Do task-irrelevant direction-associated motion verbs affect action planning? Evidence from a Stroop paradigm

**Carolin Dudschig · Martin Lachmair ·
Irmgard de la Vega · Monica De Filippis · Barbara Kaup**

Abstract Does simply seeing a word such as *rise* activate upward responses? The present study is concerned with bottom-up activation of motion-related experiential traces. Verbs referring to an upward or downward motion (e.g., *rise/fall*) were presented in one of four colors. Participants had to perform an upward or downward hand movement (Experiments 1 and 2a/2b) or a stationary up or down located keypress response (Experiment 3) according to font color. In all experiments, responding was faster if the word's immanent motion direction matched the response (e.g., upward/up response in case of *rise*); however, this effect was strongest in the experiments requiring an actual upward or downward response movement (Experiments 1 and 2a/2b). These findings suggest bottom-up activation of motion-related experiential traces, even if the task does not demand lexical access or focusing on a word's meaning.

Keywords Embodied cognition · Language comprehension · Lexical processing

Interacting with others and with the world is a basic function of life, typically accomplished via motor actions and language. Traditionally, the route from language to meaning is viewed as a building process that combines elements according to syntactic rules. The resulting meaning representations are assumed to reside within memory systems that are separate from the brain's modal systems (e.g.,

C. Dudschig (✉) · M. Lachmair · I. de la Vega · M. De Filippis · B. Kaup
Psychologisches Institut, Universität Tübingen,
Schleichstr. 4,
72076, Tübingen, Germany
e-mail: carolin.dudschig@uni-tuebingen.de

perception, action, introspection) that give rise to experience and knowledge in the first place. In contrast, some modern views of cognition do not make a strong distinction between language understanding and the modal systems in the brain. According to this framework, meaning representations resulting from language comprehension are of a nature similar to representations that result from direct experience of corresponding situations and events. Comprehension is assumed to be tantamount to mentally simulating the experience of the described situations and events (Zwaan & Madden, 2005).

The literature on sentence and discourse comprehension provides substantial evidence for an "experiential-simulations view" of language comprehension. First, neuropsychological studies indicate a considerable overlap between the mental subsystems used for representing linguistically described situations and the mental subsystems that are active during direct experience (e.g., Buccino et al., 2007). Second, behavioral studies demonstrate an interaction between the content of linguistic stimuli and nonlinguistic aspects of the experimental task. These studies suggest that text processing activates perceptual aspects of the described situations, as well as aspects of the involved actions. For example, Glenberg and Kaschak (2002) found that participants were faster to respond to a sentence such as *Close the drawer* when the required response movement matched the movement implied in the sentence (e.g., away from the body) than when it mismatched. This suggests that mechanisms recruited for action planning are also recruited when comprehending sentences describing actions (see also Taylor & Zwaan, 2008; Zwaan & Taylor, 2006). Such results fit with the idea that when comprehending a sentence, people mentally simulate the described situations and actions. One way to account for simulation effects observed with sentences is by means of an active top-down simulation process

that is initiated subsequent to meaning composition (Kaup, Lüdtke, & Steiner, in press).

However, interactions between language and perception or action are also observed using individual words only. For instance, processing action words such as *kick* activates areas of the motor cortex similar to performing the corresponding actions (Hauk, Johnsrude, & Pulvermüller, 2004). Similarly, Meteyard, Zokaei, Bahrami, and Vigliocco (2008) showed that external activation of the motion responsive visual cortex via motion patterns results in interference with word processing, if the word denotes a motion direction that mismatches the activated visual motion (e.g., a downward moving pattern results in a slower lexical decision to the word *rise*). The authors concluded that word meaning is integrated with the visual motion and thus causes interferences. Similarly, Boulenger et al. (2006) showed that action words can have an early influence on motor activation in a grasping task. Interestingly, these word-based simulation effects can be accounted for by a simple associative mechanism, whereby every interaction with the world leaves an experiential trace in the brain. When interacting with the world, people often encounter objects, states, or events together with the words used to refer to these entities. Words get associated with the experiential traces related to their referents in the world. When people later hear or read words referring to the respective objects, states, and events, the corresponding experiential traces get reactivated (Barsalou, 1999; Zwaan & Madden, 2005). This should occur in a relatively automatic fashion when words are being processed. Specifically, this suggests that simulations in language comprehension are not only top-down and subsequent to meaning composition (see above), but also include an early bottom-up component (see also Bub & Masson, 2010).

Despite increasing research on the experiential-simulations view of language comprehension, with converging evidence that single words can activate simulations, there is mixed evidence regarding the exact conditions under which these effects are observed, and the limits of interaction between linguistic stimuli and responding are still to be clarified. Indeed, there are doubts regarding the bottom-up activations of simulations given the extreme task and context dependency of the respective effects. For instance, Van Dam, Rueschemeyer, Lindemann, and Bekering (2010) found that words denoting objects for which the functional use is associated with a movement (e.g., *telephone*) facilitated congruent responses, but only when the words were presented in a context emphasizing the action feature (e.g., *conversation–telephone*). Similarly, it has been suggested that affordances are activated depending on the situational context. For example, compatibility judgements to word pairs such as *drink–glass* are faster if a glass has previously been presented within a reachable distance

(Costantini, Ambrosini, Scorolli, & Borghi, 2011). Additionally, several studies have shown that individual words activate experiential traces related to their literal meaning only when used in a literal rather than a metaphoric or idiomatic manner (Bergen, Lindsay, Matlock, & Narayanan, 2007; Raposo, Moss, Stamatakis, & Tyler, 2009). Bergen et al., for example, showed that literal sentences denoting a downward motion (*The glass dropped*) decrease object categorization performance in a lower location, whereas metaphorical sentences (*The percentage dropped*) do not affect subsequent object categorization. The authors concluded that rather than lexical association on its own triggering simulation effects, it is sentence comprehension. Similarly, the fMRI study by Raposo et al. showed increased activity in the motor cortex when participants heard individual action words (e.g., *kick*) or literal sentences describing an action (e.g., *kick the ball*), but not when presented with these words in idiomatic sentences (e.g., *kick the bucket*). The authors concluded that activation of the motor cortex is modified by the semantic context, suggesting that activation of relevant features is context dependent, which means it is top-down modulated and not automatic. However, because of the temporal limitations of fMRI, it remains unclear whether early motor activation does occur and is subsequently suppressed. Indeed, recently it has been shown that responding can be affected by single words, even if those are used in a metaphoric sentence, and this effect is particularly evident if responses have to be performed 200 ms after critical word onset (Santana & de Vega, 2011).

Taken together, the evidence regarding whether individual words automatically activate simulations is ambiguous. Although there is evidence for simulation effects when processing individual words, such effects are often limited by context (e.g. Bergen et al., 2007; Raposo et al., 2009; Van Dam et al., 2010). Moreover, the majority of studies concerning the activation of simulations in language processing adopted an experimental task that required lexical access either through lexical decision tasks, sensibility judgements, or subsequent memory tasks (e.g., Bergen et al., 2007; Boulenger et al., 2006; Raposo et al., 2009; Santana & de Vega, 2011). Importantly, when reading sentences and accessing their meaning, the context and sentential constraints reported might also result from the following: (a) a lack of single word strength to trigger simulations, and (2) overwriting of single word effects by accessing and integrating them into a larger context. Taken together, those effects do not exclude the possibility that lexical association alone can result in simulation effects. However, it also remains unclear under what conditions lexical associations do trigger simulation effects, and it's questionable whether single words trigger simulation effects if the task does not require lexical access.

In the present study, we provided a first important step toward investigating the limits of word-based experiential effects by lowering the linguistic demands in the task. From the perspective of the experiential-simulations view of language comprehension, lexical access is tantamount to activating the relevant experiential traces (e.g., Zwaan & Madden, 2005). If lexical access is automatic, then so is activating these traces. Thus, if the activation of motion-related experiential traces is automatic, compatibility effects should be observed with simple upward and downward responses even when lexical access to the word's meaning is not required to perform the task. More specifically, we expected to find faster response times to motion verbs such as *rise* and *fall* when the required response involved a compatible upward or downward movement than when the required movement was incompatible, even if lexical access was not required. In contrast, a proponent of an amodal view of language comprehension might propose that the entries in the mental lexicon are amodal in nature and that people automatically activate these amodal entries when processing a word. However, experiential traces are not an integral part of a word's meaning according to such an amodal view. Thus, experiential traces are (if at all) activated subsequent to lexical access, possibly under certain conditions only—for instance, when the task requires lexical access—and thus suggests a deeper processing of the stimulus (see also Mahon & Caramazza, 2008).

