTMS study in the embodied language literature implementing this manipulation check. Also for the first time in this literature, we preregistered the hypotheses, sample size, design, exclusion criteria, and analysis plan of the study. Moreover, we share the materials, raw data, and analysis scripts as well. These practices will help the interpretation of any possible results and assist future replications and meta-analyses.

2. Main experiment: rTMS

2.1. Transparency and open practices

Materials, raw data, and analysis scripts are available at the following Open Science Framework (OSF) repository: <https://osf.io/grhmu>. All details of the study were preregistered prior to data collection: <https://aspredicted.org/cz2y5.pdf>.

2.2. Methods

2.2.1. Sample size justification

The novelty of the experiment made it impossible to estimate an effect size based on prior literature. Therefore, considering our resources, we decided to set a maximum sample size of 40 participants (20 per group, as recommended by Simmons et al., 2011). However, we added some additional constraints. Following the logic of sequential

analyses (Wald, 2004; see also Lakens, 2014), we planned to stop data collection at $N = 30$ and test the key interaction at $\alpha = 0.03$. If it were not significant, we would compute a Bayes factor (BF) to assess the evidence towards the absence of interaction (e.g., see Dienes, 2014). If the BF supported the null hypothesis with a $BF_{01} > 3$, we would stop collecting data (Schönbrodt and Wagenmakers, 2018). If not, we would continue until reaching $N = 40$ and test the interaction at $\alpha = 0.02$. By establishing this conditional stopping rule, we keep Type I error at the desired level of 5%.

2.2.2. Participants

Thirty-three right-handed, native Spanish speakers participated in the experiment. However, as preregistered, three of them were excluded since we were unable to establish their motor threshold. This left us with the preregistered sample size of thirty participants ($M_{age} = 23.1, SD_{age} = 4.14$; 5 men). It is worth noting that this sample size is greater than most of the samples used in previous rTMS studies of embodied language comprehension, which usually range between 15 and 20 participants (see Solana and Santiago, 2022). Following our preregistered plan, we did not reach $N = 40$ since the BF indicated strong evidence towards the absence of effect with $N = 30$ (see section 2.3).

All the participants took part in the study voluntarily after signing an informed consent and were compensated with an economic reward (10€/hour). Following the standard safety requirements to undergo TMS studies (Rossi et al., 2021), all the participants were screened before the experiment to make sure that none suffered from any neurological, psychiatric, or psychological condition, or any other that precludes their participation. The study was approved by the Ethics Committee on Human Research of the University of Granada (#697/CEIH/2018).

2.2.3. Materials

The construction of the materials was inspired by the Behavioral Identification Form (BIF; Vallacher and Wegner, 1989), considered the most valid instrument to measure abstraction level (Mac Giolla et al., 2022). First of all, we created 120 sentences in Spanish. 70 of them described actions usually carried out with the hands/arms (e.g., “turning off the light”), while the remaining 50 referred to actions related to the

feet/legs (e.g., “braking a car”). Each sentence was accompanied by two interpretations: one more concrete and another more abstract. For example, the interpretations of the above-mentioned sentences were “flipping a switch” and “stepping on a pedal” (concrete), and “going to sleep” and “reducing risk while driving” (abstract).

These materials were presented to 172 native Spanish speakers in an online norming study. None of them took part in the main experiment. They were visually presented with the 120 sentences (paired with their interpretations) and asked to decide which interpretation, in their opinion, described better the meaning of each sentence. Sentence order was randomized. Two different lists were created to counterbalance the position of each response option on the screen. Then, we computed the proportion of choice of the abstract interpretation for each sentence and selected those with proportions near 50% (ranging between 40% and 60%) to create the final set of stimuli. This was done to avoid ceiling or floor effects in the rTMS experiment and thus make it easier to detect an effect in any direction. The final set of stimuli contained 40 sentences: 20 hand sentences and 20 foot sentences. Four extra sentences which did not fit our requirements were used for the practice block of experiment. All the experimental sentences can be consulted at the OSF repository.

2.2.4. Behavioral procedure

In each trial (Fig. 1 bottom), participants saw a blank screen and were presented an action description auditorily through headphones. Two hundred milliseconds after the end of the action description, the blank screen was replaced with the two interpretations, one above the other. The top interpretation was labelled as “1” and the bottom interpretation as “2”. Participants were asked to decide, without time limit, which interpretation they considered more appropriate for describing the meaning of the given action. They were instructed that both interpretations were equally valid. Once they had their choice, they verbally communicated the corresponding number to the experimenter, who inputted the number in a keyboard, out of participants’ sight. This prevented participants from using their hands, which may affect motor cortex activation. Before the next trial, another blank screen appeared during a random interval between 200 and 1000 ms.

The experiment comprised one single block of 40 trials (one per

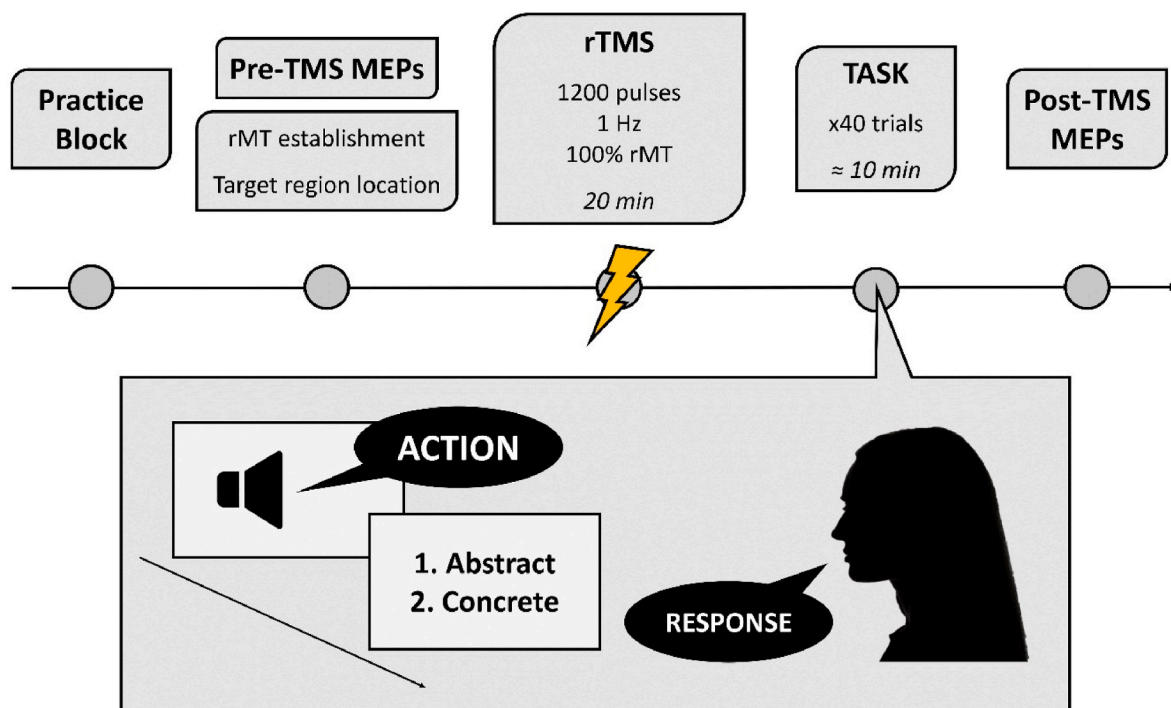


Fig. 1. Procedure of the behavioral task (bottom) and summary of the experimental session (top).

sentence). Sentence order was randomized. Half the trials presented the concrete interpretation on the upper part of the screen and the abstract interpretation below, while the other half presented the abstract interpretation above and the concrete below. The vertical location of the concrete and the abstract interpretation for each sentence was counterbalanced among participants using two lists of items.

Before undergoing the stimulation, participants completed a brief 4-trial practice block of the task. This was done to make sure the participants understood the instructions before the complete version of the task was presented and, consequently, to take the maximum advantage of the duration of cortical inhibition. After that, participants received the stimulation (section 2.2.5) and, immediately after, completed the experimental trials (as in previous rTMS studies supporting the grounding of action language in the motor cortex; e.g., Lo Gerfo et al., 2008; Repetto et al., 2013; Tremblay et al., 2012). Some studies recommend leaving a short delay between the end of the stimulation and the beginning of the subsequent cognitive task. Otherwise, the neural changes expected to occur after the stimulation may be canceled, or even reversed (see Gentner et al., 2008; Hartwigsen and Silvanto, 2023). Even if participants started the task immediately after the stimulation, we are confident about the inhibitory effects of rTMS, as they were significantly observed in the MEP manipulation check (for details, see section 2.3).

2.2.5. rTMS protocol

The stimulation session began by determining the participant's resting motor threshold (rMT). To do so, we first searched for the location of the participant's right-hand representation in left primary motor cortex (M1). We determined the hand M1 area as the hotspot evoking the highest motor evoked potential (MEPs) amplitudes for the first dorsal interosseous (FDI) muscle of the right hand. MEPs were recorded using snap surface electrodes (Natus Neurology). Then, we established the participant's rMT, defined as the minimum intensity value that elicited MEP amplitudes greater than 50 μV in five out of ten consecutive pulses (Rossini et al., 2015). The mean rMT was 63.1% ($SD = 6.55$) for the M1 group, and 68.23% ($SD = 5.34$) for the vertex group. These values are consistent with those reported in previous studies, both from our laboratory (e.g., Martín-Signes et al., 2019) and other laboratories (e.g., Boyd and Linsdell, 2009), using the same TMS stimulator and EMG recorder.

