Psychological Research
https://doi.org/10.1007/s00426-020-01309-0

ORIGINAL ARTICLE

Is 'heavy' up or down? Testing the vertical spatial representation of weight

Michele Vicovaro[1] · **Mario Dalmaso**[2]

Received: 12 October 2019 / Accepted: 19 February 2020

Abstract
Smaller numbers are typically responded to faster with a bottom than a top key, whereas the opposite occurs for larger numbers (a vertical spatial–numerical association of response codes: i.e. the vertical SNARC effect). Here, in four experiments, we explored whether a vertical spatial–magnitude association can emerge for lighter vs. heavier items. Participants were presented with a central target stimulus that could be a word describing a material (e.g. 'paper', 'iron': Experiment 1), a numerical quantity of weight (e.g. '1 g', '1 kg': Experiment 2) or a picture associated with a real object that participants weighed before the experiment (Experiments 3a/3b). Participants were asked to respond either to the weight (Experiments 1–3a) or to the size (i.e. weight was task-irrelevant; Experiment 3b) of the stimuli by pressing vertically placed keys. In Experiments 1 and 2, faster responses emerged for the lighter-bottom/heavier-top mapping—in line with a standard SNARC-like effect—whereas in Experiment 3a the opposite mapping emerged (lighter-top/heavier-bottom). No evidence of an implicit weight-space association emerged in Experiment 3b. Overall, these results provide evidence indicating a possible context-dependent vertical spatial representation of weight.

Keywords SQUARC effect · SNARC-like effect · Spatial representation · Weight judgement · SWARC effect

Introduction

A consistent finding in the numerical cognition literature is the existence of an association between numbers and space. In general, smaller numbers are associated with the left and the bottom location of space, whereas larger numbers are associated with the right and the top location of space (for reviews see Toomarian & Hubbard, 2018; Winter, Matlock, Shaki, & Fischer, 2015). Similarly, non-numerical magnitudes (e.g. time, luminance, size) also appear to be spatially coded, with smaller and larger magnitudes associated with the left/bottom and the right/top location of space, respectively. However, so far, the possible vertical spatial mappings of such non-numerical dimensions have been scarcely investigated. The current work was aimed at testing the vertical spatial representation of weight, a crucial dimension of both the physical and phenomenal world whose cognitive representation is still largely unexplored.

Horizontal and vertical representations of numerical and non-numerical quantities

One of the first findings revealing that quantities arrange along spatial dimensions comes from the seminal study by Dehaene, Dupoux, and Mehler (1990), in which participants pressed a left or a right key to classify a centrally placed number as either smaller or larger than a reference number (i.e. a magnitude comparison task). Faster responses were recorded when the mapping between numbers and response keys was compatible with a left-to-right spatial representation of numbers (i.e. smaller-left, greater-right) rather than incompatible (i.e. smaller-right, greater-left), a phenomenon known as the *Spatial-Numerical Association of Response Codes* (SNARC) effect. Interestingly, a similar pattern of results also emerged when participants classified the centrally placed number as either even or odd (i.e. a parity judgement task), thus suggesting that numerical magnitude was still processed even when it was irrelevant to the task (Dehaene, Bossini, & Giraux, 1993; for a meta-analysis and

✉ Mario Dalmaso
mario.dalmaso@unipd.it; mario.dalmaso@gmail.com

[1] Department of General Psychology, University of Padova, Padova, Italy

[2] Department of Developmental and Social Psychology, University of Padova, via Venezia 8, 35131 Padova, Italy

review of the SNARC effect see Wood, Willmes, Nuerk, & Fischer, 2008). Furthermore, the SNARC effect has been found even when the parity judgement task was performed using response keys aligned vertically rather than horizontally. For instance, Müller and Schwarz (2007) reported faster responses when the mapping between the numbers and response keys was compatible with a bottom-to-top spatial representation of numbers (i.e. smaller-bottom, greater-top) rather than incompatible (i.e. smaller-top, greater-bottom). Importantly, similar results have been also reported in other studies employing different experimental settings (e.g. Gevers, Lammertyn, Notebaert, Verguts, & Fias, 2006; Hartmann, Gashaj, Stahnke, & Mast, 2014; Hesse & Bremmer, 2017; Ito & Hatta, 2004; Schwarz & Keus, 2004; see also Winter et al., 2015).

Intriguingly, it has been observed that even non-numerical dimensions (e.g. time, luminance, size) can elicit SNARC-like effects (for a meta-analysis and review see Macnamara, Keage, & Loetscher, 2018). Indeed—in line with the SNARC effect—when participants are required to press a left or a right key to classify a non-numerical stimulus as either smaller or larger in magnitude than a reference stimulus, faster responses are typically recorded when the mapping between magnitudes and response keys is compatible with a left-to-right spatial representation of magnitudes (i.e. smaller-left, greater-right) rather than incompatible (i.e. smaller-right, greater-left). So far, horizontal SNARC-like effects have been reported for pitch height (e.g. Cho, Bae, & Proctor, 2012; Lidji, Kolinski, Lochy, & Morais, 2007; Rusconi, Kwan, Giordano, Umiltà, & Butterworth, 2006), loudness (e.g. Chang & Cho, 2015; Hartmann & Mast, 2017), luminance (e.g. Fumarola et al., 2014; Ren, Nicholls, Ma, & Chen, 2011), size (e.g. Prpic et al., in press; Ren et al., 2011; Sellaro, Treccani, Job, & Cubelli, 2015), weight (Dalmaso & Vicovaro, 2019) and time (Ishihara, Keller, Rossetti, & Prinz, 2008; Vallesi, Binns, & Shallice, 2008; Zhao et al., 2018). For instance, Ren et al. (2011) asked participants to decide whether a target shape was either smaller or larger compared to a reference shape (Experiment 2) or whether a target word referred to an object that was either smaller or larger compared to an object described by a reference word (Experiment 4). Evidence of a SNARC-like effect emerged in both experiments.

Almost surprisingly, to date only a few studies have explored the possible existence of SNARC-like effects along the vertical dimension. In this regard, Rusconi et al. (2006; see also Lidji et al., 2007) showed that high-pitch stimuli were responded to faster with a top key than with a bottom key, whereas the opposite was true for low-pitch stimuli. Similar results were reported by Bruzzi, Talamini, Priftis, and Grassi (2017) for stimuli varying in loudness rather than pitch (see also Fernandez-Prieto, Spence, Pons, & Navarra, 2017). These studies seem to suggest that both pitch and loudness are mapped along a bottom-to-top direction. However, beyond the domain of auditory stimuli, the existence of a vertical SNARC-like effect remains uncertain. Indeed, on the one hand, Ishihara et al. (2008) did not find any evidence of a vertical SNARC-like effect when participants were asked to classify the time onset of a stimulus with respect to a reference interval (i.e. 'early' vs. 'late'), suggesting that time intervals might not be represented along the vertical dimension. On the other hand, Sell and Kaschak (2012) found a SNARC-like effect for sentences describing abstract concepts of quantity. Specifically, sentences describing 'more' quantity were responded to faster with a top key than with a bottom key, whereas the opposite was true for sentences describing 'less' quantity (for other effects related to the spatial representation of concepts see also, for instance, Estes, Verges, & Barsalou, 2008).

Explaining SNARC and SNARC-like effects: ATOM, the polarity correspondence model and TEST

The nature of SNARC and SNARC-like effects is still widely debated and, so far, three main theories have been proposed to explain the relationship between magnitudes and space. Firstly, the theoretical framework known as *A Theory of Magnitude* (ATOM; Walsh, 2003, 2015) proposes that space, time and numbers are linked by a common underlying mechanism devoted to magnitude processing. Hence, in line with ATOM, the SNARC effect would be a specific instantiation of a broader *Spatial-Quantity Association of Response Codes* (SQUARC) effect, according to which a relationship between space and magnitude should emerge for any spatially- or action-coded dimension (see also Cantlon, Pratt, & Brannon, 2009; Cohen Kadosh, Lammertyn, & Izard, 2008).

Another well-known explanation for SNARC and SNARC-like effects is provided by the *polarity correspondence model* (Proctor & Cho, 2006). According to this model, when participants are required to classify stimulus magnitude through lateralized responses, both stimuli and responses would be implicitly coded. In more detail, smaller magnitudes and left/bottom responses would be coded as negative polarities, whereas larger magnitudes and right/top responses would be coded as positive polarities. Consequently, responses would be faster when the stimulus and the response polarities are identical (i.e. small-left/bottom and large-right/top) compared to when they are different (i.e. small-right/top and large-left/bottom). According to this model, the association between magnitudes and space would be a consequence of the structural features of the mental representation of conceptual dimensions (see Santiago & Lakens, 2015). In other words, SNARC and SNARC-like effects would not arise because of an intrinsic association between magnitudes and space—as suggested by ATOM—but would arise from a more general mechanism of conceptual correspondence. Despite theoretical differences, ATOM and the polarity correspondence model make identical predictions regarding SNARC and SNARC-like effects because both models predict a smaller-left/bottom and a larger-right/top association between magnitudes and space. Nevertheless, the results of recent studies have cast doubt on the hypothesis that polarity correspondence may provide an exhaustive account of the SNARC effect (see Di Rosa et al., 2017; Leth-Steensen & Citta, 2016; Santiago & Lakens, 2015).

The third theory proposed to explain the possible nature of SNARC and SNARC-like effects is the *Tropic, Embodied, and Situated Theory of cognition* (TEST: Myachykov, Scheepers, Fischer, & Kessler, 2014). This theory can be integrated within the broader theoretical framework of grounded cognition, according to which cognition would be shaped by simulations, situated actions and bodily constraints (e.g. Barsalou, 2008; see also, for instance, Shapiro, 2019). In more detail, TEST states that our cognitive representations can be hierarchically organized into three levels: tropic (or grounded), embodied and situated. Tropic representations would be shaped by stable features and constraints of the physical world and, therefore, they would be relatively unmodifiable. Embodied representations would emerge from perceptual-motor experience and be shaped both by bodily and physical constraints. Embodied representations would be less stable and universal than tropic representations because they may vary according to individual bodily features and specific constraints imposed by the specific body–environment interactions. At the bottom of the hierarchy, situated representations would be shaped by specific contextual and task features. In contrast to both ATOM and the polarity correspondence model, TEST provides two different explanations for horizontal and vertical SNARC. On the one hand, the horizontal SNARC effect would arise from a *situated* left-to-right mapping of numbers, emerging because of specific task demands such as the employment of lateralized response keys (see Bächtold, Baumüller, & Brugger, 1998; Fischer, Mills, & Shaki, 2010; Notebaert, Gevers, Verguts, & Fias, 2006; Shaki & Fischer, 2018; Sixtus, Lonnemann, Fischer, & Werner, 2019). On the other hand, the vertical SNARC effect would arise from a *tropic* representation of quantities due to constraints of the physical world. For instance, the universal action of piling up objects along the vertical dimension would lead to a stable association in which ‘more is up’ and ‘less is down’ (see Fischer, 2012; Shaki & Fischer, 2012, 2018; Sixtus et al., 2019; Winter et al., 2015; cf. Hartmann et al., 2014; Holmes & Lourenco, 2012; Hung, Hung, Tzeng, & Wu, 2008).

A possible tropic/embodied vertical representation of weight?