Traditionally, studies investigating the influence of irrelevant words on responses can be found in the Stroop literature (for reviews, see Lu & Proctor, 1995; MacLeod, 1991). In the standard Stroop paradigm, irrelevant color words (e.g., *red*) influence naming to the ink color (e.g., blue) of the word stimulus (see Stroop, 1935). Numerous variants of the Stroop task have shown the effect that irrelevant linguistic information has on responding. For example, in the spatial Stroop paradigm, naming the location at which a stimulus is presented on the screen (e.g., up) is slower if the stimulus conveys incongruent locational information (e.g., the word *down*) (e.g., Shor, 1970). Critically, in experiments investigating the spatial Stroop effect, the stimulus set typically directly overlaps with the response set (e. g., words, *up* or *down*; response, naming stimulus location by saying "up" or "down"). Interestingly, in one study, researchers investigated the effects of semantic gradient manipulations in the case of the spatial Stroop task (Fox, Shor, & Steinman, 1971). In this study, participants had to name stimulus location on a chart (up, down, left, or right) and ignore the word meaning. The study results showed an influence of direction-associated nouns (*north*, *east*, *south*, and *west*) on naming the stimulus location on the stimulus chart. However, the study results failed to show an influence of direction-associated verbs (*lift*, *drop*, *turn*, and *flow*) on naming the word's location. In the context of the Stroop

literature, it is hard for one to argue why this effect could not be found if the word's spatial information overlaps with the naming of the locations. It is well known that the response mode (naming vs. keypress) plays a crucial role for the occurrence of the spatial Stroop effect. Typically, naming responses are affected by irrelevant words, but keypress responses are not (cf. Lu & Proctor, 2001).[1] According to Kornblum's, Hasbroucq's, and Osman's (1990) dimensional-overlap model, Fox et al.'s paradigm should show rather large interference effects. This is because the irrelevant response dimension (verbal stimulus) has a direct dimensional overlap with the response mode (naming/vocalization). Thus, the irrelevant word has privileged access to the response system, and the potential for interference should be particularly large (see Lu & Proctor, 2001). In summary, to date there is converging evidence that location words (e.g., *up* and *down*) interfere with vocal responses according to stimulus location (i.e., saying "up" or "down"); however, there is no evidence that direction-associated verbs interfere with either vocal or manual responding. Thus, in the context of the spatial Stroop literature, a finding of spatial activation through direction-associated verbs would also be of interest.

From the experiential-simulation view of language, there has been one previous study by Bub, Masson, and Cree (2008) that was specifically concerned with the question of whether a certain type of information associated with the referents of particular words is activated *automatically* when these words are just being presented and no lexical task has to be performed. The study focused on gestural knowledge—that is, knowledge about how one typically interacts with a particular object. Participants were presented with pictures of objects (Experiments 1 & 2) and words denoting objects (Experiments 3 & 4). The participants' task was to respond to the color of the picture/word with hand postural gestures (see Stroop, 1935). Reaction times to produce a gestural response were faster in compatible conditions, but only for the pictures, not for the words. A compatibility effect with words was observed only when participants were required to read the words in a lexical decision experiment. Bub et al. proposed that knowledge about how one interacts with an object becomes available automatically via experiential traces when participants see a picture of an object, but not when they see a word referring to this object. However, whether the results of Bub et al. (2008) can be generalized remains unclear. Specifically, the response setting adopted within Bub et al. required complex gestural responses to be prepared (e.g., open grasp gesture to the word *nutcracker*). This may have limited the influence

---

[1] Lu and Proctor (2001) showed a significant effect of the irrelevant word *left* or *right* on left or right keypress responses according to color or an arrow symbol if the irrelevant word was preexposed to the relevant response information.

of automatically activated knowledge. In principle, it therefore seems possible that more simple aspects of word meaning, such as the typical location of their referents, do get activated automatically.

In the present experiments, we used a color-response paradigm in which participants are required to indicate font color of individual motion words (e.g., *rise* or *fall*) with an up-(ward) or down-(ward) response. Thus, analogous to the Stroop paradigm, word meaning is task irrelevant. If the processing of motion verbs automatically activates related experiential traces in a bottom-up manner, compatibility effects should be observed, even if no lexical access is required in order to perform the task. Additionally, we were interested in temporal characteristics of such a compatibility effect and whether it affects responding in general or is limited to certain subprocesses of responding (e.g., response preparation, planning, or execution). If responding in general is affected by the compatibility between response direction and motion direction in the verb, we would expect all aspects of a response, from planning to actual response execution, to be affected. Researchers in previous studies have typically reported an influence on premotor processing, such as response planning (e.g., Glenberg & Kaschak, 2002). However, although a motor plan is typically assembled before movement execution, the execution of this plan is not unshakable and is continuously modified and updated (Desmurget & Grafton, 2000). Thus, movement execution might also be affected by the compatibility between the required response and the verb to be processed.

Experiment 1

Method

Participants Thirty German native speakers (nine male; M_{age} = 23.73, SD = 5.29) participated in Experiment 1 for course credit. Three participants were excluded because of low accuracy in at least one condition (< 90 %).

Material Forty German verbs[2] such as *steigen* ("rise") and *fallen* ("fall") were presented in one of four font colors: blue (rgb, 0, 0, 255), orange (rgb, 255, 128, 0), lilac (rgb, 150, 0, 255) and brown (rgb, 140, 80, 20), on a white background. Additionally, four verbs served as stimuli during practice trials. Words were controlled for frequency with the "Wortschatz Portal" of University of Leipzig (http://wortschatz.uni-leipzig.de), for length and for motion direction (vertical axis): For this purpose, 32 volunteers—none of whom participated in the experiment proper—rated 110 verbs

_____
[2] The stimulus material can be found on our website: http://www.uni-tuebingen.de/index.php?id=26912

regarding their typical location on a 5-point Likert scale, with 1 being *down* and 5 being *up* for half of the participants, and vice versa for the other half. In a second step, word length and frequency were matched across the two categories of vertical position, resulting in 20 up words (letters: M = 9.05, SD = 2.11), and 20 down-words (letters: M = 8.65, SD = 2.03). Up- and down-words did not differ significantly with regard to frequency, $t(38)$ = 0.83, p = .41, or length, $t(38)$ = 0.61, p = .55, but did differ significantly for the rated position, M_{up} = 4.29, SD = 0.18, M_{down} = 1.72, SD = 0.13, $t(38)$ = 50.49, p < .001.

Procedure and Design Participants were instructed to respond to the colors as fast and as accurately as possible. The mapping of colors to upward and downward response was balanced across participants, whereby two colors (e.g., blue and orange) were mapped to the upward response, and two colors (e.g., lilac and brown) to the downward response. The experimental procedure was implemented using MATLAB 7.9.0 and Psychophysics Toolbox 3.0.8/Revision 1630. The stimuli were displayed centrally on a 17-in. CRT monitor, and visual angle varied according to word length between 1.43 ° and 4.65 °. Responses were recorded using a standard keyboard in an upright position in the vertical plane in front of the participants. The keyboard was placed inside a response box that could be vertically mounted at the table via a bench vice. Four buttons that were 2.5 cm in diameter were arranged on top of the response box in a line (see Fig. 1) corresponding to the two middle keys ("u," "o") and the "up" and "down" keys ("tab", "end") on the keyboard.

At the beginning of each trial, participants simultaneously held down the two middle keys with their left and right hands, respectively (left hand, lower middle key; right hand, upper middle key). After the fixation cross (800 ms), the stimulus was displayed until response. An up-response was made by releasing the upper middle key and pressing the upper key. A down-response was made by releasing the lower middle key and pressing the lower key. Participants were required to keep the nonresponding hand stationary on the starting position. Response times (RT) were measured as the time to release the initial button. Movement times (MT) were measured as the movement duration from releasing the initial middle button to reaching the up or down button. After response, the participants returned to the starting position. A total of 640 experimental trials were conducted, subdivided into eight blocks, and separated by a self-paced break with error feedback.