Participants were randomly assigned to one of two experimental groups. Both groups completed the same experimental session (Fig. 1 top) but differed in where the stimulation was applied: the hand area of the left hemisphere M1 ($N = 15$) or the vertex ($N = 15$). In alignment with previous brain stimulation studies on embodied language comprehension (e.g., Birba et al., 2020; Lo Gerfo et al., 2008; Repetto et al., 2013), the hand M1 area was defined as the same hotspot used to establish the rMT. The vertex was defined following the 10–20 system guidelines as the intersection between the midpoint of the line connecting the preauricular points of the left and right ears and the midpoint between the nasion and inion (Trans Cranial Technologies, 2012). The vertex was used as an active control site for M1 stimulation since stimulating this region is expected to have no effect on language comprehension (e.g., Harpaz et al., 2009; Tomasino et al., 2008), while mimicking tactile sensations and noise of M1 stimulation. We decided to implement a between-group design (instead of a within-participant design) since we did not want to expose our participants with the same stimuli twice, and we did not have enough sentences to create two parallel versions of our task.

The stimulation was delivered by means of a biphasic repetitive stimulator (Super Rapid 2, Magstim, Whitland UK) and a 70-mm TMS figure-of-eight coil (Magstim, Whitland UK) which was held tangentially to the skull with the axis of the coil angled 45° from the mid-sagittal axis (lateral to medial and caudal to rostral). A TMS neuronavigation system (Brainsight; Rogue Systems, Montreal, Canada) was used to maximize spatial precision. During rTMS, the coil was controlled by a robotic arm

(TMS Robot; Axilum Robotics). The stimulation sequence corresponded to a low-frequency repetitive TMS (rTMS) protocol (Chen et al., 1997). More specifically, it consisted of two trains of 600 pulses applied at a frequency of 1 Hz with an inter-pulse interval of 1 s (1200 pulses in total for 20 min). The stimulation was delivered offline (i.e., before the task) and at 100% of each participant's rMT. In spite of its long duration, we decided to use this protocol, instead of a shorter but higher-frequency protocol such as theta-burst stimulation (TBS; Huang et al., 2005), because we consider the present one much more comfortable for the participants. In fact, none of the participants reported any discomfort, and all confirmed that they felt well enough to complete the behavioral task.

Importantly, this TMS protocol is expected to reduce corticospinal excitability for at least 15 min (e.g., Chen et al., 1997; Muellbacher et al., 2000), which ought to guarantee the proper inhibition of M1 until the end of the behavioral task, which lasted around 10 min. Yet, as a direct manipulation check for M1 inhibition, we recorded pre- and post-stimulation MEPs from the FDI muscle (Chen et al., 1997; Huang et al., 2005). The ten MEPs used to set participants' rMT were taken as pre-stimulation MEPs. Post-stimulation MEPs were obtained by applying ten extra pulses over M1 immediately after finishing the behavioral task (to make sure that M1 was inhibited until the end of the task). It is known that both motor and cognitive activity can modulate MEPs (e.g., Buccino et al., 2005; Gentner et al., 2008). Yet, since participants did not produce manual responses during the task, post-stimulation MEPs were unaffected by this potential confound. More generally, any differences in post-stimulation MEPs between the experimental and control groups cannot be attributed to any aspect of the protocol other than the stimulation site (e.g., the lack of a rest period between the end of the task and the measurement of MEPs) because, besides that, both groups underwent the exact same protocol.

2.2.6. Design and data analysis

As preregistered, we conducted two different analyses: one for the behavioral results from the meaning construction task, and another for the pre-post stimulation MEP amplitudes. All the analyses were carried out in R (R Core Team, 2021). Raw data and data scripts are available at the OSF repository.

2.2.6.1. Main analysis: meaning construction task. There was one within-participant factor: Sentence Type (hand vs. foot action), and one between-participant factor: TMS Group (M1 vs. vertex stimulation). We also included the individual differences in MEP amplitudes between the pre- and the post-stimulation as a covariate (MEP Differences hereafter). The dependent measure was the proportion of choice of the abstract interpretation.

Data were analyzed following the preregistered plan by means of mixed-effects models using the *lme4* package (Bates et al., 2015). This let us take into account the variability due to both participants and sentences simultaneously and without data aggregation, which reduces the likelihood of obtaining false-positive results, increases power, and yields more generalizable results (Baayen et al., 2008). Because our dependent variable is binomial (i.e., choosing one or another interpretation), we used generalized mixed models with a logit link (Jaeger, 2008). Following Barr et al. (2013), we started with the maximal model justified by the design in order to reduce Type I errors. The fixed term of this model included Sentence Type, TMS Group, MEP Differences, and all their possible interactions. The random term included random intercepts for Items and Participants, as well as the slope of the fixed factors that varied over their corresponding random factor. From this model, we searched for the simplest model that maintained the same goodness of fit as the maximal converging model (Bates et al., 2015). Model comparisons were done using the *anova()* function of the *lmerTest* package (Kuznetsova et al., 2017). *P*-values were obtained from type II Wald chi-square tests, using the *Anova()* function from the *car* package (Fox

and Weisberg, 2019). Effect sizes were calculated as odds ratios (OR) by exponentiating the estimate (β) values from the model. 95% confidence intervals around the ORs were computed by means of the *confint.merMod()* function of *lme4*. Alpha level was set at $\alpha = 0.03$ (see section 2.2.1).

To obtain the Bayes factor (BF) for the critical interaction, we computed a BIC-based Bayes factor (Masson, 2011; Wagenmakers, 2007) by comparing the BIC (Bayesian Information Criterion) of two competing models: the best fitting model containing the key interaction (full model; H1) and the same model without the key interaction (null model; H0). To do so, we relied in the following equation: $BF_{01} = e^{(BIC_{null} - BIC_{full})/2}$. This way to compute BF assumes a minimally informative prior (the unity information criterion, see Wagenmakers, 2007) and leaves no degrees of freedom to the experimenter.

2.2.6.2. Control analysis: MEP amplitudes. There was one within-participants factor: Time (before the stimulation vs. after completing the task), and one between-participants factor: TMS Group (M1 vs. vertex). The dependent measure was the peak-to-peak amplitude of MEPs.

The visual inspection of both a density plot and a Q-Q plot (Das and Imon, 2016) revealed that the data distribution did not follow normality (see the R script at the OSF repository). Therefore, as preregistered, we used non-parametric tests. Specifically, we ran two Wilcoxon tests with the *wilcox.test()* function: one for the pre-post comparison in the M1 group and another for the pre-post comparison in the vertex group. Effect sizes were calculated as Kendall's tau-b coefficients (T_b) using the *KendallTauB()* function from the *DescTools* package (Signorelli, 2017).

2.3. Results and discussion

As expected, the application of rTMS over M1 reduced hand muscles excitability, as indexed by lower MEP amplitudes after stimulation (pre-stimulation MEPs: $M = 54.89 \pm 3.58 \mu V$; post-stimulation MEPs: $M = 40.77 \pm 3.62 \mu V$; $W = 8475$, $p < 0.001$, $T_b = 0.18$, 95% CI [0.09, 0.27]). Conversely, rTMS to the vertex did not yield any significant difference between the pre-stimulation MEPs ($M = 50.03 \pm 3.15 \mu V$) and the post-stimulation MEPs ($M = 92.74 \pm 11.41 \mu V$; $W = 11744$, $p = 0.39$, $T_b = -0.04$, 95% CI [-0.14, 0.05]). In fact, 12 out of 15 participants in the M1 group showed a lower mean amplitude after stimulation, while in the vertex group, 8 participants showed higher MEPs after stimulation, 6 showed lower MEPs, and one remained unchanged. These results confirm that the right-hand area of the primary motor cortex of the participants in the M1 group was properly inhibited until the end of the comprehension task, whereas no reliable changes in motor cortex activation occurred in the vertex group (Chen et al., 1997; Huang et al., 2005).