Even though weight can strongly affect our interactions with the physical world, the way it is treated and represented within our cognitive system is largely unknown. To the best of our knowledge, only two studies have explored the possible spatial representation of weight, but they only focused on the horizontal dimension. In the first study on this topic (Holmes & Lourenco, 2013), participants completed a standard parity judgement task (e.g. Dehaene et al., 1993) while wearing a weight on either the left or the right wrist (i.e. left and right condition, respectively). In a baseline condition, no weight was employed. Interestingly, the SNARC effect emerged in both the baseline and the *right* conditions but not the *left* condition. According to the authors, the *left* condition elicited a right-to-left representation of quantity (i.e. the right side was associated with a lighter weight, the left side was associated with a heavier weight), which counteracted the left-to-right representation of numbers, thus nullifying the SNARC effect. In the second study on this topic (Dalmaso & Vicovaro, 2019), the possible horizontal representation of weight was explored through a magnitude comparison task. Specifically, participants were asked to judge whether the weight associated with a target word (e.g. mouse) was either lighter or heavier compared to the weight associated with a reference word (e.g. sheep). As a main result, ‘heavier’ responses were faster when provided with a right rather than a left key, whereas no such difference emerged for ‘lighter’ responses, thus providing supporting evidence for the existence of a horizontal SNARC-like effect for weight (i.e. a *Space-Weight Association of Response Codes*: a SWARC effect).

Here, we systematically explored the possible vertical representation of weight. This was done within the theoretical framework outlined by TEST (Myachykov et al., 2014), which allows the impact of weight on both the environment (i.e. the tropic representation) and our bodily actions (i.e. the embodied representation) to be considered. Indeed, evidence for a vertical representation of weight emerges clearly from our everyday life experience. For instance, while relatively heavy objects (e.g. a car) are typically located on the ground (i.e. the bottom location of ‘our’ physical space), lighter objects (e.g. a leaf) are often associated with upper locations, such as when we see these objects flying in the sky (i.e. the top location of ‘our’ physical space). Similarly, when objects are piled up, the optimal equilibrium is usually attained when heavier objects are placed at the bottom and lighter objects are placed at the top, which also prevents lighter objects from being crushed by the heavier ones. Moreover, when objects are dropped to the ground from a certain height, heavier objects tend to fall faster than lighter ones, because of the effect of air resistance (see Oberle, McBeath, Madigan, & Sugar, 2005; Vicovaro, Noventa, & Battaglini, 2019). In the same vein, weight is definitely related to our perceptual-motor experience. Indeed, in a gravitational environment such as ours, the perceived heaviness of objects is a function of the downward force they exert on our body, which corresponds to weight (see Ross, 2018). As a consequence, heavier objects exert a greater downward force on our body than lighter objects, explaining why lighter objects are easier to lift than heavier ones.[1]

In summary, because weight is a genuine vertical downward force, and because this force can shape the relative position of objects in space and the way we interact with them, we predict that a possible tropic/embodied representation of weight would more likely result in a heavier-bottom/lighter-top association rather than the opposite. If confirmed, this pattern would be in sharp contrast with the typical smaller-bottom/greater-top association reported for both numerical (e.g. Müller & Schwarz, 2007) and non-numerical magnitudes (e.g. Bruzzi et al., 2017; Rusconi et al., 2006; Sell & Kaschak, 2012), and, more generally, with the predictions from ATOM (Walsh, 2003, 2015) and the polarity correspondence model (Proctor & Cho, 2006). Indeed, according to ATOM, the general mechanism for magnitude processing should treat weight and other magnitudes alike, thus leading to a classic lighter-bottom/heavier-top representation. In a similar vein, as for the polarity correspondence model, 'heavier' and 'lighter' should be coded as positive and negative polarities, respectively, thus also leading, in this case, to a heavier-top/lighter-bottom association. In line with these observations, the vertical representation of weight might constitute an important test bench for the theories of space–quantity associations.

In four experiments, participants completed a magnitude task for weight in response to a central target stimulus that could be a word describing a material (e.g. 'paper', 'iron': Experiment 1), a numerical quantity of weight (e.g. '1 g', '1 kg': Experiment 2) or a picture associated with a real object weighed by the participant before the experiment (i.e. weight as task-relevant dimension: Experiment 3a; weight as task-irrelevant dimension: Experiment 3b). As main results, we anticipate here that in Experiments 1 and 2 a lighter-bottom/heavier-top representation emerged—in line with both ATOM and the polarity correspondence model—whereas in Experiment 3a the reversed mapping emerged, in line with a tropic/embodied representation of weight.

Experiment 1: Weight comparison of words describing materials

This first experiment was inspired by and extended Dalmaso and Vicovaro (2019; Experiment 2), in which participants were presented with target words describing materials (e.g. plastic) lighter or heavier than a reference material, and responses were provided with two horizontally placed buttons. A similar approach was used here except that responses were provided with two vertically placed buttons. In line with Dalmaso and Vicovaro (2019), materials were mainly chosen as they minimize the size–weight correlation that typically characterizes objects (e.g. lighter/heavier objects are also generally smaller/bigger, respectively). For instance, plastic can be used to create both relatively small (e.g. a shirt button) and big (e.g. a table) objects. Furthermore, despite materials being typically associated with density rather than weight, there is evidence that stimuli referring to materials can be employed to manipulate weight expectations effectively (e.g. Buckingham, 2014; Buckingham, Ranger, & Goodale, 2011; Vicovaro & Burigana, 2017; for other manipulations of weight using linguistic materials see also, for instance, Schneider, Rutjens, Jostmann, and Lakens 2011 and Scorolli, Borghi, and Glenberg 2009).

Method

Participants

The sample size was estimated following the guidelines proposed by Brysbaert and Stevens (2018) for linear mixed-effects models, with subjects and items as random factors. According to these guidelines, at least 1600 observations per condition should be collected. This gives adequate statistical power for the small effect sizes that generally characterize reaction times (RT) studies. Here, because we planned to collect 70 trials per condition for each participant, at least 23 participants were required. The final sample was composed of 32 naïve students (mean age $M = 24$ years, $SD = 1.95$; seven males). All participants completed a written consent form. Four participants declared to be left-handed. Manual preference was further assessed through the Edinburgh Handedness Inventory (EHI; Oldfield, 1971), which provides a continuous handedness score on a scale ranging from − 100 (i.e. strong preference for the left hand) to 100 (i.e. strong preference for the right hand). Here, the mean EHI score was 58 (SD = 41.67; range from − 88 to 100).

Stimuli, apparatus and procedure

Stimuli were presented on a PC monitor (1600 × 1200 pixels; 75 Hz), placed 57 cm from the participant, through E-Prime

---

[1] Interestingly, it has been suggested that the correlation between mass and downward force (i.e. weight) underlies the well-known tendency—displayed by people without formal physics instruction—to overestimate the positive relationship between object mass and its falling speed (see Rohrer 2003; Vicovaro 2014; Vicovaro et al. 2019).

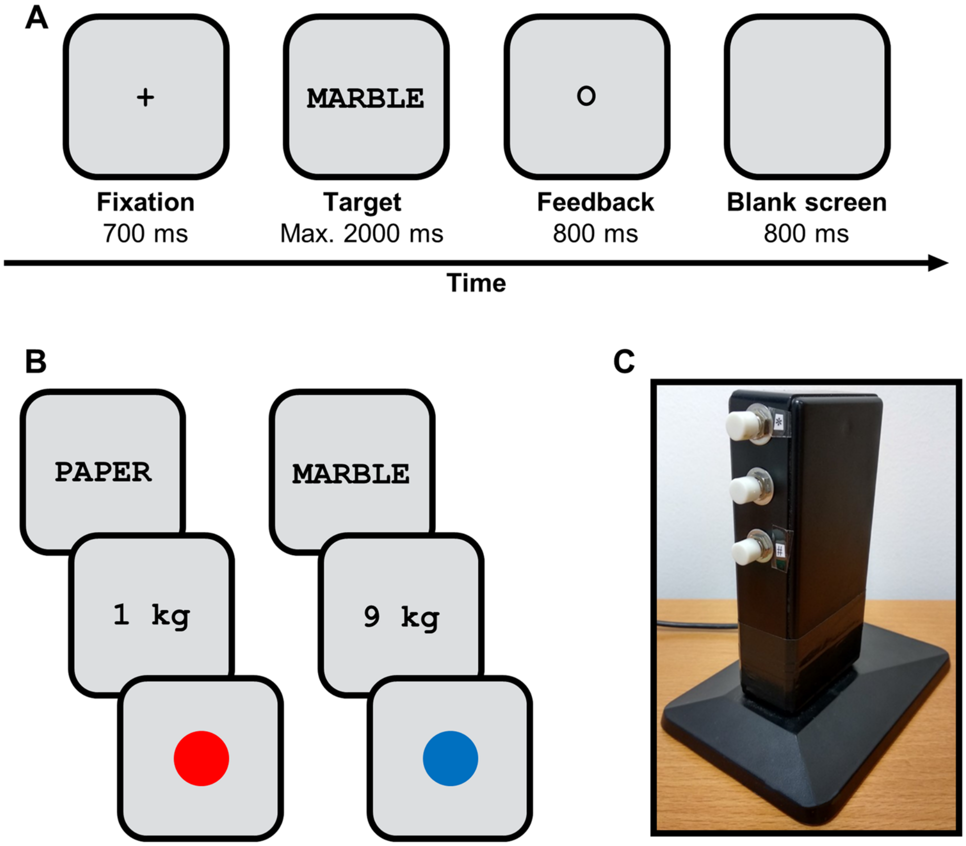

Fig. 1 **a** Example of a trial. The sequence of events was identical in all experiments. **b** Examples of target stimuli employed in Experiment 1 (paper vs. marble), Experiment 2 (1 kg vs. 9 kg) and Experiments 3a/3b (red circle vs. blue circle). **c** The custom-made response box employed in all experiments. The upper and lower keys were marked with ‘*’ and ‘#’ symbols, respectively, to avoid any mention of spatial positions when providing participants with task instructions. Please note that the central key was not used

2 (Psychology Software Tools, Pittsburgh, PA). The background was set to grey. For the comparison task, each trial started with a black central fixation cross (30-point Courier New font; see Fig. 1a). After 700 ms, the cross was replaced by a randomly selected black target word (30-point Courier New font) describing a material (for the whole stimulus set see Table 1).

Participants were asked to press a button as quickly and accurately as possible to decide whether the weight associated with the target word was either ‘lighter’ or ‘heavier’ than the reference word (i.e. wood). The reference word was only mentioned during the initial instructions. Manual responses were collected through a custom-made response box with keys arranged vertically.[2] To avoid any reference to spatial positions while providing participants with task instructions, the upper key was marked with the symbol ‘*’ and the lower key with the symbol ‘#’ (Fig. 1c). The response box was placed centrally with respect to the screen. After either a response or a 2000-ms timeout limit (whichever came first), a central black letter (30-point Courier New font) was provided as feedback for 800 ms (i.e. ‘O’ for correct responses; ‘X’ combined with an acoustic buzz for wrong responses; ‘TOO SLOW’ combined with an acoustic buzz for missed responses). Finally, a blank screen appeared for 800 ms. There were two practice blocks of 20 trials each (i.e. 40 practice trials in total) both followed by an experimental block of 140 trials (i.e. 280 experimental

Table 1 Word stimuli used in Experiment 1 (Italian words are in parentheses)

Reference word	Target words	
	Lighter weight	Heavier weight
Wood (*Legno*)	1. Cloth (*Stoffa*)	6. Cement (*Cemento*)
	2. Plastic (*Plastica*)	7. Iron (*Ferro*)
	3. Paper (*Carta*)	8. Lead (*Piombo*)
	4. Sponge (*Spugna*)	9. Marble (*Marmo*)
	5. Rubber (*Gomma*)	10. Steel (*Acciaio*)

Please note that the reference word was only mentioned during the initial instructions. Each target word was presented 14 times in each of the two experimental blocks. Lighter and heavier word stimuli did not differ neither for word length nor for log-transformed word frequency expressed as instances per million words (*ps* > .78; itWac corpus; Baroni, Bernardini, Ferraresi, & Zanchetta 2009)

---

[2] The use of a vertical response box is highly desirable in experiments investigating vertical SNARC and SNARC-like effects. Indeed, if participants’ responses are collected through a traditional keyboard placed on a horizontal plane and ‘vertically aligned’ keys are used (e.g. ‘Y’ and ‘B’; see for instance Ito and Hatta, 2004), then the response keys are actually aligned sagittally rather than vertically. In turn, as highlighted by Winter et al. (2015), this can lead to confounding between ‘bottom-top’ and ‘near-far’ dimensions.

trials in total). Each target word was presented randomly and for a fixed number of times, in both the practice (i.e. 2 times per block) and the experimental (i.e. 14 times per block) blocks. The association between response locations and 'lighter' and 'heavier' responses was inverted in the two blocks, namely in one block 'lighter' was associated with the top response and 'heavier' with the bottom response, whereas in the other block the opposite was true. Response keys were pressed with the thumbs, and the thumb–key association was inverted in the middle of each block to control for hand dominance. Block order and the initial thumb–key association were counterbalanced across participants.