A 2 (word: up vs. down) x 2 (response: up vs. down) design was implemented with repeated measurement on both variables in the by-participants analysis (F_1) and repeated measurement on response location in the by-items analysis (F_2).
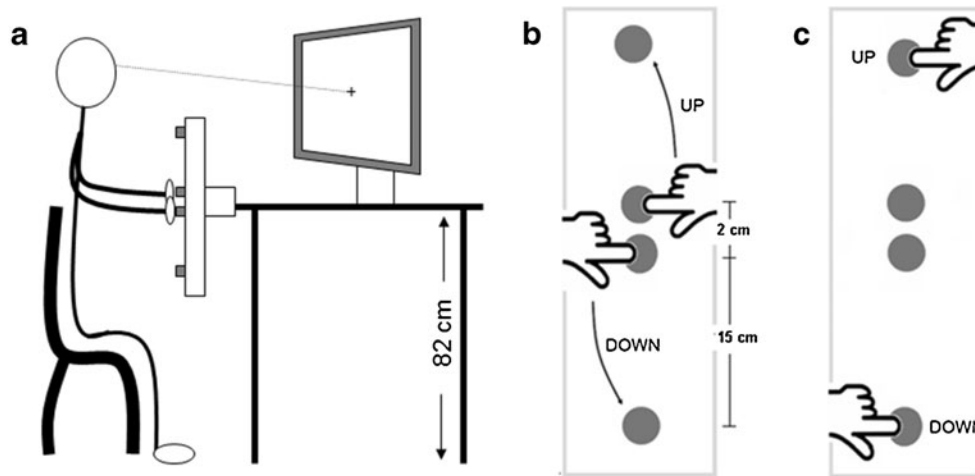
Fig. 1 Experimental setup and response key apparatus placed over a standard German keyboard. **a** Experimental setup: Participants responded with the right and left hand on a response apparatus vertically attached at the table slightly higher than a standard desk. **b** Response action for Experiments 1 and 2: Participants responded by releasing one of the middle keys and subsequently pressing the according upper or lower key. **c** Response action for Experiment 3: Participants kept their hands stationary and responded by pressing the upper or lower key

Results

Erroneous responses as well as responses below the 100-ms cut-off value were excluded from analyses. To determine outliers, we conducted z-transformations for each of the words in each of the four conditions, according to the normalized RTs of each participant. Responses with z-values above 2 or below −2 were excluded. This exclusion reduced the data set by less than 4.4 %. Mean RTs in the four conditions are displayed in Fig. 2a.

An ANOVA[3] conducted on the RTs showed a main effect of response direction, which however, was significant only in the by-items analysis $F_1(1, 26) = 1.86$, $p > .10$; $F_2(1, 38) = 14.67$, $p < .001$. Numerically, responses were faster for up- (542 ms) than for down- (551 ms) responses. This difference probably reflects the fact that up responses were performed with the dominant right hand. There was no main effect of word, $F(1, 26) < 1$; $F_2(1, 38) < 1$. Importantly, there was a significant word × response interaction, with responses being significantly faster in congruent trials (Up_{Word}–$Up_{Response}$, 534 ms; $Down_{Word}$–$Down_{Response}$, 544 ms) than in incongruent trials (Up_{Word}–$Down_{Response}$, 558 ms; $Down_{Word}$–$Up_{Response}$, 550 ms), $F_1(1, 26) = 10.54$,

$p < .01$; $F_2(1, 38) = 39.53$, $p < .001$. Post hoc tests showed that up-responses were faster for up words than for down-words, $t_1(26) = -2.83$, $p < .01$; $t_2(19) = -5.37$, $p < .001$, and that down-responses were faster for down-words than for up words, $t_1(26) = 3.37$, $p < .01$; $t_2(19) = 4.06$, $p < .001$.

An error analysis mirrored the RT results. There was no main effect of response or word (both $Fs < 1.1$). However, the interaction was significant, $F_1(1, 26) = 12.44$, $p < .01$; $F_2(1, 38) = 12.41$, $p < .01$, showing that fewer errors were committed in congruent (Up_{Word}–$Up_{Response}$, 3.03 %; $Down_{Word}$–$Down_{Response}$, 2.73 %) than in incongruent (Up_{Word}–$Down_{Response}$, 3.61 %; $Down_{Word}$–$Up_{Response}$, 3.80 %) conditions.

We conducted a further analysis of RT distributions in order to analyze the temporal characteristics of the compatibility effect (Ratcliff, 1979). Additionally, we were interested in whether compatibility between response and word affects all parts of responding or is limited to stages before movement execution (i.e., before MT). First, RTs in the compatible and incompatible conditions were grouped into deciles separately for each participant. An ANOVA with the additional factor decile confirmed the compatibility effect, $F(1, 26) = 11.19$, $p < .01$, and showed a compatibility × decile interaction, $F(9, 234) = 16.00$, $p < .001$. Analogous to the typical Stroop effect's distribution (e.g., Pratte, Rouder, Morey, & Feng, 2010), the compatibility effect observed in the present study was larger in longer RTs (see delta plot, Fig. 2b). Second, we conducted an ANOVA with the corresponding MTs as the dependent variable, which did not show any effect of compatibility, $Fs < 1$, and no interaction with decile, $F(9, 234) = 1.42$, $p = .18$.

In summary, compatibility affected slow responses more than fast responses, suggesting that the compatibility effect

---

[3] In addition to the standard ANOVA, a mixed-effects model analysis was conducted. The lmer function of the lme4 package (see Bates, Maechler, & Bolker, 2011) in the open-source statistical programming environment R was used to fit a mixed-effects model to the data set. Model comparison resulted in random intercepts for participant and item, and by-subject random slopes for reaction. The model with the fixed-effect Response*Word Location showed the best fit for RTs, $\Delta \chi2 = 61.19$, $p < .001$, and errors, $\Delta \chi2 = 8.64$, $p = .003$, but not for MTs, $\Delta \chi2 = 0.02$, $p = .88$. In summary, the results of the mixed-model analysis were consistent with the results of the F1/F2 analysis.
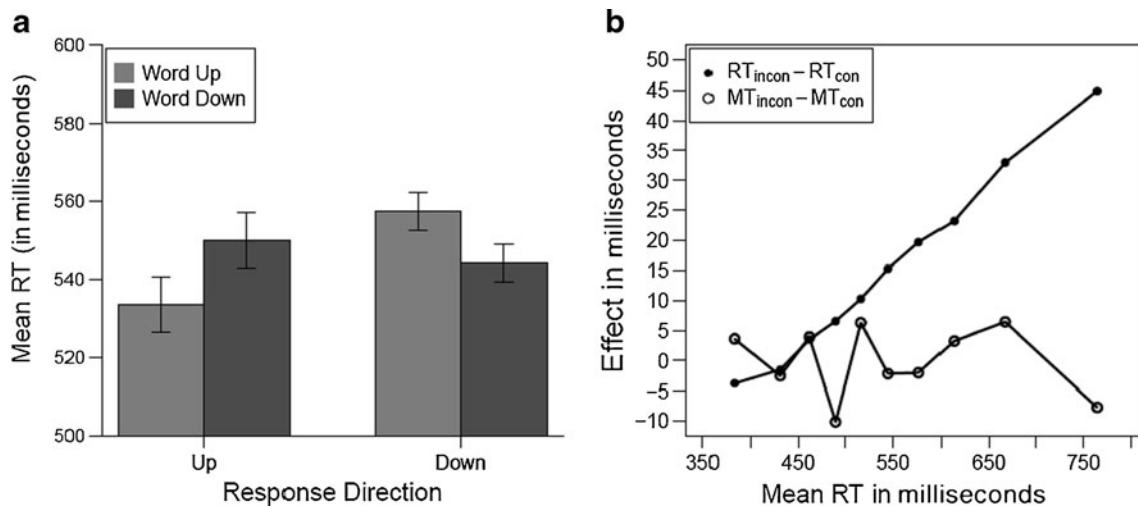
Fig. 2 Results of Experiment 1. **a** Mean RTs as a function of response direction and word direction. *Error bars* represent the 95 % confidence interval for the pairwise contrasts in within-subjects designs (Loftus & Masson, 1994). **b** Delta plot (De Jong, Liang, & Lauber, 1994) drawn from RT distributions. The *dots* in the lines represent the deciles (1st to 10th) for the RT and corresponding MT distributions. The *y*-axis shows the size of the compatibility effect, at a particular decile. The *x*-axis provides the average RT