Despite successful inhibition of M1, results of the behavioral task indicated that M1 inhibition did not lead to any significant change in meaning construction (Fig. 2). After searching for the most parsimonious model, data were analyzed using a model that contained Sentence Type, TMS Group, and their interaction as fixed factors, and random intercepts for Participant and Item as random factors¹ (the full model selection process can be consulted in detail in the OSF repository). The model revealed no significant main effects of either TMS Group ($\chi^2(1) = 0.47$, $p = 0.49$, OR = 1.39, 95% CI [0.54, 3.56]) or Sentence Type ($\chi^2(1) = 0.99$, $p = 0.32$, OR = 0.79, 95% CI [0.49, 1.26]). Most importantly, the TMS Group x Sentence Type interaction was also non-significant ($\chi^2(1) = 0.06$, $p = 0.80$, OR = 0.93, 95% CI [0.51, 1.68]). The proportion of choice of the abstract interpretation did not differ between hand and foot sentences, neither in the M1 group (hand sentences: $M = 0.25$, $SD = 0.43$; foot sentences: $M = 0.29$, $SD = 0.45$; $\chi^2(1) = 1.06$, $p = 0.30$, OR = 0.76, 95% CI [0.45, 1.28]) or the vertex group (hand sentences: $M =$

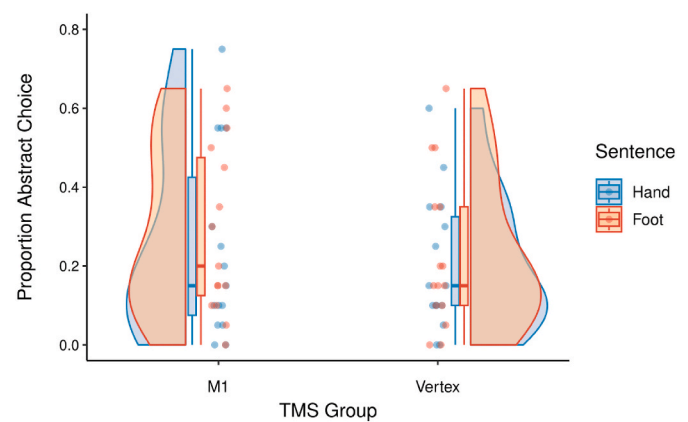


Fig. 2. Raincloud plot for the proportion of abstract choice in the main experiment as a function of TMS Group and Sentence Type. The box plots depict the median (horizontal line inside the box), upper quartile (top limit of the box), lower quartile (bottom limit of the box), maximum value (end of the top whisker), and minimum value (end of the bottom whisker) of the data. The jittered points correspond to the raw mean of each participant. The density functions represent the distribution of the data.

0.20, $SD = 0.40$; foot sentences: $M = 0.23$, $SD = 0.42$; $\chi^2(1) = 0.51$, $p = 0.48$, OR = 0.83, 95% CI [0.50, 1.38]). In fact, the Bayes factor for the TMS Group x Sentence Type interaction was $BF_{01} = 33.61$, which is interpreted as very strong evidence towards the absence of effect (Lee and Wagenmakers, 2014). Moreover, note that the fact that the MEP Differences covariate was not retained in the model means that this factor did not yield any significant main effect or interaction (all p s > 0.05). In other words: the changes in motor cortex excitability following the stimulation were not related to the expected interaction between TMS Group and Sentence Type.

In summary, these results provide no support for a causal role of primary motor cortex in construing meaning from action language, even when controlling for individual variations in M1 excitability. And importantly, the absence of effect cannot be accounted for by failure to significantly inhibit M1. Yet, several other factors could potentially explain the lack of effect. First, we consider whether the behavioral task is able to reveal changes in meaning construction (e.g., see Mac Giolla et al., 2022). Second, we consider whether the study had enough statistical power to detect the expected effect (e.g., see Button et al., 2013). The next two sections are devoted to explore these questions.

3. Control experiments: is our task valid?

Over recent years, concern is raising regarding the (non)validity of the psychological tasks to measure the target cognitive processes, which may contribute to the growing number of replication failures in cognitive science (see Schimmack, 2021). Could the lack of effect in the rTMS experiment be due to the inability of the present task to capture changes in construal level? To assess this possibility, we run two additional control experiments, one in-person ($N = 40$) and one online ($N = 101$). Both experiments were preregistered prior to data collection. Detailed information about preregistrations, sample size justifications, procedures, analyses, and results is available as Supplementary Material at the OSF repository.

Broadly, in both experiments, participants were presented with the same task and materials than in the rTMS study (without stimulation), but were induced to think about the actions either abstractly or concretely. To do so, we relied in the procedure designed by Mac Giolla et al. (2022) to validate the BIF, consisting of explicitly asking the participants to think in either abstract or concrete terms during the experiment: "Focus on the overall purpose and meaning of the actions" (abstract induction) vs. "Focus on the physical performance of the

¹ In R notation: `glmer(Choice ~ TMSGroup + SentenceType + TMSGroup: SentenceType + (1|Participant) + (1|Item), family = binomial(link = logit))`.

actions" (concrete induction). If our task is really able to capture changes in construal level, participants in the abstract condition should choose more abstract interpretations than those in the concrete condition. Moreover, we expected this effect to be present for both hand- and foot-related sentences.

The results supported these predictions. In the in-person experiment, participants in the abstract condition chose significantly more abstract interpretations ($M = 0.78$, $SD = 0.41$) than those in the concrete condition ($M = 0.29$, $SD = 0.46$; $\chi^2(1) = 34.06$, $p < 0.0001$, $OR = 27.48$, $95\% CI = [8.28, 91.21]$). Importantly, that effect was present for both hand ($\chi^2(1) = 33.75$, $p < 0.0001$, $OR = 47.74$, $95\% CI = [12.97, 175.74]$) and foot sentences ($\chi^2(1) = 31.02$, $p < 0.0001$, $OR = 18.86$, $95\% CI = [6.71, 53.00]$). The online experiment replicated these findings, with participants in the abstract group selecting more abstract interpretations ($M = 0.73$, $SD = 0.44$) than those in the concrete group ($M = 0.11$, $SD = 0.31$; $\chi^2(1) = 122.07$, $p < 0.0001$, $OR = 93.76$, $95\% CI = [39.26, 223.94]$). Once again, the effect was present for both hand ($\chi^2(1) = 88.89$, $p < 0.0001$, $OR = 418.53$, $95\% CI = [119.32, 1468.07]$) and foot sentences ($\chi^2(1) = 106.81$, $p < 0.0001$, $OR = 69.62$, $95\% CI = [31.08, 155.29]$). These results clearly support that our materials and task are valid to capture changes in abstraction level. Therefore, the absence of effects showed in the rTMS experiment cannot be accounted for by the use of an invalid task.

4. Sensitivity analysis: can we detect a small effect?

Failing to detect a true effect can also be due to low statistical power (Button et al., 2013), which mainly derives from the use of small sample sizes and few items or observations per participant (Brysbaert, 2019). In this line, although our task seems able to capture changes in construal level, it is possible that the target effect is so small that the rTMS experiment lacks power to detect it.

To test this possibility, we carried out a sensitivity power analysis, a novel approach that informs about the smallest effect size an experiment can detect considering a given power level (for a general introduction, see Lakens, 2022a). To do so, we run several series of 1000 simulations using the data from the rTMS study and the mixed model that was selected to analyze them. This secured us to work with the same matrix of variances and covariances observed in the real experiment. In each series, we maintained fixed the number of participants ($N = 30$) and items ($k = 40$), as well as the alpha level ($\alpha = 0.03$), but tried different effect sizes of the key interaction until determining the smallest effect size that could be detected with 80% power (i.e., when 80% of the simulations yielded p -values lower than 0.03). We set the desired power level at 80% following Button et al. (2013). These calculations were performed in R by means of the *simr* package (Green and MacLeod, 2016). The code for these simulations is available at the OSF repository.

The analysis revealed that the rTMS experiment was able to capture an effect size of $OR = 2.52$ with 80.6% power ($95\% CI = [78.01, 83.01]$). Crucially, following the established thresholds for interpreting the size of odds ratios, this effect size is considered small (i.e., $OR < 3.47$; Chen et al., 2010). Consequently, we can reject that the absence of effect in the rTMS experiment was due to not having enough power to detect a small effect (see section 5 for discussion). It is important to point out that the results of this sensitivity analysis should not be interpreted as meaning that any experiment with the same design as ours will be well-powered. Present results only apply to our specific experiment, constrained by its own design, materials, variability observed in the data, and analytic pipeline.

5. General discussion

Does M1 play a causal role in meaning construction? The present preregistered study aimed to assess whether perturbing motor cortex activation with rTMS can alter meaning construction from action language. Contrary to our predictions, the results failed to show a

significant change in language comprehension following the stimulation of primary motor cortex (M1). In fact, they strongly supported the absence of the key interaction. Hence, our results fail to support a central prediction of the embodiment view of language comprehension: that altering the activity of M1 should produce changes in the meaning of action sentences.

Importantly, we conducted additional analyses and control experiments that discarded several alternative causes of the lack of effect. First, pre-post stimulation MEP recordings showed that applying rTMS over M1 successfully reduced hand muscles excitability, which has been interpreted as a proxy of motor cortex inhibition (Chen et al., 1997; Huang et al., 2005). Therefore, the absence of a significant behavioral effect cannot be explained by an absence of effect at the neural level due to non-proper functioning of our manipulation. Second, across two control experiments, we observed that the employed task captured changes in abstraction level, thus confirming the construct validity of the task (Mac Giolla et al., 2022). Third, a sensitivity power analysis revealed that the experiment had 80% power to observe a small-sized effect. Moreover, the preregistered nature of this study secures that present results do not depend on deviations from our original plan (Hardwicke and Wagenmakers, 2023; Nosek et al., 2018). To the best of our knowledge, none of these quality checks have been implemented to date in any previous rTMS study within the embodied semantics field.