As in Dalmaso and Vicovaro (2019), the comparison task was followed by a rating task, used for assessing the presence of a distance effect, if any. This task consisted of estimating the weight associated with each target word on a scale in which the value '50' corresponded to the hypothetical weight of 'wood'. Only integers (e.g. 16, 1, 84, etc.) were allowed, and participants were explicitly told that numbers greater than 100 could also be used. On each trial, a black target word (30-point Courier New font) was presented at the centre of the screen with no time limits. Participants reported the weight associated with the material by typing the value on the keyboard. Then, the response was confirmed by pressing the enter key. There was a practice block in which three randomly selected words were presented, followed by an experimental block in which each word was presented three times, to obtain more reliable ratings.

Finally, participants completed the EHI.

Results and discussion

Wrong responses (1.86% of trials) were deleted and analysed no further. Correct trials with RTs three standard deviations (3SD) above or below the participant's mean were classified as outliers and removed (1.83% of trials). In this manner, each condition was associated with at least 2139 observations, thus guaranteeing adequate statistical power (see Brysbaert & Stevens, 2018).

As in Dalmaso and Vicovaro (2019), we followed the guidelines proposed by Baayen, Davidson, and Bates (2008), analysing RTs of correct trials using a linear mixed-effects model (R package lme4; Bates, Maechler, Bolker, and Walker 2015). The statistical significance of random effects was tested using the *exact Likelihood Ratio Test* (LRT; Crainiceanu & Ruppert, 2004; R package RLRsim; Scheipl, Greven, & Kuechenhoff, 2008). More specifically, as fixed effects we entered the weight associated with the target word (lighter vs. heavier than the reference word 'wood'), the response location (top vs. bottom) and the interaction term.[3] As random effects, we had intercepts for subjects (LRT = 2808.8, $p < 0.001$) and items (i.e. target words; LRT = 19.39, $p < 0.001$), as well as by-subject random slopes for the effects of weight (LRT = 2320.4, $p < 0.001$) and response location (LRT = 2623.1, $p < 0.001$). This was the model that best fitted the data according to a likelihood ratio test comparing increasingly complex models (from the null to the saturated model; for a discussion of different approaches to the analysis of linear mixed-effects model see Barr, Levy, Scheepers, & Tily, 2013). This model was then entered into a Type 1 ANOVA (Satterthwaite's approximation for degrees of freedom) for linear mixed-effects models (R package lmerTest; Kuznetsova, Brockhoff, & Christensen, 2017). Evidence of a vertical SNARC-like effect for weight would be provided by a significant two-way interaction between weight and response location. Effect sizes were calculated according to the formulas provided by Westfall, Kenny, and Judd (2014; see also Brysbaert & Stevens, 2018) for linear mixed-effects models. The main effect of weight was significant [$F(1, 13.6) = 6.54$, $p = 0.023$, $d = 0.19$] because RTs were smaller for heavier ($M = 600$ ms, SE = 14.97) than lighter ($M = 632$ ms, SE = 17.09) weights, while the main effect of response location was non-significant [$F(1, 30.9) = 0.7$, $p > 0.10$, $d = 0.03$]. Importantly, the weight × response location interaction was significant [$F(1, 8523.2) = 8.33$, $p = 0.004$, $d = 0.10$], indicating the presence of a vertical SNARC-like effect. The interaction was explored further through planned Tukey's HSD comparisons for linear mixed-effects models (R package 'lsmeans'; Lenth, 2016). These showed no significant differences between top responses ($M = 625$ ms, SE = 17.78) and bottom responses ($M = 620$ ms, SE = 16.74) for 'lighter' target words [$t(64.73) = 0.91$, $p > 0.10$, $d = 0.16$], while 'heavier' target words led to shorter RTs for top ($M = 594$ ms, SE = 15.79) than for bottom ($M = 605$ ms, SE = 14.54) responses [$t(64.13) = 2.29$, $p = 0.025$, $d = 0.40$; see also Fig. 2, left panel]. These results appear to be consistent with the hypothesis that 'heavier' target words were represented up and 'lighter' target words were represented down. This is in line with both ATOM and the polarity correspondence model, and it suggests that words associated with different materials did not elicit a spatial representation of weight coherent with a tropic/embodied representation. Interestingly, a difference in RTs between the bottom and top responses emerged for 'heavier' but not for 'lighter'

---

[3] The SNARC effect is frequently tested by computing, for each number stimulus, the mean RT difference between the right- and the left-side key, and then by testing the existence of a negative correlation between number magnitudes and mean RT difference (see Fias, Brysbaert, Geypens, and d'Ydewalle 1996). Theoretically, this approach could also be used in the current context using the mean rated weight of target words instead of number magnitude. However, when magnitude is task relevant, as in our study, the mean RT difference is not a linear but a categorical function of magnitude, which implies the violation of one basic assumption of linear regression analysis (see Gevers et al. 2006).

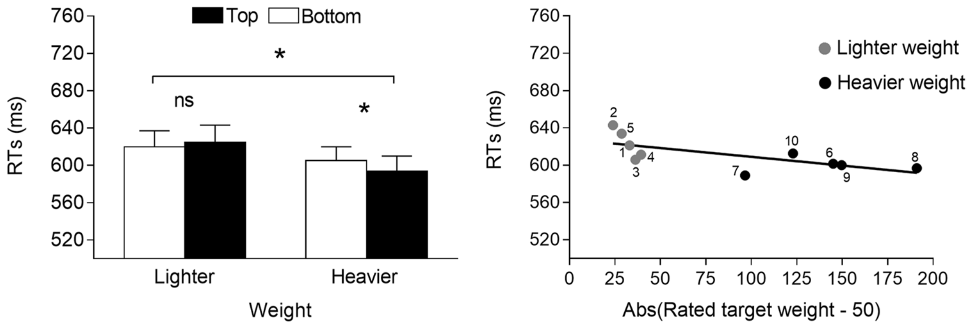

Fig. 2 Left panel: Mean RTs observed in the comparison task of Experiment 1 for bottom and top responses as a function of weight. *$p<.05$; *ns* non-significant. Error bars are standard errors of the mean. Right panel: Mean RTs for each target word as a function of the mean absolute difference between its rated weight and the weight of the reference word (i.e. 'wood'). Black and grey circles are for 'heavier' and 'lighter' target words, respectively. Numbers refer to the word list reported in Table 1

target words. This finding aligns with some previous studies that reported a similar, unbalanced pattern of results when participants were asked to discriminate smaller vs. greater magnitudes (Chang & Cho, 2015; Dalmaso & Vicovaro, 2019; Di Rosa et al., 2017). For instance, Chang & Cho, (2015) reported a SNARC-like effect only in response to higher—but not lower—levels of loudness. According to Chang and Cho (2015), this could reflect a bias in processing positive—rather than negative—polarities more efficiently, which would be particularly true for physical dimensions (e.g. Lakens, 2012). We tentatively suggest that a similar rationale could also be applied to the present context.[4] Additionally, we note that, although ATOM is consistent with a lighter-bottom/heavier-top association, it does not provide a clear explanation for the unbalanced pattern of results for 'heavier' vs. 'lighter' stimuli. Therefore, in this specific case, an explanation of the results in terms of polarity correspondence might be more suitable than an explanation in terms of ATOM, although this conclusion should be treated with caution.

Finally, we analysed the possible presence of a distance effect for weight. In the context of numerical cognition, the distance effect refers to the fact that when participants are asked to decide which number in a pair is the largest, RTs tend to decrease with increased absolute difference between the numbers (Moyer & Landauer, 1967). In a similar vein, here we tested the hypothesis that when participants are asked to decide if a target word is heavier or lighter than the reference word, RTs tend to decrease with increased absolute difference between the rated weight of the target and the weight of the reference (i.e. 50). This negative relationship was indeed confirmed by linear mixed-effects regression analysis, with fixed effects for the intercept and slope of the regression line relating the mean RTs to the untransformed mean of rating differences, and random (by-subject) effects for the intercept of the same regression line: $b=-0.054$, $SE_b=0.015$, $t(293.4)=-3.58$, $p<0.001$ (see Fig. 2, right panel). As suggested by Toomarian and Hubbard (2018), the distance effect can be interpreted as a sign of magnitude processing, therefore, the existence of a distance effect for weight suggests that participants processed the weight associated with each target word. This further confirms that words describing material can be used effectively to manipulate represented weight (see also Dalmaso & Vicovaro, 2019). Interestingly, 'lighter' target words were judged to be closer in weight to the reference word compared to 'heavier' words (see also Fig. 2, right panel). According to the distance effect, RTs tend to decrease with increased difference in weight between the target and the reference, so this probably explains why RTs were smaller for heavier than for lighter target words (i.e. the main effect of weight in the comparison task).

Experiment 2: Weight comparison of numerical quantities of weight

The main results from Experiment 1 revealed that words describing materials did not lead to a heavier-bottom/lighter-top spatial representation of weight, but rather to the opposite pattern of results (i.e. lighter-bottom/heavier-top). This aligns with both ATOM and the polarity correspondence model. In this second experiment, the target stimuli were numerical quantities of weight (e.g. 3 kg).

---

[4] According to Lakens (2012), it can be established which of two opposite categories (e.g. light vs. heavy) is positively polarized by considering which category gives the name to its dimension. Here, the positively polarized category is 'heavy' as it gives one of the names with which to refer to the 'weight' dimension (i.e. heaviness). This is also reflected in in common language, since there is a natural preference is using 'heavy' rather than 'light' when referring to heaviness/weight dimension, such as when we want to know the weight of another individual; in this case, we typically ask 'how heavy are you?' and not 'how light are you?'.

We reasoned that these stimuli, expressing weight magnitude in a more concrete and unequivocal way compared to materials, could facilitate the activation of a tropic/embodied representation of weight.

Method

Participants

As we planned to collect 56 trials per condition for each participant, at least 29 participants were required (see Brysbaert & Stevens, 2018; see also Experiment 1). In line with Experiment 1, the final sample was composed of 32 naïve students (mean age $M = 23$ years, SD = 2.14, 10 males). All participants completed a written consent form. One participant declared to be left-handed. The mean EHI score was 66 (SD = 27.18; range from − 42 to 100).