takes some time to develop. At first sight, this may seem surprising in the context of arguing in favor of automatic activation of experiential traces. However, effects developing with increasing RTs is a typical finding in the standard Stroop paradigm (Pratte et al., 2010) and is usually not taken to contradict bottom-up mechanisms (De Jong, Liang, & Lauber, 1994). Interestingly, in the analysis of the MTs, we could see that the action-word compatibility does not systematically affect movement execution at any point. This suggests that compatibility affects response processes only before movement execution, such as response planning and response preparation. Interestingly, researchers in previous studies typically reported button release times (e.g., Glenberg & Kaschak, 2002) and thus concluded that it is the planning of an action that is affected by the compatibility of the linguistic stimuli. Whether response execution is affected as well has typically remained unaddressed. Our results suggest that movement execution in a simple aiming task as in the present experiment is not affected by the compatibility between response direction and motion direction implied by the verb. This might be because of the fact that the ballistic, preprogrammed, and purely executed part of the movement is predominant in a rather simple aiming task as the one implemented in the present experiment (Desmurget & Grafton, 2000). However, a strict embodied perspective of language processing would still predict that the compatibility between language and action affects response execution in tasks that involve a larger amount of nonballistic components. Indeed, it has been shown that in a more complex reaching task motor responses can be affected even if words are presented after movement onset (Boulenger et al., 2006).

Taken together, in the present experiment, responses were faster when the response direction matched the verb's

motion direction, despite the task not requiring lexical access. Critically, one could argue that the present stimulus material might have primed directional interpretation, since only verbs implying an upward or downward motion were shown to the participants. Experiment 2a was conducted to rule out this possibility. Another problem with the Experiment 1 was that upward responses were always conducted with the right and downward responses with the left hand. Thus, the compatibility effect could in principle also be attributed to a match between up words and right hand responses, and down - words and left-hand responses (e.g. Cho & Proctor, 2002). Experiment 2b was conducted to rule out this possibility in the present experimental setup.

Experiment 2

Experiments 2a and 2b were identical to Experiment 1 with the difference that filler verbs not implying a vertical motion (e.g., *feed*, *treat*, *surround*) were introduced in order to prevent a directional interpretation through stimulus material. Moreover, in Experiment 2b, participants responded upward with their left hands and downward with their right hands.

Experiment 2a

Method

Participants Twenty-eight German native speakers (seven male; $M_{age} = 26.04$, $SD = 4.81$) participated for course credit or payment. One participant was excluded because of low accuracy in one condition (<90 %).

Material Material was identical to that in Experiment 1, except that 20 filler words were included that were rated as neutral with respect to their motion direction ($M = 3.00$ on a 5-point Likert scale; see Experiment 1).

Procedure and Design Design and procedure were identical to those in Experiment 1.

Results

Data were analyzed analogue to those in Experiment 1. Outlier elimination reduced the data by less than 4.4 %. Mean RTs in the four conditions are displayed in Fig. 3.

An ANOVA[4] conducted on the RTs showed a main effect of response direction, $F_1(1, 26) = 4.93$, $p < .05$; $F_2(1, 38) = 28.48$, $p < .001$. Responses were faster for up responses (515 ms) than for down responses (528 ms), which probably reflects the fact that up responses were performed with the dominant right hand. There was no effect of word, $F(1, 26) < 1$; $F_2(1, 38) < 1$. Importantly, there was a significant interaction between word and response direction, with responses being significantly faster in congruent trials (Up$_{Word}$–Up$_{Response}$, 509 ms; Down$_{Word}$–Down$_{Response}$, 522 ms) than in incongruent trials (Up$_{Word}$–Down$_{Response}$, 535 ms; Down$_{Word}$–Up$_{Response}$, 522 ms), $F_1(1, 26) = 24.19$, $p < .001$; $F_2(1, 38) = 26.82$, $p < .001$. Post hoc tests showed that the up responses were faster for up words than for down words, $t_1(26) = -3.74$, $p < .001$; $t_2(19) = -4.59$, $p < .001$, and that down responses were faster for down than for up words, $t_1(26) = 5.47$, $p < .001$; $t_2(19) = 3.45$, $p < .01$.

Error analysis mirrored the RT results. There was no main effect of response or word (all $Fs < 1$). However, the interaction was significant, $F_1(1, 26) = 8.87$, $p < .01$; $F_2(1, 38) = 8.05$, $p < .01$, showing that fewer errors were committed in congruent (Up$_{Word}$–Up$_{Response}$, 1.62 %; Down$_{Word}$–Down$_{Response}$, 1.55 %) than in incongruent (Up$_{Word}$–Down$_{Response}$ 2.13 %; Down$_{Word}$–Up$_{Response}$, 2.36 %) conditions.

An additional analysis of RT and MT distributions were conducted analogous to Experiment 1. First, an ANOVA with the factor decile confirmed the compatibility effect, $F(1, 26) = 15.72$, $p < .001$, and showed a compatibility × decile interaction, $F(9, 234) = 12.42$, $p < .001$. As in Experiment 1, the word–action compatibility effect was

larger in longer RTs (see delta plot, Fig. 3b). Second, we conducted an ANOVA with the corresponding MTs as the dependent variable, which did not show any effect of compatibility, $Fs < 1$, and no interaction with decile, $Fs < 1$.

In summary, in Experiment 2a, we replicated the results of Experiment 1. Thus, the compatibility effect between verbs implying a vertical motion and responding with an up or down response can be found when a more heterogeneous stimulus set is implemented. Thus, we can conclude that even if directional interpretations were not obvious, compatibility effects do emerge.

Experiment 2b

Method

Participants Thirty German native speakers (nine male; $M_{age} = 26.10$, $SD = 5.24$) participated for course credit or payment. Four participants were excluded because of low accuracy in one condition (<90 %), and one participant was excluded because of technical problems in data file storage.

Material The material was identical to that in Experiment 2a.

Procedure and Design The procedure and design were identical to those in Experiment 2a. However, upward responses were now conducted with the left hand, and downward responses with the right hand.

Results

Data were analyzed analogous to the previous experiments. Outlier elimination reduced the data by less than 4.6 %. Mean RTs in the four conditions are displayed in Fig. 4a.

An ANOVA[5] showed a main effect of response direction for RTs in the by-item analysis, $F_1(1, 24) = 2.19$, $p = 0.15$; $F_2(1, 38) = 29.97$, $p < .001$. Responses were faster for down responses (553 ms) than for up responses (566 ms), which probably reflects the fact that down responses were now performed with the dominant right hand. There was no effect for word, $F_1(1, 24) = 1.20$, $p = 0.28$; $F_2(1, 38) < 1$.

---

[4] Analogous to the previous experiment, a mixed-effects model was fitted to the data set. Model comparison resulted in random intercepts for participant and item, and by-subject and by-item random slopes for reaction. The model with the fixed-effect Response*Word Location showed the best fit for RTs, $\Delta \chi2 = 20.14$, $p < .001$, and errors, $\Delta \chi2 = 4.42$, $p < .05$, but not for MTs, $\Delta \chi2 = 1.35$, $p = .25$. In summary, the results of the mixed-model analysis were consistent with the results of the F1/F2 analysis.