According to the strongest versions of embodiment, the reactivation of the neural architecture underlying action performance is an automatic and necessary component of action language comprehension (Barsalou, 1999; Gallese and Lakoff, 2005; Pulvermüller, 2005). Present results prevent us to conclude in support of this claim. This conclusion thus stands in sharp contrast with the large number of behavioral (e.g., Glenberg et al., 2008b; Shebani and Pulvermüller, 2013; Togato et al., 2021), neuropsychological (e.g., Boulenger et al., 2008; Bernardino et al., 2013; García et al., 2018), and specially, brain stimulation studies (e.g., Birba et al., 2020; Lo Gerfo et al., 2008; Pulvermüller et al., 2005; Repetto et al., 2013; Vukovic et al., 2017; Willems et al., 2011) that claim support for a functional role of the motor system in action language comprehension.

On the contrary, present findings are most compatible with several studies on behavioral motor interference (e.g., Postle et al., 2013; Saccone et al., 2021; Strozyk et al., 2019), Parkinson patients (e.g., Aiello et al., 2022; Humphries et al., 2019; Kemmerer et al., 2013), lesion symptom mapping (e.g., Arévalo et al., 2012; Reilly et al., 2014), and TMS (e.g., Papeo et al., 2009, 2011; Tomasino et al., 2008) that suggest that the motor system does not play a functional role in action language comprehension. More broadly, our results are also in line with recent studies on modality-specific systems other than the motor system, such as vision (Ostarek et al., 2019) or olfaction (Speed et al., 2022), failing to observe causal influences of low-level simulations in language understanding. Therefore, present results could be interpreted in agreement with the idea that sensorimotor activation, rather than being a necessary feature of language comprehension, may be a by-product of non-conceptual processes (such as motor imagery; e.g., Papeo et al., 2009, 2011; Tomasino et al., 2008), reflect aspects of language processing other than semantic processing (such as ortho-phonological processing; e.g., de Zubicaray et al., 2013, 2021; see also Brandscheidt et al., 2018), or be flexibly engaged according to situational and contextual demands, as suggested by weaker and more contemporary versions of embodiment (e.g., García and Ibáñez, 2016; Ibáñez et al., 2023).

The current study also aligns with the growing body of research, including replications (Montero-Melis et al., 2022; Morey et al., 2022), reanalyses (Papesh, 2015; Witt et al., 2020), and meta-analytic works (Solana and Santiago, 2022, 2023), that raises concerns about the reliability of key findings in the embodied language literature. Indeed, as sketched in the Introduction, previous work from our group already warned that, specifically, motor cortex stimulation studies in this field do not stand on solid ground (Solana and Santiago, 2022, 2023), which

is compatible with the fact that we have not been able to observe even a small-sized effect in the present rTMS study.

Do present results mean that no part of the motor system is implicated in construing meaning from action language? According to classic embodied accounts, motor simulations underlying action language comprehension are similar to action performance in real life, which predicts a functional implication of brain regions devoted to action execution, such as M1 — the “re-enactment” view (Barsalou, 1999; Pulvermüller, 2005). On an alternative view, M1 is restricted to intentional motor imagery, while the implicit simulations expected to occur during language understanding rely on the reactivation of the neural circuits involved in action preparation, which include regions such as the premotor cortex (PMC) or the supplementary motor area (SMA) — the “pre-enactment” view (Gijssels et al., 2018; Willems et al., 2010b). In this line, there are studies that have indeed shown effects in action language comprehension after stimulating the PMC (e.g., Gijssels et al., 2018; Willems et al., 2011) and the SMA (Courson et al., 2017). The present study did fail to provide evidence for a causal role of M1, despite being the most targeted and supported region in this field (e.g., Lo Gerfo et al., 2008; Repetto et al., 2013; Vukovic et al., 2017; for review, see Solana and Santiago, 2022), but it cannot discard that stimulating secondary motor areas might change meaning construction. Future studies are needed to clarify the causal implication of motor areas outside primary motor system.

Do present results mean that the motor system is not implicated in any stage of language comprehension? As argued in the Introduction, we believe that the measures used in previous studies capture the work of mechanisms that can alter the difficulty of meaning construction, but not those that determine its final form. On the contrary, the present construal level task was intended to tackle central meaning construction processes that determine the final construed meaning. In fact, this idea was strongly supported by our validation studies (see also Mac Giolla et al., 2022). If so, present results call into question that M1 is implicated in central processes of meaning construction, but leave open the possibility of a more superficial or collateral implication of this cortical area. The latter account may be responsible of the significant findings in previous studies using measures of reaction time and accuracy. Moreover, it offers a plausible path to reconcile present findings with previous ones. More research combining both families of tasks is needed to give a more conclusive answer to this debate.

Nonetheless, there does exist a recent study (published while we were conducting the present study) that used a task comparable to ours and claimed support for embodied tenets: Togato et al. (2021). Their participants were presented with object nouns having two different meanings (homographs): one related to hand or foot movements, and another not related to movement. Then, from a list of three words, they were asked to choose the one they thought was more strongly related to the previously presented homograph. The words could be related to the motor meaning of the homograph, to the non-motor meaning, or to be unrelated. During the task, participants moved repeatedly either their hands, their feet, or they moved nothing. Results showed that, when the homograph was highly biased towards the motor meaning, moving an effector reduced the probability of choosing the motor meaning associated with the same effector being used. On the contrary, in presence of homographs with a low bias towards the motor meaning, moving an effector increased the probability of choosing the motor meaning related to that effector.

Despite procedural differences, it could be argued that Togato et al.’s work is directly comparable to the present one. Hence, their results and ours raise mutual concerns about each other’s conclusions. Can present results be reconciled with those by Togato et al.? One potential reason for this discrepancy may be related to the interference method employed (M1-rTMS vs. moving the hands/feet). As previously discussed, the embodiment of language may involve regions other than M1. Repeatedly moving a body effector might have engaged a larger set of sensorimotor regions than stimulating M1 (e.g., Scott, 2012), which may have

acted as the locus for the effect. Relatedly, it is possible that continuous limb movements produced a greater motor interference than our stimulation protocol. Finally, it should be noted that our participants were completely blind to the stimulation site, but participants in Togato et al.’s study overtly moved their hands or feet. This could have led to priming effects consistent with an amodal view (Ostarek and Bottini, 2021). However, although this account may explain their facilitatory effect on low-biased homographs, it can hardly account for their inhibitory effect on highly-biased homographs.

Several concerns can be raised regarding the interpretation of the present null result. A first concern is the possibility that the motor system is implied in representing the meaning of both concrete and abstract concepts. Here, we built upon the distinction between concrete (embodied) and abstract (disembodied) concepts handed down from classic embodied theories (e.g., Barsalou, 1999; Pulvermüller, 2005). However, more contemporary theories argue that both concrete and abstract construals are grounded in sensorimotor experience, thereby blurring the abstract-concrete distinction (e.g., Banks et al., 2023; Buccino et al., 2019). As an example, a recent investigation by Banks and Connell (2023) concluded that participant-generated exemplars for abstract concepts contained traces of sensorimotor experience. Most importantly, some neurophysiological (e.g., Glenberg et al., 2008a; Harpaintner et al., 2022) and neuroimaging (e.g., Harpaintner et al., 2020; Sakreida et al., 2013) studies have found that abstract language recruits motor regions as well. Accordingly, disrupting M1 could equally affect both the concrete and abstract features of meaning, leading to no observable changes in the task. These studies are, nonetheless, correlational, and there remains the fact that the published rTMS literature agrees that perturbing motor cortex activation selectively alters the processing of concrete but not abstract verbs (e.g., Lo Gerfo et al., 2008; Repetto et al., 2013; Willems et al., 2011). Only Vukovic et al. (2017) reported an unpredicted facilitation for abstract verbs, but they interpreted it as a product of spreading effects of TMS over prefrontal areas. Therefore, although this issue deserves more attention in the future, in light of the current state of the literature, we do not consider this a likely explanation for present results.

A second set of concerns relates to aspects of the use of brain stimulation in the present study. First, according to some relevant approaches, the working mechanism of brain stimulation can be understood as inducing/decreasing noise in the neural system, which reduces/amplifies the overall signal-to-noise ratio, thereby interfering/facilitating cognitive processing (Miniussi et al., 2013; Hartwigsen and Silvanto, 2023). This is why, for years, most neurostimulation studies have used reaction time and accuracy as dependent variables. If so, could rTMS qualitatively change how meaning is represented? Recent studies combining TMS and multivariate pattern analysis (MVPA) of fMRI data point out in this direction. For instance, Jackson et al. (2021) demonstrated that stimulation of the dorsolateral prefrontal cortex enhances the neural representation of task-irrelevant visual information. Similarly, Rose and collaborators (2016) reported that TMS was able to reactivate the representation of unattended stimuli, including words. By this logic, if M1 represents the motoric (i.e., concrete) features of language, then we believe that applying inhibitory rTMS to this area could potentially hinder the representation of this kind of information, thereby biasing participants to construe meaning in a less motoric (i.e., more abstract) way.

Second, previous neurophysiological studies typically used at least 15–20 MEPs and an intensity of 110–130% rMT to assess changes in corticospinal excitability (e.g., Bastani and Jaberzadeh, 2013; Goldsworthy et al., 2016). Here, we adopted a different strategy: to compare the participants’ motor threshold before and after the stimulation. Since pre-stimulation rMT is commonly assessed using 10 pulses (Rossini et al., 2015), we also used 10 pulses for the post-stimulation rMT. In a similar vein, the intensity for the post-TMS recording was set at 100% rMT to match the intensity used to elicit pre-TMS MEPs. Although the procedure could be improved, our results do show the expected

significant reduction in MEP amplitudes following M1 stimulation.