Stimuli, apparatus and procedure

Everything was identical to Experiment 1 but with the following exceptions: the reference stimulus was '5 kg'; 1, 2, 3, 4 kg and 6, 7, 8, 9 g were the lighter stimuli, whereas 1, 2, 3, 4 t (i.e. tons) and 6, 7, 8, 9 kg were the heavier stimuli (see also Fig. 1b); there were two practice blocks of 16 trials each (i.e. 32 practice trials in total; each target stimulus appeared one time per block) both followed by an experimental block of 224 trials (i.e. 448 experimental trials in total; each target stimulus appeared 14 times per block); the rating task was not included. Please note that to perform the task accurately, participants had to pay attention to both the digit and the weight unit. Indeed, digits smaller than 5 could be associated with weights lighter or heavier than 5 kg (e.g. 3 g or 3 t), and the same was true for digits larger than 5 (e.g. 7 g or 7 t). Moreover, 'kg' could be associated with weights lighter or heavier than 5 kg (e.g. 3 kg or 7 kg). Therefore, the response behaviour could not be based on either digit or weight unit processing alone but had to be based on the combination of both dimensions (for a similar approach with temporal stimuli see Zhao et al., 2018).

Experiment 2 was also aimed to test whether the spatial representation of the stimuli employed here was actually shaped by weight magnitude itself or simply reflected the automatic processing of number magnitude, in line with a classic vertical SNARC effect (e.g. Müller & Schwarz, 2007). Indeed, for trials in which target stimuli referred to kilograms (hereafter, *kg trials*), both digits (1–9) and weight magnitudes (1–9 kg) correlated positively. The opposite was true for trials in which target stimuli referred to both grams and tons (hereafter, *g-t trials*), that is, larger/smaller digits were linked to lighter/heavier weights, respectively (e.g. 9 g vs. 4 t). If the spatial representation of the stimuli reflected the automatic processing of number magnitude, then a lighter-bottom/heavier-top association should emerge in *kg trials*, whereas the opposite association should emerge in *g-t trials*. Instead, if the spatial representation of the stimuli was shaped by weight magnitude, then the space-weight association should be the same in both types of trials. We anticipate here that the results provided support to the latter hypothesis.

Results and discussion

For the aforementioned reasons, *kg trials* and *g-t trials* were analysed separately.

Analyses of kg trials

Wrong responses (5.08% of trials) and RTs 3SD above or below the participant's mean (2.5% of trials) were deleted and not analysed further. In this manner, each condition was associated with at least 1588 observations, thus guaranteeing reasonable statistical power (see Brysbaert & Stevens, 2018).

The model that best fitted the RT data included, as fixed effects, the relative weight implied by the target stimulus (lighter vs. heavier than the reference '5 kg'), the response location (top vs. bottom) and the interaction term. As random effects, it included intercepts for subjects (LRT = 1725.7, $p < 0.001$) and items (i.e. target stimuli; LRT = 62.23, $p < 0.001$), and by-subject random slopes for the effects of weight (LRT = 1592.1, $p < 0.001$) and response location (LRT = 1671.6, $p < 0.001$). The main effect of weight was non-significant [$F(1, 8.0) = 1.36$, $p > 0.10$, $d = 0.25$] but the main effect of response location was significant [$F(1, 29.8) = 5.74$, $p = 0.023$, $d = 0.07$] due to smaller RTs for top responses ($M = 702$ ms, SE = 17.3) than bottom responses ($M = 713$ ms, SE = 18.04). Importantly, the weight × response location interaction was significant [$F(1, 6542.6) = 41.78$, $p < 0.001$, $d = 0.27$], indicating the presence of a vertical SNARC-like effect. In line with Experiment 1 (see also Chang & Cho, 2015; Lakens, 2012), for 'lighter' stimuli the difference between top responses ($M = 724$ ms, SE = 18.95) and bottom responses ($M = 712$ ms, SE = 19.56) was not statistically significant [$t(68.37) = 1.83$, $p = 0.072$, $d = 0.32$], whereas 'heavier' stimuli led to significantly shorter RTs for top responses ($M = 680$ ms, SE = 20.04) than for bottom responses ($M = 715$ ms, SE = 20.84): $t(64.32) = 5.63$, $p < 0.001$, $d = 0.99$ (see Fig. 3, left panel).

These results are consistent with those observed in Experiment 1. Indeed, also in this case, evidence of a lighter-bottom/heavier-top association emerged and the difference between bottom vs. top responses was limited to 'heavier' items. Overall, these results are consistent with both ATOM and the polarity correspondence model. However, as suggested in Experiment 1, the unbalanced pattern of results for 'heavier' and 'lighter' items seems to provide tentative

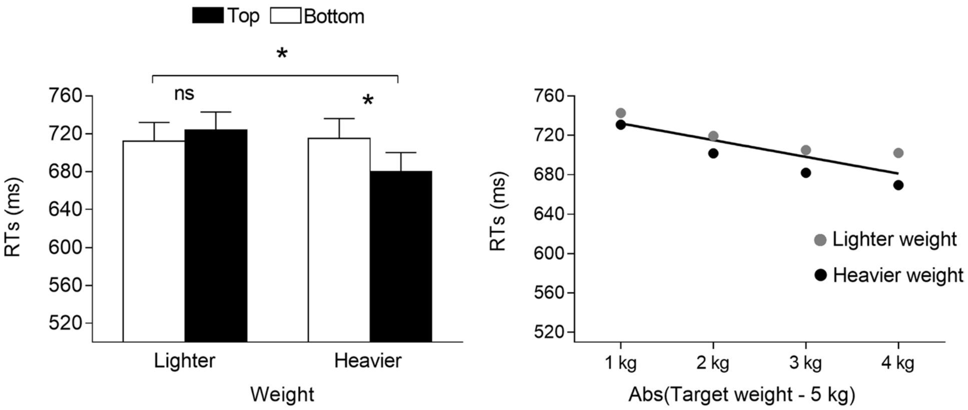

Fig. 3 Left panel: Mean RTs observed in the comparison task of Experiment 2 (*kg trials* only) for bottom and top responses as a function of weight. $*p<.05$; *ns* non-significant. Error bars are standard errors of the mean. Right panel: Mean RTs for each numerical target weight (*kg trials* only) as a function of the mean absolute difference between its magnitude and the magnitude of the reference weight (i.e. 5 kg). Black and grey circles are for 'heavier' and 'lighter' target weights, respectively

support for the hypothesis of a benefit in processing positive polarities (see footnote 4; see also Chang & Cho, 2015; Lakens, 2012).

As for the distance effect, we employed a linear mixed-effects regression analysis with fixed effects for the intercept and slope of the regression line relating RTs to absolute differences between the weight of the target and that of the reference, and random (by-subject) effects for the intercept and slope of the same regression line. A clear negative relationship emerged between RTs and weight difference [$b=-17.25$, $SE_b=2.35$, $t(31.28)=-7.35$, $p<0.001$; see Fig. 3, right panel], indicating that participants processed weight magnitudes effectively (see Toomarian & Hubbard, 2018).

Analyses of g-t trials

Wrong responses (2.29% of trials) and RTs 3SD above or below the participant's mean (1.28% of trials) were deleted and not analysed further. In this manner, each condition was associated with at least 1703 observations, thus guaranteeing adequate statistical power (see Brysbaert & Stevens, 2018).

The model that best fitted the RT data included, as fixed effects, the relative weight implied by the target stimulus (lighter vs. heavier than the reference '5 kg'), the response location (top vs. bottom) and the interaction term. As random effects, it included intercepts for subjects (LRT = 1533.8, $p<0.001$) and items (i.e. target stimulus; LRT = 1.81, $p=0.024$), and by-subject random slopes for the effects of weight (LRT = 1389.1, $p<0.001$) and response location (LRT = 1492.2, $p<0.001$). The main effect of weight was non-significant [$F(1, 11.5)=0.35$, $p>0.10$, $d=0.18$], whereas the main effect of response location was significant [$F(1, 30.8)=6.13$, $p=0.019$, $d=0.14$] due to smaller RTs for top responses ($M=638$ ms, SE = 11.68) than bottom responses ($M=648$ ms, SE = 16.65). Importantly, the weight × response location interaction was significant [$F(1, 6810.9)=93.56$, $p<0.001$, $d=0.41$], indicating the presence of a vertical SNARC-like effect. For 'lighter' stimuli, the RTs were significantly shorter for bottom responses ($M=632$ ms, SE = 12.61) than for top responses ($M=651$ ms, SE = 11.94): $t(80.01)=4.01$, $p<0.001$, $d=0.71$. Instead, for 'heavier' stimuli the RTs were significantly shorter for top responses ($M=626$ ms, SE = 12.42) than for bottom responses ($M=664$ ms, SE = 13.59): $t(80.05)=7.91$, $p<0.001$, $d=1.4$ (see also Fig. 4, left panel). Moreover, it is important to note that the difference between bottom and top responses was larger for 'heavier' than for 'lighter' stimuli (i.e. 38 vs. 19 ms, respectively), which is still consistent with the hypothesis of a benefit in processing positive polarities (see Chang & Cho, 2015; Lakens, 2012).

In summary, these results replicated the weight–space representation already observed both in Experiment 1 and in the *kg trials*, because a lighter-bottom/heavier-top association emerged also in this case. Importantly, the likeness between the *kg* and *g-t trials* indicates that participants' responses were actually based on the weight implied by the stimuli rather than on number magnitude per se. Indeed, if participants' responses were solely based on number magnitude, then a smaller-bottom/greater-top digit–space association should emerge, in line with a vertical SNARC effect. However, this was clearly not the case.

Finally, the distance effect was tested independently for *grams* and *tons*, given their unsymmetrical distance from the reference. In both cases, we employed linear mixed-effects regression analysis with fixed effects for the intercept and slope of the regression line relating RTs to the weight of the target, and random (by-subject) effects for the intercept of the same regression line. Note that, in line with the distance effect, RTs should decrease with the number of tons and

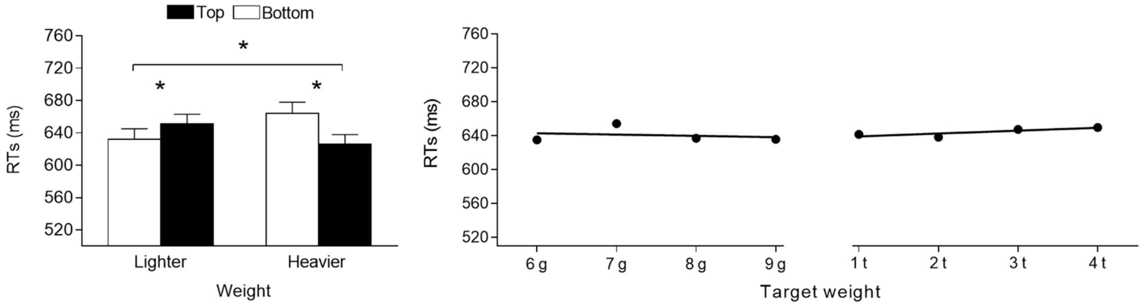

Fig. 4 Left panel: Mean RTs observed in the comparison task of Experiment 2 (g-t trials only) for bottom and top responses as a function of weight. *$p < .05$. Error bars are standard errors of the mean. Right panels: Mean RTs for each numerical target weight (g-t trials only) as a function of their magnitude

increase with the number of grams (i.e. the larger the number of grams, then the closer the target stimulus to the reference stimulus and the larger the RTs). No statistically significant relationship emerged between RTs and weight, neither for *tons* [$b = 3.37$, $SE_b = 1.9$, $t(3419) = 1.77$, $p = 0.077$] nor *grams* [$b = -1.43$, $SE_b = 1.9$, $t(3422.8) = -0.75$, $p > 0.10$; see Fig. 4, right panels]. The lack of a distance effect in the *g–t trials* can perhaps be attributed to a sort of range-restriction phenomenon: the subjective distance between the targets and the reference was probably too large to allow the emergence of distance differences between the targets.