[5] Analogous to the previous experiments, a mixed-effects model was fitted to the data set. Model comparison resulted in random intercepts for participant and item, and by-subject and by-item random slopes for reaction. The model with the fixed-effect Response*Word Location showed the best fit for RTs, $\Delta \chi2 = 36.02$, $p < .001$, and errors, $\Delta \chi2 = 5.69$, $p < .05$, but not for MTs, $\Delta \chi2 = 1.17$, $p = .28$. In summary, the results of the mixed model analysis were consistent with the results of the F1/F2 analysis.
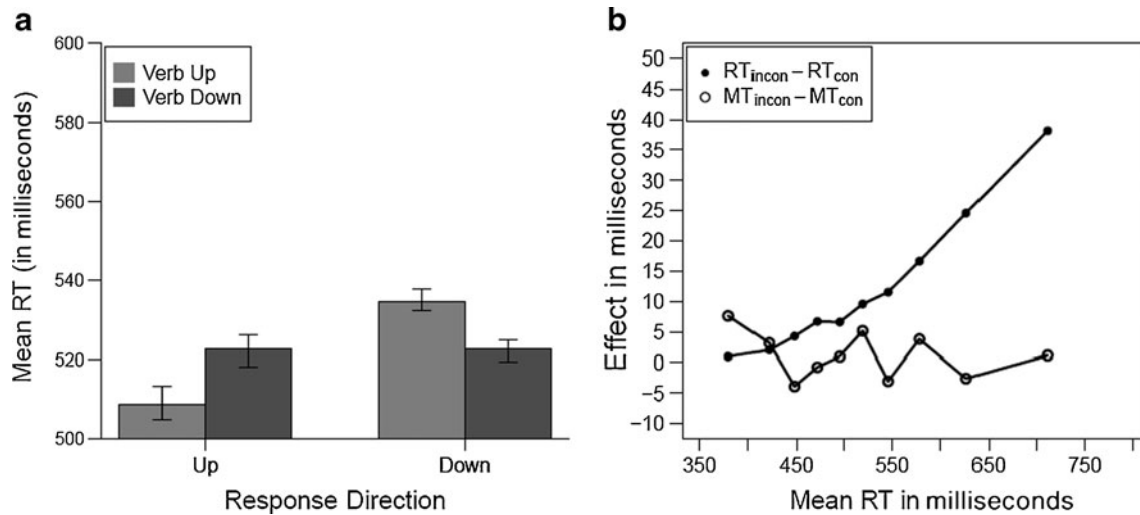
Fig. 3 Results of Experiment 2. **a** Mean RTs as a function of response direction and word direction. *Error bars* represent the 95 % confidence interval for the pairwise contrasts for within-subjects designs (Loftus & Masson, 1994). **b** Delta plot (De Jong et al., 1994) drawn from RT

distributions. The *dots* in the lines represent the deciles (1st to 10th) for the RT distributions and corresponding MTs for each decile. The *y*-axis shows the size of the compatibility effect, at a particular decile. The *x*-axis provides the average RT

Importantly, there was a significant interaction between word and response direction, with responses being significantly faster in congruent trials (Up$_{Word}$–Up$_{Response}$, 561 ms; Down$_{Word}$–Down$_{Response}$, 545 ms) than in incongruent trials (Up$_{Word}$–Down$_{Response}$, 562 ms; Down$_{Word}$–Up$_{Response}$, 572 ms), $F_1(1, 24) = 18.79$, $p < .001$; $F_2(1, 38) = 36.47$, $p < .001$. Post hoc tests showed that the up responses were faster for up words than for down words, $t_1(24) = −2.91$, $p < 0.01$; $t_2(19) = −3.77$, $p < .01$, and that down responses were faster for down than for up words, $t_1(24) = 4.20$, $p < .001$; $t_2(19) = 5.12$, $p < .001$.

Error analysis confirmed the RT results. There was no main effect of response or word (all Fs < 1). However, the interaction between response and word was significant, $F_1(1, 24) = 5.52$, $p < .05$; $F_2(1, 38) = 5.80$, $p < .05$, showing that fewer errors were committed in congruent (Up$_{Word}$–Up$_{Response}$, 1.67 %; Down$_{Word}$–Down$_{Response}$, 1.60 %) than in incongruent (Up$_{Word}$–Down$_{Response}$, 1.90 %; Down$_{Word}$–Up$_{Response}$, 2.37 %) conditions.

An additional analysis of RT and MT distributions were conducted. First, an ANOVA with the factor decile confirmed the compatibility effect, $F(1, 24) = 12.95$, $p < .01$,
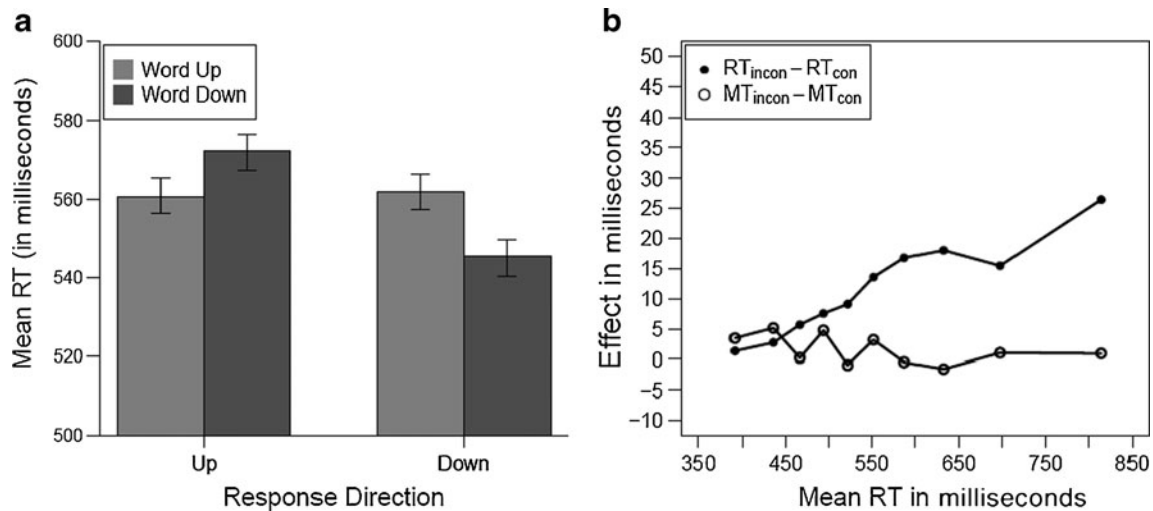

Fig. 4 Results of Experiment 2b. **a** Mean RTs as a function of response direction and word direction. *Error bars* represent the 95 % confidence interval for the pairwise contrasts for within-subjects designs (Loftus & Masson, 1994). **b** Delta plot (De Jong et al., 1994)

drawn from RT distributions. The *dots* in the lines represent the deciles (1st to 10th) for the RT distributions and corresponding MTs for each decile. The *y*-axis shows the size of the compatibility effect, at a particular decile. The *x*-axis provides the average RT

and showed a compatibility × decile interaction, $F(9, 216) =$ 3.91, $p < .001$. Again the word-action compatibility effect develops with increasing RT (see delta plot, Fig. 4b). Second, the ANOVA with the corresponding MTs as the dependent variable did not show an effect of compatibility, $F(1, 24) = 1.06$, $p = .30$, and no interaction with decile, $F < 1$.

In summary, upward responses with the left hand were fastest if paired with words implying an upward motion, and downward response with the right hand were fastest if paired with words implying a downward motion. Comparisons between Experiment 2a and 2b showed no effect of experiment on the interaction between word and action ($Fs < 1$). This shows that the congruency effect observed in the present experiments is not due to compatibility between words implying an upward motion and a right hand action, or words implying a downward motion and a left hand action. In contrast, the response motion (upward or downward) rather than response hand determines the congruency effect.

Experiment 3

In Experiment 3, we investigated whether the congruency effect is dependent on an experimental task that requires a response movement, or alternatively, whether it also occurs with a stationary up/down response. In the previous experiment, we showed that the actual response execution was not affected by compatibility; thus, it remains questionable whether the planning of an upward or downward response movement is needed in order to find the compatibility effect. In the study of Glenberg and Kaschak (2002), the action-sentence compatibility effect emerged only when participants responded with a movement, but not with a stationary keypress. Thus, if there is no relevant interfering action according to the experiential-simulation view, we should not observe a compatibility effect. However, if spatial congruency is sufficient to cause our effects, then resting in a location compatible to the motion's aiming point should result in the same effects as in the previous experiments. In the spatial Stroop literature, Fox et al. (1971) did not find a spatial congruency effect with direction-associated verbs. However, the authors used words for both the left–right (turn, flow) and the up–down (lift, drop) axis, whereby the left–right associations might be rather weak, potentially weakening a spatial congruency effect.