A final topic to discuss is statistical power. May the failure to find a significant effect be just the result of a lack of power? Although our sample size is larger than most published TMS studies on the topic (see Solana and Santiago, 2022), we agree that it does not appear large, not even reaching the recommended value of 20 participants per condition that we used at preregistration (Simmons et al., 2011). However, rules of thumb are often “overly simplistic recommendations” (Lakens, 2022b, p. 9). Both rules of thumb and analytic procedures such as those implemented in G*power are particularly difficult to apply to designs including within-participant factors, because of the central role that the correlations between repeated measures play in the calculation of power (see, e.g., Cohen, 1988; Brysbaert, 2019). As internal correlations are an aspect of the data that very rarely receives any attention in published reports, researchers usually do not have useful a priori intuitions, but their impact on power is large. As an example, while a two-tailed *t*-test between groups requires 200 participants (100 per group) to detect a small effect of $d = 0.4$ with 80% power, the necessary sample size decreases to 30 participants when the comparison is within-participants and there is a correlation of 0.8 between the dependent measures (correlations of this size or even higher are actually not uncommon, see Brysbaert, 2019). This problem is exacerbated when data are collected in a design that includes more than one random factor (e.g., participants and items) as in the present case. In these cases, Westfall et al. (2014) show that estimating power analytically in a design with a single fixed factor requires the following parameters: (1) effect size, (2) number of participants and items, (3) the specific design to be used, and (4) six variance partitioning components: the variances of each of the random factors (the variance in random intercepts) and their interactions with the conditions (the variance in random slopes), the variance of the interaction between the two random factors, and the residual error variation. We did not feel able to provide reasonable a priori estimations of these variance partitioning parameters, but again, their impact on power calculations is too large to be dismissed. Moreover, when the design includes more than one fixed factor (as in our case), the number of variance-partitioning components to be specified increases, and the analytic solution proposed by Westfall et al. (2014) has not yet been extended to this case.

Together with practical limitations on the total number of participants that was feasible to run in our lab in the available time, we instead adopted a combined approach that overcame these problems, while leaving minimum room for experimenter freedom. A priori, we preregistered a maximum number of participants per condition, which was both feasible and reasonable based on a rule of thumb, plus a sequential analysis strategy with two stopping points and a running Bayesian test of a kind that does not depend on priors set by the experimenter. This allowed us to keep alpha at the desired 5% level but to stop data collection earlier depending on the results. Crucially, after data analysis, we implemented a simulation-based sensitivity analysis using the linear mixed model derived from our own data. Simulations have the advantage of flexibility: they can be adapted to assess any main effect or interaction within a complex design, with any number of participants and items, and even to generalized linear models such as the logit model we used for our binomial measure (see Brysbaert and Stevens, 2018; Kumle et al., 2021). In a word, they provide a power estimation that is tailored to the experiment. When simulations are run using the same model that has been used to analyze the data, it is possible to vary some of its parameters (e.g., effect size, sample size, or number of items) while keeping others constant (variances and covariances). By using this approach, we were able to show that the present study had 80% power to detect what is conventionally considered a small effect size in the critical interaction. We can therefore conclude that there was not a small-sized interaction, or larger, in the data (with a 20% chance of having missed it). This conclusion is also supported by the additional Bayesian analyses. Can we rule out an even smaller interaction? It is of course possible that, with a larger sample and/or stimulus set (and

thereby more power), the experiment might have detected an even smaller effect. In that case, the debate should focus on whether the effect is of any relevance (Anvari et al., 2023), likewise challenging the proposedly essential role of primary motor brain areas in language comprehension. The progress in this debate will, no doubt, benefit from future direct or conceptual replications of the present work and extended discussion about the range of effect sizes that may be considered of theoretical relevance.

For years, psychology and cognitive neuroscience have been on a surprising “winning streak” (Haefffel, 2022), meaning almost all the published studies yielded support for researchers’ predictions. However, replications and meta-analyses showed a radically different reality (e.g., Klein et al., 2018; Open Science Collaboration, 2015). The cause seems to be the combination of several questionable research practices such as small sample sizes (e.g., Button et al., 2013), *p*-hacking (e.g., Simmons et al., 2011), HARKING (e.g., Kerr, 1998), or publication bias (e.g., Rosenthal, 1979). The difficulties to replicate key findings supporting embodiment (Montero-Melis et al., 2022; Morey et al., 2022), together with meta-analytic techniques showing low power and evidence for publication bias (Solana and Santiago, 2022, 2023), reveals that this quandary is also present within the embodied language field. For that reason, we believe that studies like the present one are timely, necessary, and deserve more attention from the embodiment community. The field should start moving towards open and robust research practices such as implementing preregistrations (e.g., Nosek et al., 2018) and well-powered designs (e.g., Lakens, 2022a), replicating previous findings (e.g., Zwaan et al., 2018), validating psychological measures (e.g., Schimmack, 2021), publishing null results (e.g., Munafò and Neill, 2016), and also citing them (e.g., Serra-García and Gneezy, 2021) to reach more credible, valuable, and interpretable outcomes (for similar proposals, see Montero-Melis et al., 2022; Ostarek and Huettig, 2019; Solana, 2023; Solana and Santiago, 2022, 2023).

In conclusion, the present study shows that rTMS to M1 did not induce any significant effect on meaning construction from action language. These results challenge the embodiment tenet that the primary motor system is causally involved in action language understanding.

Funding

This work was supported by the Project PGC2018–096096-B-I00, funded by FEDER/Ministerio de Ciencia e Innovación - Agencia Estatal de Investigación, to J.S. (PI) and D.C.; and by the Project PY20_00689, funded by FEDER/Junta de Andalucía-Consejería de Transformación Económica, Industria, Conocimiento y Universidades, to J.S. (PI). A.B.C. was supported by the Junta de Andalucía (I + D + I Programa 1149 Operativo FEDER Andalucía, 2014–2020, B.SEJ.570.UGR20), MCIN/AEI/(research projects PSI2017-88136 and PID2020- 1147 119033 GB-I00) and by ERDF A way of making Europe. P.S. was supported by a FPU predoctoral grant from the Spanish Ministry of Education (FPU20/01946).

CRedit authorship contribution statement

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

Declaration of competing interest

None.

Data availability

Data, analysis scripts, materials, and supplementary information are publicly available at the OSF repository: <https://osf.io/qrhmu>

Acknowledgements

We thank Cristina Narganes-Pineda for her valuable help with data collection, and Celia Barnés-Castaño for recording the audio files used in the study. We are also grateful to the University of Granada for covering the Open Access fee for this article. The present work is part of the doctoral dissertation of PS at the Psychology PhD Program of the University of Granada under the supervision of JS.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.neuropsychologia.2024.108832>.