Experiment 3a: Weight comparison of real objects differing in both weight and size (weight as task-relevant dimension)

In Experiments 1 and 2, evidence for a lighter-bottom/heavier-top association emerged. This is coherent both with ATOM and the polarity correspondence model, rather than with the tropic/embodied representation of weight. Nevertheless, we reasoned that in the previous two experiments weight was manipulated at a 'conceptual' level—namely, based on abstract representations of magnitudes evoked by words (Experiment 1) or numerical quantities (Experiment 2)—rather than on a 'concrete' physical exposure to real weights. For this reason, in Experiments 3a and 3b weight was manipulated through real objects that were weighed by participants before the comparison task. In so doing, we aimed to enforce the notion that physically heavier objects exert a stronger downward force with respect to lighter objects. This, in turn, was expected to activate a heavier-bottom/lighter-top association, in line with a tropic/embodied representation of weight. Moreover, we took the advantage of employing real objects to explore also whether the expected SNARC-like effect was still detectable even when weight was a task-irrelevant dimension, following the approach that is typically adopted in classic SNARC tasks based on parity rather than on magnitude discrimination (e.g. Deahene et al. 1993). For this reason, the objects employed in the following two experiments were orthogonally manipulated for both weight and size, and participants classified these stimuli as either 'light' vs. 'heavy' (i.e. weight was task-relevant and size was task-irrelevant; Experiment 3a) or 'small' vs. 'big' (i.e. size was task-relevant and weight was task-irrelevant; Experiment 3b).

Method

Participants

As we planned to collect 56 trials per condition for each participant, at least 29 participants were required (see Brysbaert & Stevens, 2018; see also Experiment 1). In line with previous experiments, the final sample was composed of 32 naïve students (mean age $M = 20$ years, $SD = 1.5$, nine males). All participants completed a written consent form. Four participants declared to be left-handed. The mean EHI score was 51 ($SD = 45.74$; range from -89 to 100).

Stimuli, apparatus and procedure

Conceptually, Experiment 3a was similar to Experiments 1 and 2 but weight was manipulated through real items and four different tasks comprised the whole experimental procedure. In more detail, the stimuli were four hollow polystyrene spheres: two with a diameter of 15 cm ('small' spheres) and two with a diameter of 30 cm ('big' spheres; see Fig. 5b). Each sphere was filled with play dough so that the weights of one small sphere (695 g) and one big sphere (1100 g) were larger than the weights of the other two spheres (97 g for the small sphere, 154 g for the large sphere). Please note that the weight of each sphere was regulated so that the two 'light' spheres were perceived to be similar in weight, irrespective of their size; this was also the case for the two 'heavy'

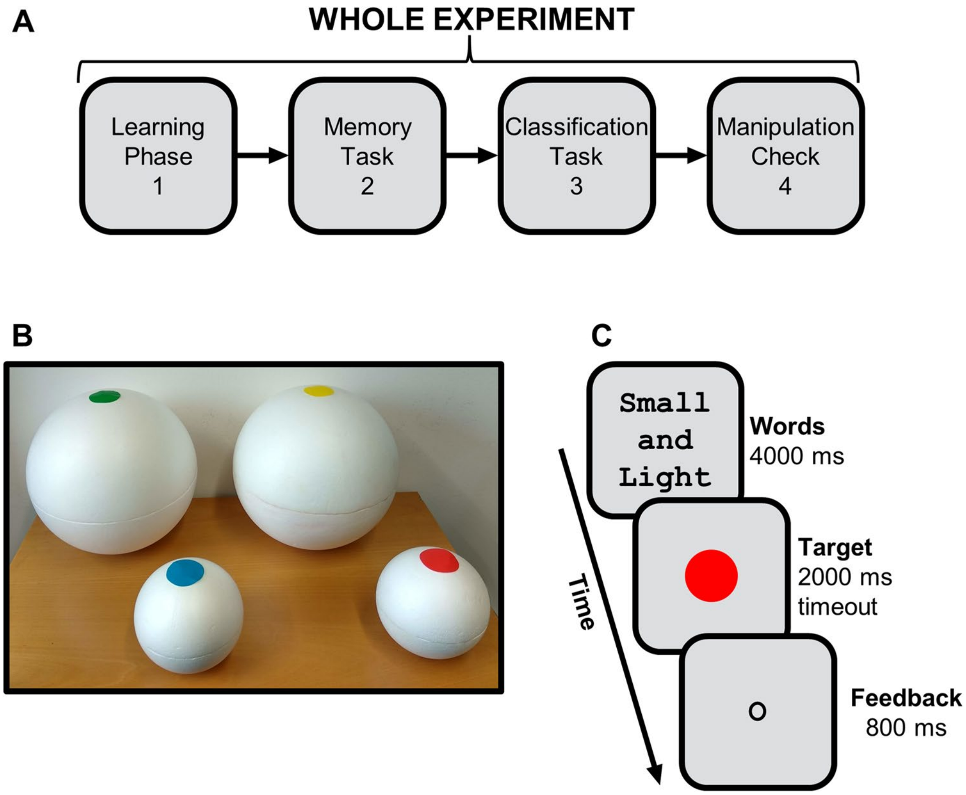

Fig. 5 **a** The experimental design employed in Experiments 3a/3b. After the learning phase (1), a memory task (2) was administered. This was followed by a classification task (3) and a manipulation check (4). Please note that a similar design has been successfully employed previously but with social stimuli (see Carraro et al. 2016; Dalmaso, Galfano, Coricelli, & Castelli 2014). **b** Photograph of the four spheres employed during the learning phase. **c** Example of a trial employed during the memory task. The same trial structure was also employed during the manipulation check, but no feedback was provided

spheres.[5] Then, four 5-cm diameter circular stickers of different colours (red, green, blue, yellow) were stuck on the top of the four spheres (Fig. 5b). The spheres were placed on a table placed inside the experimental room and a physical barrier hid them from view when the participant was sat in front of the PC monitor. The colour–sphere association and the relative position of the spheres on the table were randomized for each participant and remained unaltered for the whole experimental session.

As for the experimental procedure, the first task was a learning phase in which participants were asked to stay in front of the table on which the four spheres were placed and to lift each of them with both hands, memorizing the associations between weight (light or heavy), size (small or big) and colour (e.g. 'the red sphere is light and small'). The spheres were lifted in a left-to-right order. No time limits were imposed. After the learning task, to further consolidate learning the participants sat in front of the PC monitor to complete a memory task (see Fig. 5c). This consisted of two centrally-placed adjectives presented for 4000 ms, referring to the properties of a given sphere (e.g. 'light and small'; 30-point Courier New font). Then, a central circle (diameter 4 cm; 4° visual angle) coloured red, green, blue or yellow was presented for 2000 ms. Participants were

---

[5] A well-known phenomenon underlying weight perception is the so-called 'size–weight illusion': When two objects of identical physical weight but different size are lifted, the smaller object is typically perceived to be heavier than the larger object (e.g. Buckingham, 2014; Vicovaro and Burigana, 2014). This explains why, in the present context, the physical weight of the two bigger spheres (light and heavy) exceeded that of the corresponding smaller spheres. Moreover, the weight difference between the small and the big sphere was larger for the heavy spheres (695 g and 1100 g, respectively) than for the light spheres (97 g and 154 g, respectively). This was done on purpose to comply with Weber's law, according to which differences in perceived weight are related to weight ratios rather than to weight differences. Since a weight ratio of about 1.58 (i.e. 154 g/97 g) nullified the size-weight illusion in the case of light spheres, the same weight ratio was also maintained for heavy spheres (i.e. 1100 g/695 g ≈ 1.58), to nullify the size-weight illusion likewise. For completeness, the perceived weight of the four spheres was also pre-tested by a sample of 24 individuals (mean age $M = 23$ years, $SD = 2.5$; five males, one left-handed). In more detail, participants were asked to lift each sphere with two hands, and to estimate its weight by providing an integer numerical value. The integer value was recorded manually by the experimenter. Sphere order was counterbalanced across participants. A 2 (weight: light vs. heavy) × 2 (size: small vs. big) repeated-measures ANOVA was conducted on standardized estimates. The main effect of weight was significant [$F(1, 23) = 7213.5$, $p < .001$, $\eta^2_g = .899$], confirming that light and heavy spheres were perceived as different, whereas the main effect of size was non-significant ($F < 1$). The interaction between the two factors was significant [$F(1, 23) = 5.85$, $p = .024$, $\eta^2_g = .107$]. Nevertheless, two-tailed paired

Footnote 5 (continued)

t-tests revealed that the two light and the two heavy spheres were perceived to be similar in weight independently of their size ($ts < 1.88$, $ps > .072$).

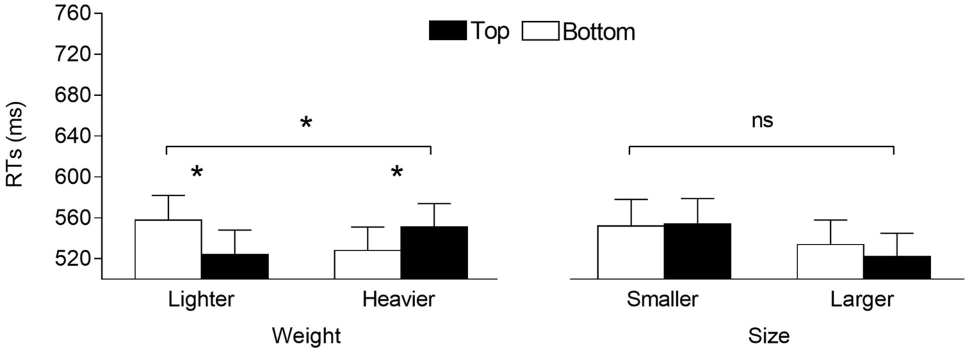

Fig. 6 Mean RTs observed in the classification task of Experiment 3a for bottom and top responses as a function of weight (i.e. the task-relevant dimension; left panel) and size (i.e., the task-irrelevant dimension; right panel). *$p < .05$; *ns* non-significant. Error bars are standard errors of the mean

asked to press a button on a standard keyboard (‘d’ and ‘k’ keys; counterbalanced across participants) to decide whether the association between the two adjectives (e.g. ‘light and small’) and the colour of the circle (e.g. red) matched one of the memorized spheres. The keyboard was placed centrally with respect to the screen. Eight trials were provided: four correct associations plus four randomly chosen wrong associations. Feedback was provided for 800 ms (i.e. a central ‘O’ for correct responses; a central ‘X’ combined with an acoustic buzz for wrong responses; the words ‘TOO SLOW’ combined with an acoustic buzz for missed responses). If at least one wrong or one missing response was detected, then the learning phase was administered again, followed by the memory task. When the memory task was successfully completed, participants lifted the spheres again and the classification task was started. This was identical to the comparison task employed in Experiments 1 and 2, with the following exceptions: participants were presented with a randomly-selected centrally-placed circle and they were asked to decide whether it was associated with a ‘light’ or a ‘heavy’ sphere (for a similar task see also Vallesi et al. 2008); there were two practice blocks of 12 trials each (i.e. 24 practice trials in total; each target stimulus appeared 3 times per block) both followed by an experimental block of 112 trials (i.e. 224 experimental trials in total; each target stimulus appeared 28 times per block); every 56 trials, participants were asked to lift the spheres. Finally, the classification task was followed by a manipulation check with the aim to assess whether the information associated with each sphere was correctly retained in memory for the whole duration of the classification task. The manipulation check was similar to the memory task but with the following exceptions: a single cycle was presented; no feedback was provided; there was no time limit for responding to maximize accuracy.

Results and discussion

Data were analysed as in previous experiments.

The mean number of cycles needed to successfully complete the learning phase was three (SD = 1.26; range 1–6). For the manipulation check, the mean accuracy was particularly high ($M = 95\%$, SD = 11.82), confirming that participants were accurate overall in remembering both the weight and size of each sphere. With the classification task, wrong responses (3.11% of trials) and RT outliers (1.97% of trials) were eliminated and not analysed further. In this manner, each condition was associated with at least 1687 observations, thus guaranteeing adequate statistical power (see Brysbaert & Stevens, 2018).