Method

Participants Thirty German native speakers (seven male; $M_{age} = 22.8$, $SD = 5.72$) participated for course credit. Two participants were excluded because of low accuracy in one condition (< 90 %).

Material The material was the same as in Experiment 1.

Procedure and Design The design and procedure were identical to those in Experiment 1, except that participants positioned their right and left hands on the upper and lower keys throughout the experiment (see Fig. 1b).

Results

Data were analyzed as in previous experiments. Outlier elimination reduced the data by less than 4.5 %. Mean RTs are displayed in Fig. 5a. The ANOVA[6] for the RTs showed that responses were significantly faster for up (536 ms) than for down (561 ms) responses, $F_1(1, 27) = 30.56$, $p < .001$; $F_2(1, 38) = 149.82$, $p < .001$. There was no main effect of word (both $Fs < 1$). Importantly, there was a significant interaction between word and response, with faster responses in congruent (Up$_{Word}$–Up$_{Response}$, 534 ms; Down$_{Word}$–Down$_{Response}$, 558 ms) than in incongruent (Up$_{Word}$–Down$_{Response}$, 563 ms; Down$_{Word}$–Up$_{Response}$, 538 ms) conditions, $F_1(1, 27) = 4.68$, $p < .05$; $F_2(1, 38) = 4.85$, $p < .05$. Post hoc tests showed that the down responses were faster for down words than for up words, $t_1(27) = 2.10$, $p < .05$; $t_2(19) = 1.99$, $p = .06$, and that up responses had a tendency to be faster for up than for down words, $t_1(27) = -1.45$, $p = .16$; $t_2(19) = -1.24$, $p = .23$.

Error rates were higher for up than for down responses, $F_1(1, 27) = 5.15$, $p < .05$; $F_2(1, 38) = 5.17$, $p < .05$, and were higher for down than for up words, $F_1(1, 27) = 4.42$, $p < .05$; $F_2(1, 38) = 6.28$, $p < .05$. The interaction was significant, $F_1(1, 27) = 5.00$, $p < .05$; $F_2(1, 38) = 6.54$, $p < .05$, with fewer errors in congruent (Up$_{Word}$–Up$_{Response}$, 1.99 %; Down$_{Word}$–Down$_{Response}$, 1.90 %) than in incongruent (Up$_{Word}$–Down$_{Response}$, 2.05 %; Down$_{Word}$–Up$_{Response}$, 3.04 %) conditions.

Additional analysis of the RT distribution was conducted analogous to the previous experiments. First, RTs in the compatible and incompatible conditions were grouped into deciles separately for each participant (see Fig. 5b). An ANOVA with the factor decile showed no compatibility effect, $F(1, 27) = 2.73$, $p = .11$, but importantly showed a significant compatibility × decile interaction, $F(9, 243) = 2.76$, $p < .01$. As can been seen in the delta plot (Fig. 5b), the compatibility effect increases with decile.

In summary, responses were faster when the response location matched the verb's direction of the described

---

[6] Analogous to the previous experiments, a mixed-effects model was fitted to the data set. Model comparison resulted in random intercepts for participant and item, and by-subject random slopes for reaction. The model with the fixed-effect Response*Word Location showed the best fit for RTs, $\Delta \chi 2 = 5.81$, $p = .02$, and errors, $\Delta \chi 2 = 12.52$, $p < .001$. In summary, the results of the mixed-model analysis were consistent with the results of the F1/F2 analysis
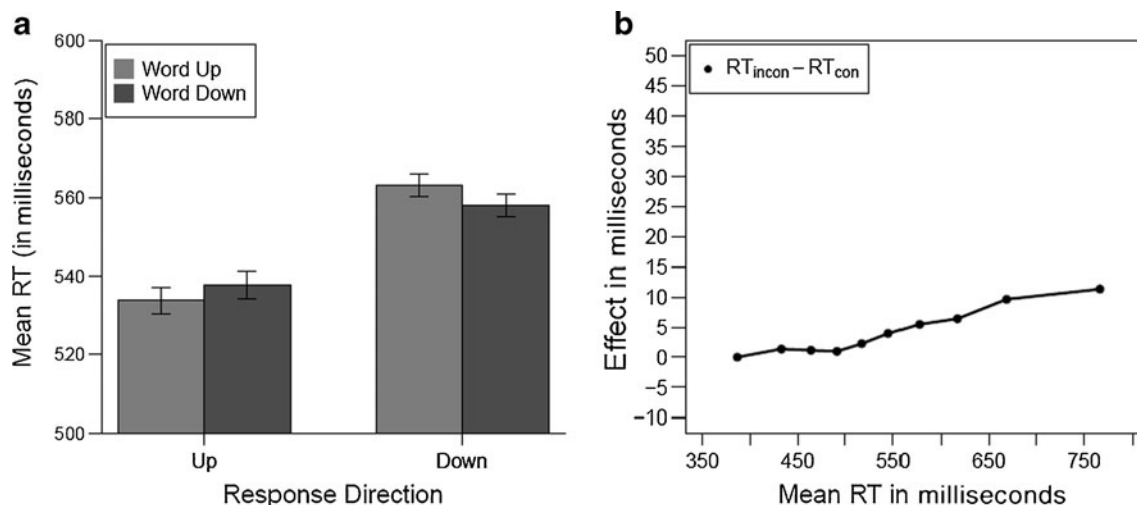
Fig. 5 Results of Experiment 3. **a** Mean RTs as a function of response direction and word direction. *Error bars* represent the 95 % confidence interval for the pairwise contrasts for within-subjects designs (Loftus & Masson, 1994). **b** Delta plot (De Jong et al., 1994) drawn from RT distributions. The *dots* in the lines represent the deciles (1st to 10th) for the RT distributions for each decile. The *y*-axis shows the size of the compatibility effect, at a particular decile. The *x*-axis is an average of the congruent and incongruent conditions

motion, even though only physical location of the response was manipulated. This suggests that stationary responses can be influenced by the motion words and that spatial congruency indeed plays a role in explaining the observed compatibility effects. However, a between-experiment analysis (Experiments 1 and 3) showed a significant three-way interaction between experiment, word, and response, $F(1, 53) = 4.34$, $p < .05$, indicating that the compatibility effect was stronger if a response movement was implemented. In contrast, comparisons among Experiments 1, 2a, and 2b showed no effect of experiment on the word-response interaction ($Fs < 1$). In summary, this suggests that the compatibility effect can be partially explained by spatial congruency. However, action planning as demanded in Experiments 1 and 2 is affected even more by irrelevant words than spatially congruent or incongruent responses are (Experiment 3), suggesting that the observed compatibility effect is not solely due to spatial congruency.

Discussion

Despite behavioral studies providing strong evidence for the experiential-simulations view of language processing, the underlying mechanisms are still underspecified. Controlled top-down simulations probably occur subsequent to comprehension and capture the meaning of the sentence or larger phrase. In addition, there may be an early bottom-up component by which individual words or combinations of words activate experiential traces in memory. However, the literature has not provided a clear picture with respect to this proposed bottom-up component. On the one side, it has been shown that certain experiential-simulation effects

occur only when the sentence context suggests a literal interpretation of the target words (Bergen et al., 2007; Raposo et al., 2009). This speaks against an early and automatic bottom-up component and suggests that top-down processes are predominant. On the other hand, word-based simulation effects have been shown to be very robust, and those could be explained by automatic reactivation of experiential traces (Barsalou, 1999; Zwaan & Madden, 2005). However, simulation effects using single words have typically been reported in paradigms demanding lexical access, and could therefore still be the result of a top-down initiation if word meaning is task relevant. More specifically, it is possible that activation of experiential traces is an additional optional process, which is independent of grasping the word's meaning. Thus, experiential traces might be activated only in scenarios in which a higher level of language processing is demanded. For example, if we need to perform a lexical decision task, we might actively call experiential traces or check for their availability since they help us with the decision. In contrast, experiential traces do not have that significance in a color-response task.