References

- Aiello, E.N., Grosso, M., Di Liberto, A., Andriulo, A., Buscone, S., Caracciolo, C., et al., 2022. Disembodying language: actionality does not account for verb processing deficits in Parkinson's disease. *J. Neurolinguistics* 61, 101040. <https://doi.org/10.1016/j.jneuroling.2021.101040>.
- Anvari, F., Kievit, R., Lakens, D., Pennington, C.R., Przybylski, A.K., Tiokhin, L., et al., 2023. Not all effects are indispensable: psychological science requires verifiable lines of reasoning for whether an effect matters. *Perspect. Psychol. Sci.* 18 (2), 503–507. <https://doi.org/10.1177/17456916221091565>.
- Arévalo, A.L., Baldo, J.V., Dronkers, N.F., 2012. What do brain lesions tell us about theories of embodied semantics and the human mirror neuron system? *Cortex* 48 (2), 242–254. <https://doi.org/10.1016/j.cortex.2010.06.001>.
- Argiris, G., Budai, R., Maierson, M., Ius, T., Skrap, M., Tomasino, B., 2020. Neurosurgical lesions to sensorimotor cortex do not impair action verb processing. *Sci. Rep.* 10 (1), 523. <https://doi.org/10.1038/s41598-019-57361-3>.
- Barsalou, L.W., 1999. Perceptual symbol systems. *Behav. Brain Sci.* 22 (4), 577–660. <https://doi.org/10.1017/S0140525X99002149>.
- Barsalou, L.W., 2008. Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645. <https://doi.org/10.1146/annurev.psych.59.103006.093639>.
- Baayen, R.H., Davidson, D.J., Bates, D.M., 2008. Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59 (4), 390–412. <https://doi.org/10.1016/j.jml.2007.12.005>.
- Banks, B., Borghi, A.M., Fargier, R., Fini, C., Jonauskaitė, D., Mazzuca, C., Montalti, M., Villani, C., Woodin, G., 2023. Consensus paper: current perspectives on abstract concepts and future research directions. *Journal of Cognition* 6 (1), 62. <https://doi.org/10.5334/joc.238>.
- Banks, B., Connell, L., 2023. Multi-dimensional sensorimotor grounding of concrete and abstract categories. *Philosophical Transactions of the Royal Society B* 378, 20210366. <https://doi.org/10.1098/rstb.2021.0366>, 1870.
- Barr, D.J., Levy, R., Scheepers, C., Tily, H.J., 2013. Random effects structure for confirmatory hypothesis testing: keep it maximal. *J. Mem. Lang.* 68 (3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>.
- Bastani, A., Jaberzadeh, S., 2013. Differential modulation of corticospinal excitability by different current densities of anodal transcranial direct current stimulation. *PLoS One* 8 (8), e72254. <https://doi.org/10.1371/journal.pone.0072254>.
- Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. *J. Stat. Software* 67 (1), 1–48. <https://doi.org/10.18637/jss.v067.i01>.
- Bergen, B., 2015. Embodiment, simulation and meaning. In: Riemer, N. (Ed.), *The Routledge Handbook of Semantics*. Routledge.
- Binder, J.R., Desai, R.H., 2011. The neurobiology of semantic memory. *Trends Cognit. Sci.* 15 (11), 527–536. <https://doi.org/10.1016/j.tics.2011.10.001>.
- Birba, A., Vitale, F., Padrón, I., Dottori, M., de Vega, M., Zimerman, M., et al., 2020. Electrifying discourse: anodal tDCS of the primary motor cortex selectively reduces action appraisal in naturalistic narratives. *Cortex* 132, 460–472. <https://doi.org/10.1016/j.cortex.2020.08.005>.
- Brybaert, 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>.
- Brybaert, M., Stevens, M., 2018. Power analysis and effect size in mixed effects models: a tutorial. *Journal of Cognition* 1 (1), 9. <https://doi.org/10.5334/joc.10>.
- Boulenger, V., Mechtouff, L., Thobois, S., Broussolle, E., Jeannerod, M., Nazir, T.A., 2008. Word processing in Parkinson's disease is impaired for action verbs but not for concrete nouns. *Neuropsychologia* 46 (2), 743–756. <https://doi.org/10.1016/j.neuropsychologia.2007.10.007>.
- Boyd, L.A., Lindsell, M.A., 2009. Excitatory repetitive transcranial magnetic stimulation to the left dorsal premotor cortex enhances motor consolidation of new skills. *BMC Neurosci.* 10, 72. <https://doi.org/10.1186/1471-2202-10-72>.
- Buccino, G., Riggio, L., Melli, G., Binkofski, F., Gallese, V., Rizzolatti, G., 2005. Listening to action-related sentences modulates the activity of the motor system: a combined TMS and behavioral study. *Cognit. Brain Res.* 24 (3), 355–363. <https://doi.org/10.1016/j.cogbrainres.2005.02.020>.
- Buccino, G., Colagè, I., Silipo, F., D'Ambrosio, P., 2019. The concreteness of abstract language: an ancient issue and a new perspective. *Brain Struct. Funct.* 224 (4), 1385–1401. <https://doi.org/10.1007/s00429-019-01851-7>.
- 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. *Nat. Rev. Neurosci.* 14 (5), 365–376. <https://doi.org/10.1038/nrn3475>.
- Casasanto, D., 2023. Embodied semantics. In: Li, F.T. (Ed.), *Handbook of Cognitive Semantics*. Brill, Leiden.
- Chatterjee, A., 2010. Disembodying cognition. *Lang. Cognit.* 2 (1), 79–116. <https://doi.org/10.1515/langcog.2010.004>.
- Chen, H., Cohen, P., Chen, S., 2010. How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. *Commun. Stat. Simulat. Comput.* 39 (4), 860–864. <https://doi.org/10.1080/03610911003650383>.
- 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>.
- Cohen, J., 1988. *Statistical Power Analysis for the Behavioral Sciences*. L. Erlbaum Associates.
- Courson, M., Macoir, J., Tremblay, P., 2017. Role of medial premotor areas in action language processing in relation to motor skills. *Cortex* 95, 77–91. <https://doi.org/10.1016/j.cortex.2017.08.002>.
- Das, K.R., Imon, A.H.M.R., 2016. A brief review of tests for normality. *Am. J. Theor. Appl. Stat.* 5 (1), 5–12. <https://doi.org/10.11648/j.ajtas.20160501.12>.
- de Zubicaray, G.I., Arciuli, J., McMahon, J., 2013. Putting an “end” to the motor cortex representations of action words. *J. Cognit. Neurosci.* 25 (11), 1957–1974. https://doi.org/10.1162/jocn_a_00437.
- de Zubicaray, G.I., McMahon, K.L., Arciuli, J., 2021. A sound explanation for motor cortex engagement during action word comprehension. *J. Cognit. Neurosci.* 33 (1), 129–145. https://doi.org/10.1162/jocn_a_01640.
- Dienes, Z., 2014. Using Bayes to get the most out of non-significant results. *Front. Psychol.* 5, 781. <https://doi.org/10.3389/fpsyg.2014.00781>.
- Dreyer, F.R., Picht, T., Frey, D., Vajkoczy, P., Pulvermüller, F., 2020. The functional relevance of dorsal motor systems for processing tool nouns—evidence from patients with focal lesions. *Neuropsychologia* 141, 107384. <https://doi.org/10.1016/j.neuropsychologia.2020.107384>.
- Fernandino, L., Conant, L.L., Binder, J.R., Blindauer, K., Hiner, B., Spangler, K., Desai, R.H., 2013. Parkinson's disease disrupts both automatic and controlled processing of action verbs. *Brain Lang.* 127 (1), 65–74. <https://doi.org/10.1016/j.bandl.2012.07.008>.
- Fodor, J.A., 1975. *The Language of Thought*. Harvard University Press.
- Fox, J., Weisberg, S., 2019. *An R Companion to Applied Regression*. SAGE.
- Gallese, V., Lakoff, G., 2005. The brain's concepts: the role of the sensory-motor system in conceptual knowledge. *Cogn. Neuropsychol.* 22 (3–4), 455–479. <https://doi.org/10.1080/02643290442000310>.
- García, A.M., Bocanegra, Y., Herrera, E., Moreno, L., Carmona, J., Baena, A., et al., 2018. Parkinson's disease compromises the appraisal of action meanings evoked by naturalistic texts. *Cortex* 100, 111–126. <https://doi.org/10.1016/j.cortex.2017.07.003>.
- García, A.M., Ibáñez, A., 2016. A touch with words: dynamic synergies between manual actions and language. *Neurosci. Biobehav. Rev.* 68, 59–95. <https://doi.org/10.1016/j.neubiorev.2016.04.022>.
- Gentner, R., Wankerl, K., Reinsberger, C., Zeller, D., Classen, J., 2008. Depression of human corticospinal excitability induced by magnetic theta-burst stimulation: evidence of rapid polarity-reversing metaplasticity. *Cerebr. Cortex* 18 (9), 2046–2053. <https://doi.org/10.1093/cercor/bhm239>.
- Gijssels, T., Ivry, R.B., Casasanto, D., 2018. tDCS to premotor cortex changes action verb understanding: complementary effects of inhibitory and excitatory stimulation. *Sci. Rep.* 8 (1), 1–7. <https://doi.org/10.1038/s41598-018-29600-6>.
- Glenberg, A.M., Sato, M., Cattaneo, L., Riggio, L., Palumbo, D., Buccino, G., 2008a. Processing abstract language modulates motor system activity. *Q. J. Exp. Psychol.* 61 (6), 905–919. <https://doi.org/10.1080/17470210701625550>.
- Glenberg, A.M., Sato, M., Cattaneo, L., 2008b. Use-induced motor plasticity affects the processing of abstract and concrete language. *Curr. Biol.* 18 (7), R290–R291. <https://doi.org/10.1016/j.cub.2008.02.036>.
- Goldsworthy, M.R., Hordacre, B., Ridding, M.C., 2016. Minimum number of trials required for within- and between-session reliability of TMS measures of corticospinal excitability. *Neuroscience* 320, 205–209. <https://doi.org/10.1016/j.neuroscience.2016.02.012>.
- Green, P., MacLeod, C.J., 2016. simr: an R package for power analysis of generalised linear mixed models by simulation. *Methods Ecol. Evol.* 7 (4), 493–498. <https://doi.org/10.1111/2041-210X.12504>.
- Haefel, G.J., 2022. Psychology needs to get tired of winning. *R. Soc. Open Sci.* 9 (6), 220099. <https://doi.org/10.1098/rsos.220099>.
- Hardwicke, T.E., Wagenmakers, E.J., 2023. Reducing bias, increasing transparency and calibrating confidence with preregistration. *Nat. Human Behav.* 7 (1), 15–26. <https://doi.org/10.1038/s41562-022-01497-2>.
- Harpaintner, M., Sim, E.J., Trumpp, N.M., Ulrich, M., Kiefer, M., 2020. The grounding of abstract concepts in the motor and visual system: an fMRI study. *Cortex* 124, 1–22. <https://doi.org/10.1016/j.cortex.2019.10.014>.
- 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. *Psychol. Res.* 86 (8), 2560–2582. <https://doi.org/10.1007/s00426-020-01374-5>.
- Harpaz, Y., Levkovitz, Y., Lavidor, M., 2009. Lexical ambiguity resolution in Wernicke's area and its right homologue. *Cortex* 45 (9), 1097–1103. <https://doi.org/10.1016/j.cortex.2009.01.002>.
- Hartwigsen, G., Silvanto, J., 2023. Noninvasive brain stimulation: multiple effects on cognition. *Neuroscientist* 29 (5), 639–653. <https://doi.org/10.1177/10738584221113806>.
- 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](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. *Hum. Brain Mapp.* 21 (3), 191–201. <https://doi.org/10.1002/hbm.10157>.