Analysis of the task-relevant dimension (i.e. weight) showed that the model that best fitted RT data included, as fixed effects, the relative weight implied by the target stimulus (light vs. heavy), the response location (top vs. bottom) and the interaction term. As random effects, it included intercepts for subjects (LRT = 2274.6, $p < 0.001$) and for the task-irrelevant dimension (i.e. size; LRT = 21.69, $p < 0.001$), and by-subject random slopes for the effects of weight (LRT = 1933.7, $p < 0.001$). The main effects of weight [$F(1, 31) = 0.04$, $p > 0.10$, $d = 0.13$] and response location [$F(1, 6759.9) = 1.64$, $p > 0.10$, $d = 0.17$] were both non-significant. Importantly, the weight × response location interaction was significant [$F(1, 6758.9) = 52.65$, $p < 0.001$, $d = 0.28$], indicating the presence of a vertical SNARC-like effect (see Fig. 6, left panel). The pattern of results was in sharp contrast to what was observed in Experiments 1 and 2. Indeed, as for lighter stimuli, top responses ($M = 524$ ms, SE = 23.98) were faster than bottom responses ($M = 558$ ms, SE = 23.98): $t(6759) = 6.04$, $p < 0.001$, $d = 1.07$. As for heavier stimuli, the opposite pattern emerged, with bottom responses ($M = 528$ ms, SE = 22.68) faster than top responses ($M = 551$ ms, SE = 22.68): $t(6759.8) = 4.22$, $p < 0.001$, $d = 0.75$. To the best of our knowledge, these results provide the first evidence of a heavier-bottom/lighter-top association between weight and space, which is consistent with the hypothesis of a tropic/embodied vertical representation of weight and inconsistent with both ATOM and the polarity correspondence model. Interestingly, the unbalanced pattern of results for ‘heavier’ and ‘lighter’ stimuli reported in Experiments 1–2 did not emerge in Experiment 3a, likely reflecting the fact that, in this third experiment, weight was processed differently as compared to the previous two experiments.

The same analyses as described above were also performed on the task-irrelevant dimension (i.e. size). The model that best fitted the RT data included, as fixed effects, the relative size implied by the target stimulus (small vs. big), the response location (top vs. bottom) and the interaction term. As random effects, it included intercepts for subjects (LRT = 2269.8, $p < 0.001$) and for the task-relevant dimension (i.e. weight; LRT = 24.71, $p < 0.001$), and by-subject random slopes for the effects of size (LRT = 2269.8, $p < 0.001$) and response location (LRT = 2162, $p < 0.001$). The main effect of size was significant [$F(1, 31.1) = 13.12$, $p = 0.001$, $d = 0.16$] because participants were faster in responding to bigger stimuli ($M = 528$ ms, SE = 18.30) compared to smaller stimuli ($M = 553$ ms, SE = 20.7). The main effect of response location [$F(1, 31) = 1.07$, $p > 0.10$, $d = 0.06$] and the two-way interaction [$F(1, 6728.2) = 2.93$, $p = 0.086$, $d = 0.07$] were not significant (see Fig. 6, right panel). In summary, no evidence of a vertical representation for the task-irrelevant dimension (i.e. size) emerged.

Experiment 3b: Size comparison of real objects differing in both weight and size (weight as task-irrelevant dimension)

In Experiment 3a, we observed a heavier-bottom/lighter-top association between weight (the task-relevant dimension) and space. On the other hand, no evidence of a spatial representation of size (the task-irrelevant dimension) emerged. In Experiment 3b, we wanted to explore whether the SNARC-like effect observed in Experiment 3a emerged even when weight was the task-irrelevant dimension, and participants were asked to provide a response based on the task-relevant dimension of size.

Method

Participants

In line with Experiment 3a, the sample was composed of 32 naïve students (mean age $M = 22$ years, SD = 2.6, nine males). All participants completed a written consent form. Six participants declared to be left-handed. The mean EHI score was 45 (SD = 44.81; range from -58 to 100).

Stimuli, apparatus and procedure

Everything was identical to Experiment 3a, with the following exception: in the classification task, participants were asked to decide whether the target colour was associated with a small or a big sphere. In so doing, size became the task-relevant dimension and weight was the task-irrelevant dimension.

Results and discussion

Data were analysed as in previous experiments.

The mean number of cycles needed to successfully complete the learning phase was three (SD = 3.31; range 1–13). For the manipulation check the mean accuracy was high ($M = 94\%$, SD = 10.52), confirming that participants were accurate overall in remembering both the weight and size of each sphere. With the classification task, wrong responses (2.69% of trials) and RT outliers (1.92% of trials) were eliminated and not analysed further. In this manner, each condition was associated with at least 1702 observations, thus guaranteeing adequate statistical power (see Brysbaert & Stevens, 2018).

Analysis for the task-relevant dimension (i.e. size) showed that the model that best fitted the RT data included, as fixed effects, the relative size implied by the target stimulus (small vs. big), the response location (top vs. bottom) and the interaction term. As random effects, it included intercepts for subjects (LRT = 1788, $p < 0.001$) and for the task-irrelevant dimension (i.e. weight; LRT = 0.60, $p = 0.049$), and by-subject random slopes for the effects of size (LRT = 1638.1, $p < 0.001$) and response location (LRT = 1697, $p < 0.001$). The main effect of response location was significant [$F(1, 31.1) = 5.69$, $p = 0.023$, $d = 0.08$] because top responses were faster ($M = 535$ ms, SE = 16.01) than bottom responses ($M = 547$ ms, SE = 15.70). The main effects of size and the size × response location interaction were non-significant: $F(1, 31.1) = 0.19$, $p > 0.10$, $d = 0.02$ and $F(1, 6754.2) = 0.35$, $p > 0.10$, $d = 0.02$, respectively; see Fig. 7, right panel).

Overall, the results emerging from the present context seem to suggest that size might not be represented along the vertical dimension because no evidence of a size–space association emerged in Experiments 3a/3b. This constitutes a valuable addition to the existing literature on SNARC-like effects because previous studies only focused on the representation of size along the horizontal dimension, reporting evidence of left-to-right mapping (see Prpic et al., in press; Ren et al., 2011; Sellaro et al., 2015). That said, future studies are needed to explore whether a vertical spatial representation of size could emerge by employing different stimuli and/or procedures from those employed here.

The same analyses as described above were also performed on the task-irrelevant dimension (i.e. weight). In this case, the model that best fitted the RT data included, as fixed effects, the relative weight implied by the target stimulus (light vs. heavy), the response location (top vs. bottom) and the interaction term. As random effects, it

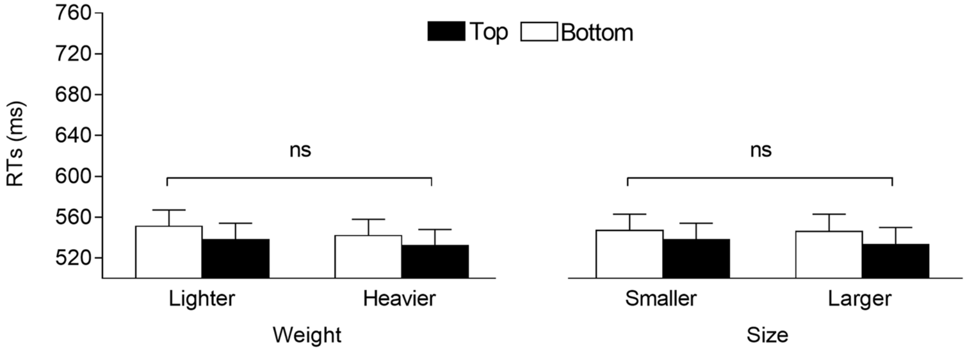

Fig. 7 Mean RTs observed in the classification task of Experiment 3b for bottom and top responses as a function of weight (i.e. the task-irrelevant dimension; left panel) and size (i.e. the task-relevant dimension; right panel). ns = non-significant. Error bars are standard errors of the mean

included intercepts for subjects (LRT = 1788.6, $p < 0.001$) and for the task-relevant dimension (i.e. size; LRT = 17.4, $p < 0.001$), and by-subject random slopes for the effects of response location (LRT = 1697.6, $p < 0.001$). The main effect of weight was significant [$F(1, 6777.1) = 4.66$, $p = 0.031$, $d = 0.04$] due to the lower RTs for heavier stimuli ($M = 537$ ms, SE = 15.40) than lighter stimuli ($M = 545$ ms, SE = 15.40). The main effect of response location was also significant [$F(1, 31.1) = 5.73$, $p = 0.023$, $d = 0.07$] due to the lower RTs for top responses ($M = 535$ ms, SE = 15.65) than bottom responses ($M = 547$ ms, SE = 15.33). However, the weight × response location interaction was non-significant [$F(1, 6777.8) = 0.07$, $p > 0.10$, $d = 0.01$] (see Fig. 7, left panel).[6]

Taken together, the results of Experiments 3a and 3b suggest that a heavier-bottom/lighter-top weight–space representation can actually emerge, but only when weight is a task-relevant dimension. In other words, the vertical representation of weight—and in particular the embodied vertical representation of weight—appears to be linked to the explicit processing of weight magnitude. Overall, this is consistent with the notion that SNARC-like effects are more evident when magnitude is task-relevant than when it is task-irrelevant (for a meta-analysis and review see Macnamara et al., 2018).

General discussion

The bulk of the evidence reported a link between numbers and space—the smaller numbers typically associated with the left-bottom locations of space and larger numbers typically associated with the right-top locations of space (i.e. the SNARC effect; e.g. Dehaene et al., 1993; Müller & Schwarz, 2007). Interestingly, a similar link between magnitude and space also has been reported for non-numerical magnitudes (i.e. SNARC-like effects; e.g. Macnamara et al., 2018). However, the possible vertical representation of non-numerical magnitudes is still largely unknown and mainly confined to auditory stimuli (e.g. Bruzzi et al., 2017). In the present study, we conducted four experiments to explore the possible vertical representation of weight, a crucial variable that heavily shapes our interactions with the physical world.

From an experimental perspective, exploring the possible vertical representation of weight is of great interest for two main reasons. Firstly, to date, only two studies (Dalmaso & Vicovaro, 2019; Holmes & Lourenco, 2013) have explored the possible spatial representations of weight, focusing only on the horizontal dimension, and both studies provided support for a left-to-right spatial representation of weight, in line with the literature on SNARC-like effects. Secondly – and more importantly—the spatial representation of weight along the vertical dimension, as well as being a novel addition to the SNARC-like literature, might represent an important test-bench for the main theories of space–quantity associations. In this regard, both ATOM (Walsh, 2003, 2015) and the polarity correspondence model (Proctor & Cho, 2006) would predict that lighter and heavier weights should be associated with the bottom and the top location of space, respectively. On the other hand, the opposite pattern of results might be expected by considering TEST (Myachykov et al., 2014), which states that spatial representations can be affected by the constraints of the physical world (i.e. tropic representations) and of the human body (i.e. embodied representations). Under a tropic perspective, weight is intrinsically linked to the force of gravity, by which heavier objects are typically associated with the lower parts

---

[6] As suggested by one reviewer, data of Experiment 3b were also analysed excluding the 6 left-handed participants, since a previous study reported an association between stimulus type and response location in right-handers but not in left-handers (see Huber et al., 2015). However, the results of these explorative analyses showed that the two interactions between response location and either size (i.e. the task-relevant dimension) or weight (i.e. the task-irrelevant dimension) were both still non-significant ($ps > .10$).

of our physical world (e.g. generally a car is firmly placed on the ground) and lighter objects are often associated with upper locations (e.g. a leaf can be seen flying in the sky). Moreover, embracing an embodied perspective, the weight of an object is typically appraised through our body—such as when we hold an object in our hands—thus making the sense of weight a genuine bodily experience. Hence, TEST can reasonably predict a heavier-bottom/lighter-top association.