In the present study, we investigated the bottom-up component of motion-related experiential traces by using a task for which word meaning was irrelevant (i.e., a color-response paradigm). If experiential traces get active in a bottom-up manner, then the motion implied by the word should influence responding, despite lexical access being task irrelevant. In Experiments 1 and 2, participants had to perform either an upward or downward movement. As expected, RTs were slower when the motion implied by the verb mismatched the response (e.g., downward response to *rise*). A smaller but still significant congruency effect was found in Experiment 3, in which participants had to perform

stationary responses in an upper or lower location. Taken together, those results support that direction-associates verbs interfere with manual keypress response, even if the task does not require the participants to lexically access the words meaning. These results are in line with the findings of a recent study employing nouns (Lachmair, Dudschig, De Filippis, de la Vega, & Kaup, 2011).

The results from a previous study concerned with the activation of experiential traces clearly contrasts with our results. Bub et al. (2008) showed that simulations reflecting a grasping action occur only if the task requires lexical access to the presented words. It is possible that Bub et al.'s use of rather complex grasping actions limited the impact of bottom-up activations of experiential traces. Alternatively, the discrepant findings may result from differences in the experimental setup. RTs in their study were calculated by taking the time it took participants to initiate a response (i.e., to remove the hand from a single resting key). Thus, in principle, participants could leave the key before having decided which response to perform. In other words, participants might see the words, instantly release the response key (i.e., RT is measured at that point), and subsequently decide what response to perform. Potentially, this may result in a number of wrongly initiated but corrected responses. Given that wrong responses are generally more likely to occur in incongruent conditions, this would result in an artificial shortening of mean RTs in incongruent trials (i.e., wrongly initiated [fast RT] but corrected responses). This may possibly have led to a null result despite the fact that participants activated the relevant knowledge during word processing. The same does not hold for our experimental setup, which forced participants to decide which response to perform before leaving the start key. Here, participants had two release keys: one for the right hand (e.g., upward response) and one for the left hand (e.g., downward response). Once participants released this initial key, they could not correct the response any more (see Fig. 1). Even if they tried to do so, the response was recorded as wrong (see above). Thus, our data set contains only correctly initiated responses. In other words, in our upward RT, we pooled only those trials in which the response was planned, programmed, and initiated as an upward motion, the same being true for our downward RT. In summary, Bub et al.'s procedure might have shown similar results to ours if wrongly initiated responses had been excluded from their analysis.

In the Stroop literature, researchers in one previous study investigated the influence of direction-associated verbs on responding (Fox et al., 1971). In this experiment, the words were presented in a spatial location on a chart, and participants had to respond verbally by indicating the words location ("up," "down," "left," "right"). The results did not show any congruency effects. In contrast, in our present experiments, words were centrally presented, and

participants responded manually. However, if spatial features implied by the words cause the congruency effects, one would have expected the effect to have been even larger in the setup Fox et al. implemented. This is because in Fox et al.'s study, a task-irrelevant verbal stimulus was paired with a vocal response. Pairing irrelevant verbal stimuli with a vocal response typically results in larger spatial Stroop interference than pairing a verbal stimulus with a keypress response, because of dimensional overlap between stimulus and response (for details, see Lu & Proctor, 2001). Thus, the influence of the spatial information conveyed by the words should have been stronger in Fox et al.'s experiments. On the other hand, if the interference effects observed in our present experiments result from activation of motion-related experiential traces rather than spatial congruency, those effects cannot be found with vocal responses, because no action is required in order to respond. However, we also find an interference effect with stationary responses (Experiment 3), in which no direct action is involved in responding. This finding suggests that spatial congruency at least partially plays a role in explaining our effects, and this spatial congruency effect should be also found with vocal responses. Indeed, Fox et al. implemented only four words (lift, drop, turn, flow) and suggested that the left- and right-associated words might have less potential to interact with the naming response because of a lack of activation of a left or right associate. Because our bodies are left–right symmetric, the words in the left–right axis might result in less spatial activation than the activation resulting from the asymmetric up–down dimension (Turner, 1994). Indeed, it is well known that some people have problems differentiating between left and right. Thus, Fox et al.'s results might look different if repeated with only the up and down words—that is, the vertical dimension.

Critically, as was mentioned above, one could argue that our compatibility effect reflects spatial congruency effects rather than activation of experiential traces. Indeed, a small congruency effect was observed in Experiment 3, in which the response involved a stationary keypress only and no upward or downward response action. The congruency effect reflected an interaction between the aiming point of the motion implied by the word and the response location. However, between-experiments comparisons showed that the word-response interaction in Experiment 3 was significantly smaller than in the other experiments, whereby the experiments involving a response action did not differ in the size of the congruency effect. This finding suggests that the effects might be partly due to spatial congruency effects; however, a substantial part is also driven by interferences with action planning.

Alternatively, one could also propose that the smaller congruency effect with stationary responses (Experiment 3) is due to the fact that the words activate more than one spatial position. One could propose that a word such as rise

contains a reference to both the upper and the lower positions. For example, the word *rise* typically refers to a movement from a lower to an upper position. Thus, one could suggest that this word should similarly activate both locations. However, even if that is the case, the present findings suggest that the end position of a motion is dominantly represented over the starting point. One could argue that the anticipation of a future position is of special relevance for an optimal interaction with the world. For instance, anticipating the future position of, say, a falling pen, may allow us to optimally prepare for future actions (e.g., picking up the pen), rather than focus on its past location. This would suggest that the future position or end position of a motion has a high importance and thus might be preliminary represented when processing motion words. Indeed, there is evidence that the endpoint of a motion has a high functional relevance. In a recent TMS study, it was found that pictures of action motions but not pictures of stationary hands increased corticospinal excitability for the according muscles (Urgesi, Moro, Candidi, & Aglioti, 2006). The authors proposed that only pictures of ongoing motions convey information about what happens next and thus allow anticipatory representations. Similarly, Coventry et al. (2010) showed the importance of end states of falling objects (e. g., rain) when processing language and building situational representations. However, even if our words do activate one or more spatial locations, our congruency effects are still strongest if the response involves an action, suggesting that language-action interferences are at least partially the source of our compatibility effects. In summary, we suggest that action planning does play a crucial role in explaining our effects.

Another question that needs to be addressed: In what way did visual motion play a role in our present experiments? In the first place, we must state that we measured RTs at a time point at which the participants released the middle buttons. Thus, at the critical time point of RT measurement, participants did not yet perceive any visual motion of their hands but were just planning their response actions. Second, our response buttons were rather large and easy to hit when performing the experiment. Indeed, mean button release times in Experiments 1 (546 ms), 2a (522 ms), and 2b (560 ms) are comparable to stationary keypresses in Experiment 3 (548 ms), suggesting that there was no additional visual search involved before participants released the middle keys as compared with when participants responded stationary. This is additional evidence in favor of action planning being the crucial stage of language interference. However, note that in the area of grounded language processing, there are findings of both language interfering with action (e.g. Glenberg & Kaschak, 2002) and language interfering with perception (e.g. Meteyard et al., 2008). Thus, future studies are needed to further investigate the differential role of those two modalities

and how those two modalities are integrated during language processing.

Previously, there has been a debate about whether words in isolation trigger the activation of experiential traces or simulations at all, or whether integration processes involved in sentence processing are required for finding compatibility effects between processes involving the modal systems (e.g., action, perception) and language comprehension. Bergen et al. (2007) showed that motion words resulted in interferences in sentences only when used literally (*The glass dropped*), not metaphorically (*The percentage dropped*). This finding has led to the conclusion that sentence integration processes are needed in order to find simulation effects. Given the strong activation of up-down locations just when seeing words in our present study, this suggests that individual words might always show automatic bottom-up activation of experiential traces, which are then subsequently suppressed or overwritten by sentence-based top-down processes (see also Raposo et al., 2009). Consistent with this view, Santana and de Vega (2011) demonstrated a congruency effect in metaphoric sentences when RTs were measured close to critical word onset (e.g., *rise*). However, the authors themselves suggested that their upward metaphors typically conveyed positive valence, and their downward metaphors negative valence. Thus, the observed effects might reflect a grounding of emotional valence on vertical space rather than word-based simulation effects.