- Huang, Y.Z., Edwards, M.J., Rounis, E., Bhatia, K.P., Rothwell, J.C., 2005. Theta burst stimulation of the human motor cortex. *Neuron* 45 (2), 201–206. <https://doi.org/10.1016/j.neuron.2004.12.033>.
- Humphries, S., Klooster, N., Cardillo, E., Weintraub, D., Rick, J., Chatterjee, A., 2019. From action to abstraction: the sensorimotor grounding of metaphor in Parkinson's disease. *Cortex* 121, 362–384. <https://doi.org/10.1016/j.cortex.2019.09.005>.
- 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>.
- Jackson, J.B., Feredoes, E., Rich, A.N., Lindner, M., Woolgar, A., 2021. Concurrent neuroimaging and neurostimulation reveals a causal role for dlPFC in coding of task-relevant information. *Commun. Biol.* 4 (1), 588. <https://doi.org/10.1038/s42003-021-02109-x>.
- Jaeger, T.F., 2008. Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed models. *J. Mem. Lang.* 59 (4), 434–446. <https://doi.org/10.1016/j.jml.2007.11.007>.
- Kemmerer, D., Rudrauf, D., Manzel, K., Tranel, D., 2012. Behavioral patterns and lesion sites associated with impaired processing of lexical and conceptual knowledge of actions. *Cortex* 48 (7), 826–848. <https://doi.org/10.1016/j.cortex.2010.11.001>.
- Kemmerer, D., Miller, L., MacPherson, M.K., Huber, J., Tranel, D., 2013. An investigation of semantic similarity judgments about action and non-action verbs in Parkinson's disease: implications for the Embodied Cognition Framework. *Front. Hum. Neurosci.* 7, 146. <https://doi.org/10.3389/fnhum.2013.00146>.
- Kerr, N.L., 1998. HARKing: hypothesizing after the results are known. *Pers. Soc. Psychol. Rev.* 2 (3), 196–217. https://doi.org/10.1207/s15327957pspr0203_4.
- Kiefer, M., Pulvermüller, F., 2012. Conceptual representations in mind and brain: theoretical developments, current evidence and future directions. *Cortex* 48 (7), 805–825. <https://doi.org/10.1016/j.cortex.2011.04.006>.
- Klein, R.A., Vianello, M., Hasselman, F., Adams, B.G., Adams Jr., R.B., Alper, S., et al., 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>.
- Kumle, L., Vö, M.L.-H., Draschkow, D., 2021. Estimating power in (generalized) linear mixed models: an open introduction and tutorial in R. *Behav. Res. Methods* 53 (6), 2528–2543. <https://doi.org/10.3758/s13428-021-01546-0>.
- Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B., 2017. lmerTest package: tests in linear mixed effects models. *J. Stat. Software* 82 (13), 1–26. <https://doi.org/10.18637/jss.v082.i13>.
- Lakens, D., 2014. Performing high-powered studies efficiently with sequential analyses. *Eur. J. Soc. Psychol.* 44 (7), 701–710. <https://doi.org/10.1002/ejsp.2023>.
- Lakens, D., 2022a. Improving your statistical inferences. <https://doi.org/10.5281/ZENODO.6409077>.
- Lakens, D., 2022b. Sample size justification. *Collabra Psychology* 8 (1), 33267. <https://doi.org/10.1525/collabra.33267>.
- Lee, M.D., Wagenmakers, E.J., 2014. *Bayesian Cognitive Modeling: A Practical Course*. Cambridge University Press.
- Lo Gerfo, E., Oliveri, M., Torriero, S., Salerno, S., Koch, G., Caltagirone, C., 2008. The influence of rTMS over prefrontal and motor areas in a morphological task: grammatical vs. semantic effects. *Neuropsychologia* 46 (2), 764–770. <https://doi.org/10.1016/j.neuropsychologia.2007.10.012>.
- Mac Giolla, E., Luke, T.J., Calderon, S., Ask, K., 2022. Validating Measures of Mental Abstraction. <https://doi.org/10.31234/osf.io/v6xt4>. PysArXiv.
- Mahon, B.Z., Caramazza, A., 2008. A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *J. Physiol. Paris* 102 (1–3), 59–70. <https://doi.org/10.1016/j.jphysparis.2008.03.004>.
- Martín-Signes, M., Pérez-Serrano, C., Chica, A.B., 2019. Causal contributions of the SMA to alertness and consciousness interactions. *Cerebr. Cortex* 29 (2), 648–656. <https://doi.org/10.1093/cercor/bhx346>.
- Masson, M.E., 2011. A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behav. Res. Methods* 43, 679–690. <https://doi.org/10.3758/s13428-010-0049-5>.
- 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>.
- Miniussi, C., Harris, J.A., Ruzzoli, M., 2013. Modelling non-invasive brain stimulation in cognitive neuroscience. *Neurosci. Biobehav. Rev.* 37 (8), 1702–1712. <https://doi.org/10.1016/j.neubiorev.2013.06.014>.
- 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>.
- Morey, R.D., Kaschak, M.P., Díez-Álamo, A.M., Glenberg, A.M., Zwaan, R.A., Lakens, D., Ibáñez, A., García, A., Gianelli, C., Jones, J.L., Madden, J., Alifano, F., Bergen, B., Bloxson, N.G., Bub, D.N., Cai, Z.C., Chartier, C.R., Chatterjee, A., Conwell, E., et al., 2022. A pre-registered, multi-lab non-replication of the action-sentence compatibility effect (ACE). *Psychonomic Bull. Rev.* 29, 613–626. <https://doi.org/10.3758/s13423-021-01927-8>.
- Muellbacher, W., Ziemann, U., Boroojerdi, B., Hallett, M., 2000. Effects of low-frequency transcranial magnetic stimulation on motor excitability and basic motor behavior. *Clin. Neurophysiol.* 111 (6), 1002–1007. [https://doi.org/10.1016/S1388-2457\(00\)00284-4](https://doi.org/10.1016/S1388-2457(00)00284-4).
- Munafò, M., Neill, J., 2016. Null is beautiful: on the importance of publishing null results. *J. Psychopharmacol.* 30 (7) <https://doi.org/10.1177/0269881116638813>, 585–585.
- Nosek, B.A., Ebersole, C.R., DeHaven, A.C., Mellor, D.T., 2018. The preregistration revolution. *Proc. Natl. Acad. Sci. USA* 115 (11), 2600–2606. <https://doi.org/10.1073/pnas.1708274114>.
- Open Science Collaboration, 2015. Estimating the reproducibility of psychological science. *Science* 349 (6251). <https://doi.org/10.1126/science.aac4716>.
- Ostarek, M., Bottini, R., 2021. Towards strong inference in research on embodiment - possibilities and limitations of causal paradigms. *Journal of Cognition* 4 (1), 5. <https://doi.org/10.5334/joc.139>.
- Ostarek, M., Huettig, F., 2019. Six challenges for embodiment research. *Curr. Dir. Psychol. Sci.* 28 (6), 593–599. <https://doi.org/10.1177/0963721419866441>.
- Ostarek, M., Joosen, D., Ishag, A., De Nijs, M., Huettig, F., 2019. Are visual processes causally involved in “perceptual simulation” effects in the sentence-picture verification task? *Cognition* 182, 84–94. <https://doi.org/10.1016/j.cognition.2018.08.017>.
- Papeo, L., Vallesi, A., Isaja, A., Rumiati, R.I., 2009. Effects of TMS on different stages of motor and non-motor verb processing in the primary motor cortex. *PLoS One* 4 (2), e4508. <https://doi.org/10.1371/journal.pone.0004508>.
- Papeo, L., Corradi-Dell'Acqua, C., Rumiati, R.I., 2011. “She” is not like “I”: the tie between language and action is in our imagination. *J. Cognit. Neurosci.* 23 (12), 3939–3948. https://doi.org/10.1162/jocn_a.00075.
- Papesh, M.H., 2015. Just out of reach: on the reliability of the action-sentence compatibility effect. *J. Exp. Psychol. Gen.* 144 (6), e116. <https://doi.org/10.1037/xge0000125>.
- Polanía, R., Nitsche, M.A., Ruff, C.C., 2018. Studying and modifying brain function with non-invasive brain stimulation. *Nat. Neurosci.* 21, 174–187. <https://doi.org/10.1038/s41593-017-0054-4>.
- Postle, N., McMahon, K.L., Ashton, R., Meredith, M., de Zubicaray, G.I., 2008. Action word meaning representations in cytoarchitecturally defined primary and premotor cortices. *Neuroimage* 43 (3), 634–644. <https://doi.org/10.1016/j.neuroimage.2008.08.006>.
- 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. *Front. Hum. Neurosci.* 7, 11. <https://doi.org/10.3389/fnhum.2013.00011>.
- Pulvermüller, F., 2005. Brain mechanisms linking language and action. *Nat. Rev. Neurosci.* 6 (7), 576–582. <https://doi.org/10.1038/nrn1706>.
- Pulvermüller, F., Hauk, O., Nikulin, V.V., Ilmoniemi, R.J., 2005. Functional links between motor and language systems. *Eur. J. Neurosci.* 21 (3), 793–797. <https://doi.org/10.1111/j.1460-9568.2005.03900.x>.
- R Core Team, 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Reilly, J., Harnish, S., Garcia, A., Hung, J., Rodriguez, A.D., Crosson, B., 2014. Lesion symptom mapping of manipulable object naming in nonfluent aphasia: can a brain be both embodied and disembodied? *Cogn. Neuropsychol.* 31 (4), 287–312. <https://doi.org/10.1080/02643294.2014.914022>.
- Repetto, C., Colombo, B., Cipresso, P., Riva, G., 2013. The effects of rTMS over the primary motor cortex: the link between action and language. *Neuropsychologia* 51 (1), 8–13. <https://doi.org/10.1016/j.neuropsychologia.2012.11.001>.
- Rose, N.S., LaRocque, J.J., Riggall, A.C., Gosseries, O., Starrett, M.J., Meyering, E.E., Postle, B.R., 2016. Reactivation of latent working memories with transcranial magnetic stimulation. *Science* 354 (6316), 1136–1139. <https://doi.org/10.1126/science.aah7011>.
- Rosenthal, R., 1979. The file drawer problem and tolerance for null results. *Psychol. Bull.* 86 (3), 638–641. <https://doi.org/10.1037/0033-2909.86.3.638>.
- Rossi, S., Antal, A., Bestmann, S., Bikson, M., Brewer, C., Brockmüller, J., et al., 2021. Safety and recommendations for TMS use in healthy subjects and patient populations, with updates on training, ethical and regulatory issues: expert Guidelines. *Clin. Neurophysiol.* 132 (1), 269–306. <https://doi.org/10.1016/j.clinph.2020.10.003>.
- Rossini, P.M., Burke, D., Chen, R., Cohen, L.G., Daskalakis, Z., Di Iorio, R., et al., 2015. Non-invasive electrical and magnetic stimulation of the brain, spinal cord, roots and peripheral nerves: basic principles and procedures for routine clinical and research application. An updated report from an IFCN Committee. *Clin. Neurophysiol.* 126 (6), 1071–1107. <https://doi.org/10.1016/j.clinph.2015.02.001>.
- Saccone, E.J., Thomas, N.A., Nicholls, M.E., 2021. One-handed motor activity does not interfere with naming lateralized pictures of tools. *J. Exp. Psychol. Hum. Percept. Perform.* 47 (4), 529. <https://doi.org/10.1037/xhp0000863>.
- Sakreida, K., Scorolli, C., Menz, M.M., Heim, S., Borghi, A.M., Binkofski, F., 2013. Are abstract action words embodied? An fMRI investigation at the interface between language and motor cognition. *Front. Hum. Neurosci.* 7, 125. <https://doi.org/10.3389/fnhum.2013.00125>.
- Schimmack, U., 2021. The validation crisis in psychology. *Meta-Psychology* 2019 (5), 1645. <https://doi.org/10.15626/MP.2019.1645>.