Interestingly, our set of experiments provided supporting evidence for both of these two opposing scenarios. Indeed, on the one hand, the pattern of results from Experiments 1 and 2 clearly indicated a lighter-bottom/heavier-top weight–space association that aligns with previous SNARC and SNARC-like effects and, more generally, with predictions from both ATOM and the polarity correspondence model. On the other hand, Experiment 3a provided supporting evidence for the opposite mapping, with the data indicating a heavier-bottom/lighter-top association. This is a novel result that contrasts sharply with previous literature on the vertical spatial representation of quantities and aligns with the hypothesis of a tropic/embodied representation of weight. A potential explanation for this latter finding might be found in the direct perceptual-motor manipulation of weight adopted in Experiment 3a. More specifically, the request to lift real stimuli may have enforced the notion that heavier weights are associated with a greater downward force than lighter weights, leading to an embodied representation of weight in which ‘heavy is down’ and ‘light is up’. On the contrary, in both Experiments 1 and 2, the perceptual-motor experience was missing, and weight was, therefore, manipulated at a more ‘conceptual’ level. This, in turn, may have favoured a spatial representation of quantities in which ‘less is down’ and ‘more is up’, in line with previous literature on SNARC and SNARC-like effects (e.g. Bruzzi et al., 2017; Müller & Schwarz, 2007).

For the sake of clarity, it is important to point out that further methodological differences characterised Experiments 1/2 and 3a, besides the potential ‘conceptual’ vs. ‘embodied’ distinction. More precisely, in Experiments 1 and 2 weight was manipulated by asking participants to access to long-term knowledge (e.g. the notion that 1 t is heavier than 1 kg), whereas in Experiment 3a this was achieved through an episodic learning procedure. Moreover, in Experiment 3a no reference stimulus was used. Nevertheless, we see no theoretical reasons to consider these differences as responsible for the divergent results reported in Experiments 1/2 vs. 3a. In this regard, it is worth noting that, even if Experiment 3a and 3b were virtually identical, no evidence for a SNARC-like effect emerged in the latter. Hence, this makes it very unlikely that the heavier-bottom/lighter-top association observed in Experiment 3a could be merely due to the methodological differences with Experiments 1/2 discussed above.

The comparison between the results of Experiments 1 and 2 and Experiment 3 shows that the SNARC-like effect for weight reported here is a flexible phenomenon, in that both a lighter-bottom/heavier-top and a lighter-top/heavier-bottom association can emerge depending on the type of stimuli employed to manipulate weight magnitudes. To some extent, these findings are similar to those of previous studies on the flexibility of the SNARC effect. Initial evidence in this regard was provided by Bächtold et al. (1998), who observed a classic left-to-right SNARC effect when participants were asked to mentally represent numbers as distance measures on a ruler, where smaller/larger numbers are placed leftwards/rightwards, respectively. Intriguingly, the direction of the SNARC effect inverted when numbers were mentally represented as time measures on an analogical clock, where smaller/larger numbers are placed rightwards/leftwards, respectively. In a similar vein, Fischer et al. (2010) showed that the direction of the SNARC effect was modulated by the relative position of smaller and larger numbers that participants encountered in a text they read before performing a parity judgement task. More related to the vertical dimension, Holmes and Lourenco (2012) reported a vertical SNARC effect when numbers were mentally represented by participants as building floors or depth levels of a swimming pool, but not when they were represented as items on a shopping list. Furthermore, Hartmann et al. (2014) showed that the typical smaller-bottom/greater-top association was inverted when responses to smaller numbers were provided by hand and greater numbers by foot, likely reflecting a smaller-hand/greater-foot association. To the best of our knowledge, our study provides the first evidence that the flexibility of space–magnitude associations extends beyond numbers and embraces even non-numerical magnitudes. Moreover, it is important to note that in the aforementioned studies the SNARC direction was altered by means of artificially-induced (Bächtold et al., 1998; Fischer et al., 2010; Holmes & Lourenco, 2012) or effector-related (Hartmann et al., 2014) manipulations, whereas the direction of the vertical SNARC-like effect reported here was more likely shaped by the nature (i.e. conceptual vs. concrete) of the weight stimuli. More precisely, our results suggest that the embodied/tropic representation of weight can only be activated through bodily interactions with real objects. In the near future, it will be interesting to explore whether alternative experimental manipulations (e.g. asking participants to watch a confederate weighing real stimuli) are also capable to activate a vertical representation of weight and if perceptual-motor experience with real stimuli can also affect the spatial representation of other magnitudes.

Finally, another important finding emerging from this work is that the SNARC-like effect reported in Experiment

3a was completely absent in Experiment 3b. This could reflect the fact that in Experiment 3a weight was a task-relevant dimension (i.e. participants were asked to classify the coloured circle related to either a light or a heavy sphere) whereas in Experiment 3b weight was task-irrelevant (participants were asked to classify the coloured circle related to either a small or a big sphere). This pattern of results seems to suggest that the vertical SNARC-like effect for weight is related to the explicit processing of weight magnitude. The results of Experiments 3a and 3b also suggest that size might not be represented along a vertical dimension, although we cannot rule out the hypothesis that a vertical spatial representation for this physical variable could emerge by adopting different stimuli, such as geometrical shapes (see Ren et al., 2011) or words referring to real objects (see Sellaro et al., 2015) differing in size.

Conclusion

Four experiments explored the possible vertical representation of weight. Interestingly, we observed that lighter and heavier weights can be associated with both the lower and the upper location of space, depending on whether participants had a direct physical experience with the weight dimension or not. This is novel evidence suggesting that the spatial representation of weight, and potentially other non-numerical quantities can be context-dependent and shaped by physical world constraints and bodily interactions with real objects.

Acknowledgements We thank Giada Alessi, Eleonora Baldini and Giacomo Fedrigo for assistance in data collection. We also thank Massimo Grassi for providing us with the vertical response box employed here.

Data availability The datasets generated and analysed during the current study are available in the Open Science Framework repository, https://osf.io/43s8u/?view_only=7d78c3cc5e06461282c6a1f93c66a308.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

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

Baroni, M., Bernardini, S., Ferraresi, A., & Zanchetta, E. (2009). The WaCky wide web: A collection of very large linguistically processed web-crawled corpora. *Language resources and evaluation, 43*, 209–226. https://doi.org/10.1007/s10579-009-9081-4.

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

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

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

Bächtold, D., Baumüller, M., & Brugger, P. (1998). Stimulus-response compatibility in representational space. *Neuropsychologia, 36*, 731–735. https://doi.org/10.1016/S0028-3932(98)00002-5.

Bruzzi, E., Talamini, F., Priftis, K., & Grassi, M. (2017). A SMARC effect for loudness. *iPerception, 8*, 2041669517742175. https://doi.org/10.1177/2041669517742175.

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

Buckingham, G. (2014). Getting a grip on heaviness perception: A review on weight illusions and their probable causes. *Experimental Brain Research, 232*, 1623–1629. https://doi.org/10.1007/s00221-014-3926-9.

Buckingham, G., Ranger, N. S., & Goodale, M. A. (2011). The material-weight illusion induced by expectations alone. *Attention, Perception, & Psychophysics, 73*, 36–41. https://doi.org/10.3758/s13414-010-0007-4.

Cantlon, J. F., Platt, M. L., & Brannon, E. M. (2009). Beyond the number domain. *Trends in Cognitive Sciences, 13*, 83–91. https://doi.org/10.1016/j.tics.2008.11.007.

Carraro, L., Dalmaso, M., Castelli, L., Galfano, G., Bobbio, A., & Mantovani, G. (2017). The appeal of the devil's eye: social evaluation affects social attention. *Cognitive Processing, 18*, 97–103. https://doi.org/10.1007/s10339-016-0785-2.

Chang, S., & Cho, Y. S. (2015). Polarity correspondence effect between loudness and lateralized response set. *Frontiers in Psychology, 6*, 683. https://doi.org/10.3389/fpsyg.2015.00683.

Cho, Y. S., Bae, G. Y., & Proctor, R. W. (2012). Referential coding contributes to the horizontal SMARC effect. *Journal of Experimental Psychology: Human Perception and Performance, 38*, 726–734. https://doi.org/10.1037/a0026157.

Cohen Kadosh, R., Lammertyn, J., & Izard, V. (2008). Are numbers special? An overview of chronometric, neuroimaging, developmental, and comparative studies of magnitude representation. *Progress in Neurobiology, 84*, 132–147. https://doi.org/10.1016/j.pneurobio.2007.11.001.

Crainiceanu, C., & Ruppert, D. (2004). Likelihood ratio tests in linear mixed models with one variance component. *Journal of the Royal Statistical Society: Series B, 66*, 165–185. https://doi.org/10.1111/j.1467-9868.2004.00438.x.

Dalmaso, M., Galfano, G., Coricelli, C., & Castelli, L. (2014). Temporal dynamics underlying the modulation of social status on social attention. *PLoS ONE, 9*, e93139. https://doi.org/10.1371/journal.pone.0093139.

Dalmaso, M., & Vicovaro, M. (2019). Evidence of SQUARC and distance effects in a weight comparison task. *Cognitive Processing, 20*, 163–173. https://doi.org/10.1007/s10339-019-00905-2.

Dehaene, S., Bossini, P., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General, 122*, 371–396. https://doi.org/10.1037/0096-3445.122.3.371.

Dehaene, S., Dupoux, E., & Mehler, J. (1990). Is numerical comparison digital? Analogic and symbolic effects in two-digit number comparison. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 626–641. https://doi.org/10.1037/0096-1523.16.3.626.

Di Rosa, E., Bardi, L., Umiltà, C., Masina, F., Forgione, M., & Mapelli, D. (2017). Transcranial direct current stimulation (tDCS) reveals a dissociation between SNARC and MARC effects: Implication for the polarity correspondence account. *Cortex, 93*, 68–78. https://doi.org/10.1016/j.cortex.2017.05.002.

Estes, Z., Verges, M., & Barsalou, L. W. (2008). Head up, foot down: Object words orient attention to the objects' typical location. *Psychological Science, 19*, 93–97. https://doi.org/10.1111/j.1467-9280.2008.02051.x.

Fernandez-Prieto, I., Spence, C., Pons, F., & Navarra, J. (2017). Does language influence the vertical representation of auditory pitch and loudness? *i-Perception*. https://doi.org/10.1177/2041669517716183.

Fias, W., Brysbaert, M., Geypens, F., & d'Ydewalle, G. (1996). The importance of magnitude information in numerical processing: Evidence from the SNARC effect. *Mathematical Cognition, 2*, 95–110. https://doi.org/10.1080/135467996387552.

Fischer, M. H. (2012). A hierarchical view of grounded, embodied, and situated numerical cognition. *Cognitive Processing, 13*, 161–164. https://doi.org/10.1007/s10339-012-0477-5.

Fischer, M. H., Mills, R. A., & Shaki, S. (2010). How to cook a SNARC: Number placement in text rapidly changes spatial-numerical associations. *Brain and Cognition, 72*, 333–336. https://doi.org/10.1016/j.bandc.2009.10.010.

Fumarola, A., Prpic, V., Da Pos, O., Murgia, M., Umiltà, C., & Agostini, T. (2014). Automatic spatial association for luminance. *Attention, Perception, & Psychophysics, 76*, 759–765. https://doi.org/10.3758/s13414-013-0614-y.