It remains to be answered why studies that do not involve sentence contexts find a strong context dependency of word-based effects. For example, Van Dam et al. (2010) showed that context words play a crucial role for simulation effects observed in word processing. A word such as *telephone* implying an action toward the body facilitated responses only when presented in the context of a word that strengthened this aspect of the word's meaning (e.g., conversation). However, because the study did not include a condition in which the target words were presented without a context, it remains unclear whether the words alone triggered the corresponding experiential simulation. Thus, the association between the words and the proposed movement may simply not be strong enough to result in solid simulation effects in the absence of strengthening contexts. Or, it is possible that *telephone* simply triggers both a toward motion (picking it up) and an away motion (typing the numbers), and thus compatibility effects can be found only if a context is provided.

Taken together, our findings provide an important step supporting an early bottom-up activation account of experiential traces. Processing a verb such as *rise* or *fall* activates experiential traces stemming from experiencing events involving a rising or falling motion. In particular, lexical task demands are not a precondition

for their activation, which fits well with the view that activating experiential traces may be an integral part of lexical access. In previous studies, researchers have shown that words explicitly referring to an up versus a down location (e.g., *up*, *upwards*, *down*, *downwards)* affect responding even when presented subliminally (Ansorge, Kiefer, Khalid, Grassl, & Koenig, 2010). Our results would suggest that similar compatibility effects should be found with subliminally presented motion words relating to an upward or downward movement—a topic we will address in future studies. In conclusion, we suggest that in the future, a distinction needs to be made between top-down sentence-based simulations that follow meaning composition and bottom-up word-based simulation effects. Specifically, word-based effects are probably due to automatic activation of experiential traces, which can be subsequently suppressed (e.g., in case of negation) or strengthened when these traces are combined to yield simulations consistent with the meaning of the sentence or larger phrase.

References

Ansorge, U., Kiefer, M., Khalid, S., Grassl, S. I., & Koenig, P. (2010). Testing the theory of embodied cognition with subliminal words. *Cognition, 116,* 303–320.

Barsalou, L. W. (1999). Perceptions of perceptual symbols. *Behavioral and Brain Sciences, 22,* 637–660.

Bates, D., & Maechler, M., & Bolker, B. (2011). R Software Package lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-4. Retrieved from http://lme4.r-forge.r-project.org/

Bergen, B. K., Lindsay, S., Matlock, T., & Narayanan, S. (2007). Spatial and linguistic aspects of visual imagery in sentence comprehension. *Cognitive Science, 31,* 733–764.

Boulenger, V., Roy, A. C., Paulignan, Y., Deprez, V., Jeannerod, M., & Nazir, T. A. (2006). Cross-talk between language processes and overt motor behavior in the first 200 ms of processing. *Journal of Cognitive Neuroscience, 18,* 1607–1615.

Bub, D. N., & Masson, M. E. J. (2010). On the nature of hand-action representations evoked during written sentence comprehension. *Cognition, 116,* 394–408.

Bub, D. N., Masson, M. E. J., & Cree, G. S. (2008). Evocation of functional and volumetric gestural knowledge by objects and words. *Cognition, 106,* 27–58.

Buccino, G., Baumgaertner, A., Colle, L., Buechel, C., Rizzolatti, G., & Binkofski, F. (2007). The neural basis for understanding non-intended actions. *NeuroImage, 36,* 119–127.

Cho, Y. S., & Proctor, W. R. (2002). Influences of hand posture and hand position on compatibility effects for up-sown stimuli mapped to left–right responses: Evidence for a hand referent hypothesis. *Perception & Psychophysics, 64,* 1301–1315.

Costantini, M., Ambrosini, E., Scorolli, C., & Borghi, A. M. (2011). When objects are close to me: Affordances in the peripersoal space. *Psychonomic Bulletin & Review, 18,* 302–308.

Coventry, K. R., Lynott, D., Cangelosi, A., Monrouxe, L., Joyce, D., & Richardson, D. C. (2010). Spatial language, visual attention, and perceptual simulation. *Brain & Language, 112,* 202–213.

De Jong, R., Liang, C., & Lauber, E. (1994). Conditional and unconditional automaticity: A dual process model of effects of spatial stimulus–response correspondence. *Journal of Experiment Psychology: Human Perception and Performance, 20,* 731–750.

Desmurget, M., & Grafton, S. (2000). Forward modelling allows feedback control for fast reaching movements. *Trends in Cognitive Sciences, 4,* 423–431.

Fox, L. A., Shor, R. E., & Steinman, R. J. (1971). Semantic gradients and interference in naming color, spatial direction, and numerosity. *Journal of Experimental Psychology, 91,* 59–65.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review, 9,* 558–565.

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron, 41,* 303–307.

Kaup, B., Lüdtke, J., & Steiner, I. (in press). Word- vs. sentence-based simulation effects during language comprehension. In B. Stolterfoht & S. Featherston (Eds.), *Current work in linguistic evidence: The fourth Tübingen meeting*. Amsterdam, the Netherlands: de Gruyter.

Kornblum, S., Hasbroucq, T., & Osman, A. (1990). Dimensional overlap: Cognitive basis for stimulus-response compatibility - A model and taxonomy. *Psychological Review, 97,* 253–270.

Lachmair, M., Dudschig, C., De Filippis, M., de la Vega, I., & Kaup, B. (2011). Root versus roof: Automatic activation of location information during word processing. *Psychonomic Bulletin & Review, 18,* 1180–1188.

Loftus, G. R., & Masson M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review, 1,* 476–490.

Lu, C.-H., & Proctor, R. W. (1995). The influence of irrelevant location information on performance: A review of the Simon and spatial Stroop effects. *Psychonomic Bulletin & Review, 2,* 174–207.

Lu, C.-H., & Proctor, R. W. (2001). Influence of irrelevant information on human performance: Effects of S–R association strength and relative timing. *Quarterly Journal of Experimental Psychology, 51A,* 95–136.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109,* 163–203.

Mahon, B. Z., & Caramazza, C. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology—Paris, 102,* 59–70.

Meteyard, L., Zokaei, N., Bahrami, B., & Vigliocco, G. (2008). Visual motion interferes with lexical decision on motion words. *Current Biology, 18,* R732–R733.

Pratte, M. S., Rouder, J. N., Morey, R. D., & Feng, C. (2010). Exploring the differences in distributional properties between

Stroop and Simon effects using delta plots. *Attention, Perception, & Psychophysics, 72,* 2013–2025.

Raposo, A., Moss, H. E., Stamatakis, E. A., & Tyler, L. K. (2009). Modulation of motor and premotor cortices by action, action words and action sentences. *Neuropsychologia, 47,* 388–396.

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin, 86,* 446–461.

Santana, E., & de Vega, M. (2011). Metaphors are embodied, and so are their literal counterparts. *Frontiers in Psychology, 2,* 1–12.

Shor, R. E. (1970). The processing of conceptual information on spatial directions from pictoral and linguistic symbols. *Acta Psychologica, 32,* 346–365.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18,* 643–662.

Taylor, L. J., & Zwaan, R. A. (2008). Motor resonance and linguistic focus. *Quarterly Journal of Experimental Psychology, 61,* 896–904.

Turner, M. (1994). Design for a theory of meaning. In W. F. Overton & D. S. Palermo (Eds.), *The nature and ontogenesis of meaning* (pp. 91–107). Hillsdale, NJ: Erlbaum.

Urgesi, C., Moro, V., Candidi, M., & Aglioti, S. M. (2006). Mapping implied body actions in the human motor system. *Journal of Neuroscience, 26,* 7942–7949.

Van Dam, W. O., Rueschemeyer, S.-A., Lindemann, O., & Bekkering, H. (2010). Context effects in embodied lexical-semantic processing. *Frontiers in Psychology, 1,* 1–6.

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

Zwaan, R. A., & Taylor, L. (2006). Seeing, acting, understanding: Motor resonance in language comprehension. *Journal of Experimental Psychology: General, 135,* 1–11.