- Schönbrodt, F.D., Wagenmakers, E.J., 2018. Bayes factor design analysis: planning for compelling evidence. *Psychonomic Bull. Rev.* 25 (1), 128–142. <https://doi.org/10.3758/s13423-017-1230-y>.
- Scott, S.H., 2012. The computational and neural basis of voluntary motor control and planning. *Trends Cognit. Sci.* 16 (11), 541–549. <https://doi.org/10.1016/j.tics.2012.09.008>.
- Serra-García, M., Gneezy, U., 2021. Nonreplicable publications are cited more than replicable ones. *Sci. Adv.* 7 (21), eabd1705 <https://doi.org/10.1126/sciadv.abd1705>.
- Shebani, Z., Pulvermüller, F., 2013. Moving the hands and feet specifically impairs working memory for arm-and leg-related action words. *Cortex* 49 (1), 222–231. <https://doi.org/10.1016/j.cortex.2011.10.005>.
- Shtyrov, Y., Butorina, A., Nikolaeva, A., Stroganova, T., 2014. Automatic ultrarapid activation and inhibition of cortical motor systems in spoken word comprehension. *Proc. Natl. Acad. Sci. USA* 111 (18), E1918–E1923. <https://doi.org/10.1073/pnas.1323158111>.
- Signorell, A., 2017. *DescTools: Tools for Descriptive Statistics. R Package Version 0.99.20. CRAN.*
- Simmons, J.P., Nelson, L.D., Simonsohn, U., 2011. False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol. Sci.* 22 (11), 1359–1366. <https://doi.org/10.1177/0956797611417632>.
- Speed, L.J., Iravani, B., Lundström, J.N., Majid, A., 2022. Losing the sense of smell does not disrupt processing of odor words. *Brain Lang.* 235, 105200 <https://doi.org/10.1016/j.bandl.2022.105200>.
- Solana, P., 2023. The embodiment and disembodiment of language. *Nature Reviews Psychology* 2, 391. <https://doi.org/10.1038/s44159-023-00203-x>.
- 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. *Neurosci. Biobehav. Rev.* 141, 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. *Psicologica* 44 (2), e15661. <https://doi.org/10.20350/digitalCSIC/15661>.
- Strozyk, J.V., Dudschig, C., Kaup, B., 2019. Do I need to have my hands free to understand hand-related language? Investigating the functional relevance of experiential simulations. *Psychol. Res.* 83, 406–418. <https://doi.org/10.1007/s00426-017-0900-8>.
- Togato, G., Andras, F., Miralles, E., Macizo, P., 2021. Motor processing modulates word comprehension. *Br. J. Psychol.* 112 (4), 1028–1052. <https://doi.org/10.1111/bjop.12507>.
- Tomasino, B., Fink, G.R., Sparing, R., Dafotakis, M., Weiss, P.H., 2008. Action verbs and the primary motor cortex: a comparative TMS study of silent reading, frequency judgments, and motor imagery. *Neuropsychologia* 46 (7), 1915–1926. <https://doi.org/10.1016/j.neuropsychologia.2008.01.015>.
- Trans Cranial Technologies, 2012. 10/20 System Positioning Manual. https://transcranial.com/docs/10_20_pos_man_v1_0.pdf.pdf.
- Tremblay, P., Sato, M., Small, S.L., 2012. TMS-induced modulation of action sentence priming in the ventral premotor cortex. *Neuropsychologia* 50 (2), 319–326. <https://doi.org/10.1016/j.neuropsychologia.2011.12.002>.
- Trope, Y., Liberman, N., 2010. Construal-level theory of psychological distance. *Psychol. Rev.* 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. *Psychol. Rev.* 94 (1), 3–15. <https://doi.org/10.1037/0033-295X.94.1.3>.
- Vallacher, R.R., Wegner, D.M., 1989. Levels of personal agency: individual variation in action identification. *J. Pers. Soc. Psychol.* 57 (4), 660–671. <https://doi.org/10.1037/0022-3514.57.4.660>.
- Vitale, F., Padrón, I., Avenanti, A., de Vega, M., 2021. Enhancing motor brain activity improves memory for action language: a tDCS study. *Cerebr. Cortex* 31 (3), 1569–1581. <https://doi.org/10.1093/cercor/bhaa309>.
- 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>.
- Wagenmakers, E.-J., 2007. A practical solution to the pervasive problems of *p* values. *Psychonomic Bull. Rev.* 14, 779–804. <https://doi.org/10.3758/BF03194105>.
- Wald, A., 2004. *Sequential Analysis.* Courier Corporation.
- Watson, C.E., Cardillo, E.R., Ianni, G.R., Chatterjee, A., 2013. Action concepts in the brain: an activation likelihood estimation meta-analysis. *J. Cognit. Neurosci.* 25 (8), 1191–1205. https://doi.org/10.1162/jocn_a.00401.
- Westfall, J., Kenny, D. a, Judd, C.M., 2014. Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *J. Exp. Psychol. Gen.* 143 (5), 2020–2045. <https://doi.org/10.1037/xge0000014>.
- Willems, R.M., Hagoort, P., Casasanto, D., 2010a. Body-specific representations of action verbs: neural evidence from right-and left-handers. *Psychol. Sci.* 21 (1), 67–74. <https://doi.org/10.1177/0956797609354072>.
- Willems, R.M., Toni, I., Hagoort, P., Casasanto, D., 2010b. Neural dissociations between action verb understanding and motor imagery. *J. Cognit. Neurosci.* 22 (10), 2387–2400. <https://doi.org/10.1162/jocn.2009.21386>.
- 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. *Psychol. Sci.* 22 (7), 849–854. <https://doi.org/10.1177/0956797611412387>.
- Witt, J.K., Kemmerer, D., Linkenauger, S.A., Culham, J.C., 2020. Reanalysis suggests evidence for motor simulation in naming tools is limited: a commentary on Witt, kemmerer, linkenauger, and culham (2010). *Psychol. Sci.* 31 (8), 1036–1039. <https://doi.org/10.1177/0956797620940555>.
- Zwaan, R.A., Etz, A., Lucas, R.E., Donnellan, M.B., 2018. Making replication mainstream. *Behav. Brain Sci.* 41, E120. <https://doi.org/10.1017/S0140525X17001972>.