Gevers, W., Lammertyn, J., Notebaert, W., Verguts, T., & Fias, W. (2006). Automatic response activation of implicit spatial information: Evidence from the SNARC effect. *Acta Psychologica, 122*, 221–233. https://doi.org/10.1016/j.actpsy.2005.11.004.

Hartmann, M., & Mast, F. W. (2017). Loudness counts: interactions between loudness, number magnitude, and space. *Quarterly Journal of Experimental Psychology, 70*, 1305–1322. https://doi.org/10.1080/17470218.2016.1182194.

Hartmann, M., Gashaj, V., Stahnke, A., & Mast, F. W. (2014). There is more than "more is up": Hand and foot responses reverse the vertical association of number magnitudes. *Journal of Experimental Psychology: Human Perception and Performance, 40*, 1401–1414. https://doi.org/10.1037/a0036686.

Hesse, P. N., & Bremmer, F. (2017). The SNARC effect in two dimensions: Evidence for a frontoparallel mental number plane. *Vision Research, 130*, 85–96. https://doi.org/10.1016/j.visres.2016.10.007.

Holmes, K. J., & Lourenco, S. F. (2012). Orienting numbers in mental space: Horizontal organization trumps vertical. *Quarterly Journal of Experimental Psychology, 65*, 1044–1051. https://doi.org/10.1080/17470218.2012.685079.

Holmes, K. J., & Lourenco, S. F. (2013). When numbers get heavy: Is the mental number line exclusively numerical? *PLoS ONE, 8*, e58381. https://doi.org/10.1371/journal.pone.0058381.

Huber, S., Klein, E., Graf, M., Nuerk, H. C., Moeller, K., & Willmes, K. (2015). Embodied markedness of parity? Examining handedness effects on parity judgments. *Psychological Research, 79*, 963–977. https://doi.org/10.1007/s00426-014-0626-9.

Hung, Y., Hung, D. L., Tzeng, O. J. L., & Wu, D. H. (2008). Flexible spatial mapping of different notations of numbers in Chinese readers. *Cognition, 106*, 1441–1450. https://doi.org/10.1016/j.cognition.2007.04.017.

Ishihara, M., Keller, P. E., Rossetti, Y., & Prinz, W. (2008). Horizontal spatial representations of time: Evidence for the STEARC effect. *Cortex, 44*, 454–461. https://doi.org/10.1016/j.cortex.2007.08.010.

Ito, Y., & Hatta, T. (2004). Spatial structure of quantitative representation of numbers: Evidence from the SNARC effect. *Memory & Cognition, 32*, 662–673. https://doi.org/10.3758/BF03195857.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*, 1–26. https://doi.org/10.18637/jss.v082.i13.

Lakens, D. (2012). Polarity correspondence in metaphor congruency effects: Structural overlap predicts categorization times for bipolar concepts presented in vertical space. *Journal of Experimental Psychology: Learning, Memory and Cognition, 38*, 726–736. https://doi.org/10.1037/a0024955.

Lenth, R. V. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Software, 69*, 1–33. https://doi.org/10.18637/jss.v069.i01.

Leth-Steensen, C., & Citta, R. (2016). Bad-good constraints on a polarity correspondence account for the spatial-numerical association of response codes (SNARC) and markedness association of response codes (MARC) effects. *Quarterly Journal of Experimental Psychology, 69*, 482–494. https://doi.org/10.1080/17470218.2015.1055283.

Lidji, P., Kolinsky, R., Lochy, A., & Morais, J. (2007). Spatial associations for musical stimuli: A piano in the head? *Journal of Experimental Psychology: Human Perception and Performance, 33*, 1189–1207. https://doi.org/10.1037/0096-1523.33.5.1189.

Macnamara, A., Keage, H. A., & Loetscher, T. (2018). Mapping of non-numerical domains on space: A systematic review and metaanalysis. *Experimental Brain Research, 236*, 335–346. https://doi.org/10.1007/s00221-017-5154-6.

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgments of numerical inequality. *Nature, 215*, 1519–1520. https://doi.org/10.1038/2151519a0.

Müller, D., & Schwarz, W. (2007). Is there an internal association of numbers to hands? The task set influences the nature of the SNARC effect. *Memory & Cognition, 35*, 1151–1161. https://doi.org/10.3758/BF03193485.

Myachykov, A., Scheepers, C., Fischer, M. H., & Kessler, K. (2014). TEST: A tropic, embodied, and situated theory of cognition. *Topics in Cognitive Science, 6*, 442–460. https://doi.org/10.1111/tops.12024.

Notebaert, W., Gevers, W., Verguts, T., & Fias, W. (2006). Shared spatial representations for numbers and space: The reversal of the SNARC and the Simon effects. *Journal of Experimental Psychology: Human Perception and Performance, 32*, 1197–1207. https://doi.org/10.1037/0096-1523.32.5.1197.

Oberle, C. D., McBeath, M. K., Madigan, S. C., & Sugar, T. G. (2005). The Galileo bias: A naive conceptual belief that influences people's perceptions and performance in a ball-dropping task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 643–653. https://doi.org/10.1037/0278-7393.31.4.643.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*, 97–113. https://doi.org/10.1016/0028-3932(71)90067-4.

Proctor, R. W., & Cho, Y. S. (2006). Polarity correspondence: a general principle for performance of speeded binary classification tasks. *Psychological Bulletin, 132*, 416–442. https://doi.org/10.1037/0033-2909.132.3.416.

Prpic, V., Soranzo, A., Santoro, I., Fantoni, C., Galmonte, A., Agostini, T., et al. (2018). SNARC-like compatibility effects for physical and phenomenal magnitudes: A study on visual illusions. *Psychological Research*. https://doi.org/10.1007/s00426-018-1125-1.

Ren, P., Nicholls, M. E. R., Ma, Y., & Chen, L. (2011). Size matters: non-numerical magnitude affects the spatial coding of response. *PLoS ONE, 6*, e23553. https://doi.org/10.1371/journal.pone.0023553.

Rohrer, D. (2003). The natural appearance of unnatural incline speed. *Memory & Cognition, 31*, 816–826. https://doi.org/10.3758/BF03196119.

Ross, H. E. (2018). Command signals and feedback in the perception of weight and mass. In F. Müller, L. Ludwigs & M. Kupper (Eds.), *Fechner Day 2018 Proceedings of the 34th annual meeting of the International Society for Psychophysics* (pp. 174–178). Lüneburg : International Society for Psychophysics.

Rusconi, E., Kwan, B., Giordano, B. L., Umiltà, C., & Butterworth, B. (2006). Spatial representation of pitch height: The SMARC effect. *Cognition, 99*, 113–129. https://doi.org/10.1016/j.cognition.2005.01.004.

Santiago, J., & Lakens, D. (2015). Can conceptual congruency effects between number, time, and space be accounted for by polarity correspondence? *Acta Psychologica, 156*, 179–191. https://doi.org/10.1016/J.ACTPSY.2014.09.016.

Scheipl, F., Greven, S., & Kuechenhoff, H. (2008). Size and power of tests for a zero random effect variance or polynomial regression in additive and linear mixed models. *Computational Statistics and Data Analysis, 52*, 3283–3299. https://doi.org/10.1016/j.csda.2007.10.022.

Schneider, I. K., Rutjens, B. T., Jostmann, N. B., & Lakens, D. (2011). Weighty matters: Importance literally feels heavy. *Social Psychological and Personality Science, 2*, 474–478. https://doi.org/10.1177/1948550610397895.

Schwarz, W., & Keus, I. M. (2004). Moving the eyes along the mental number line: Comparing SNARC effects with saccadic and manual responses. *Perception & Psychophysics, 66*, 651–664. https://doi.org/10.3758/BF03194909.

Scorolli, C., Borghi, A. M., & Glenberg, A. (2009). Language-induced motor activity in bi-manual object lifting. *Experimental Brain Research, 193*, 43–53. https://doi.org/10.1007/s00221-008-1593-4.

Sell, A. J., & Kaschak, M. P. (2012). The comprehension of sentences involving quantity information affects responses on the up-down axis. *Psychonomic Bulletin & Review, 19*, 708–714. https://doi.org/10.3758/s13423-012-0263-5.

Sellaro, R., Treccani, B., Job, R., & Cubelli, R. (2015). Spatial coding of object typical size: Evidence for a SNARC-like effect. *Psychological Research, 79*, 950–962. https://doi.org/10.1007/s00426-014-0636-7.

Shaki, S., & Fischer, M. H. (2012). Multiple spatial mappings in numerical cognition. *Journal of Experimental Psychology: Human Perception and Performance, 38*, 804–809. https://doi.org/10.1037/a0027562.

Shaki, S., & Fischer, M. H. (2018). Deconstructing spatial-numerical associations. *Cognition, 175*, 109–113. https://doi.org/10.1016/j.cognition.2018.02.022.

Shapiro, L. (2019). *Embodied cognition*. New York: Routledge.

Sixtus, E., Lonnemann, E., Fischer, M. H., & Werner, K. (2019). Mental number representations in 2D space. *Frontiers in Psychology, 10*, 172. https://doi.org/10.3389/fpsyg.2019.00172.

Toomarian, E. Y., & Hubbard, E. M. (2018). On the genesis of spatial-numerical associations: Evolutionary and cultural factors co-construct the mental number line. *Neuroscience and Biobehavioral Reviews, 90*, 184–199. https://doi.org/10.1016/j.neubiorev.2018.04.010.

Vallesi, A., Binns, M. A., & Shallice, T. (2008). An effect of spatial-temporal association of response codes: Understanding the cognitive representations of time. *Cognition, 107*, 501–527. https://doi.org/10.1016/j.cognition.2007.10.011.

Vicovaro, M. (2014). Intuitive physics of free fall: An information-integration approach to the mass-speed belief. *Psicológica, 35*, 463–477.

Vicovaro, M., & Burigana, L. (2014). Properties of the size-weight illusion as shown by lines of subjective equality. *Acta Psychologica, 149*, 52–59. https://doi.org/10.1016/j.actpsy.2014.03.001.

Vicovaro, M., & Burigana, L. (2017). Contribution of surface material and size to the expected versus the perceived weight of objects. *Attention, Perception, & Psychophysics, 79*, 306–319. https://doi.org/10.3758/s13414-016-1212-6.

Vicovaro, M., Noventa, S., & Battaglini, L. (2019). Intuitive physics of gravitational motion as shown by perceptual judgment and prediction-motion tasks. *Acta Psychologica, 194*, 51–62. https://doi.org/10.1016/j.actpsy.2019.02.001.

Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences, 7*, 483–488. https://doi.org/10.1016/j.tics.2003.09.002.

Walsh, V. (2015). A theory of magnitude: The parts that sum to number. In R. Cohen Kadosh & A. Dowker. *The Oxford handbook of numerical cognition* (pp. 552–565). https://doi.org/10.1093/oxfordhb/9780199642342.013.64

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. *Journal of Experimental Psychology: General, 143*, 2020–2045. https://doi.org/10.1037/xge0000014.

Winter, B., Matlock, T., Shaki, S., & Fischer, M. H. (2015). Mental number space in three dimensions. *Neuroscience and Biobehavioral Reviews, 57*, 209–219. https://doi.org/10.1016/j.neubiorev.2015.09.005.

Wood, G., Willmes, K., Nuerk, H. C., & Fischer, M. H. (2008). On the cognitive link between space and number: A meta-analysis of the SNARC effect. *Psychology Science Quarterly, 50*, 489–525. https://doi.org/10.1027/1618-3169.52.3.187.

Zhao, T., He, X., Zhao, X., Huang, J., Zhang, W., Wu, S., et al. (2018). The influence of time units on the flexibility of the spatial numerical association of response codes effect. *British Journal of Psychology, 109*, 299–320. https://doi.org/10.1111/bjop.12273.