On the Embodied Representation of Words, Sentences, and Sentence Polarity

Dissertation
der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
Emanuel Schütt
aus Balingen

Tübingen
2025

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	11.12.2025
Dekan:	Prof. Dr. Thilo Stehle
1. Berichterstatter/-in:	Dr. Carolin Dudschig
2. Berichterstatter/-in:	Prof. Dr. Hartmut Leuthold

Contents

ABSTRACT (ENGLISH) 5

ABSTRACT (GERMAN) 7

OVERVIEW OF ARTICLES 9

1 INTRODUCTION 11

1.1 The traditional view of cognition and language comprehension 11

1.2 The embodied view of cognition and language comprehension 15

2 ISSUE 1: WORD-BASED SIMULATIONS OF LOCATION INFORMATION 18

2.1 Empirical background, motivation, and research objectives 18

2.2 Summary of the research articles 22

3 ISSUE 2: SENTENCE-BASED SIMULATIONS OF ENTITY SHAPES 27

3.1 Empirical background, motivation, and research objectives 27

3.2 Summary of the research article 31

4 ISSUE 3: EMBODIMENT OF SENTENCE POLARITY 35

4.1 Empirical background, motivation, and research objectives 35

4.2 Summary of the research article 37

5 GENERAL DISCUSSION 40

5.1 Context-dependence of embodiment effects 41

5.2 Sentence-based mental simulations 42

5.3 Functional relevance of mental simulations 43

5.4 Embodiment of sentence polarity 44

5.5 Methodological aspects 45

5.6 Conclusions and outlook 47

6 REFERENCES 48

APPENDIX A: ARTICLE 1 63

APPENDIX B: ARTICLE 2 80

APPENDIX C: ARTICLE 3 116

APPENDIX D: ARTICLE 4 131

Abstract (English)

In cognitive science, the question of how humans understand the meaning of language has been intrinsically linked to investigating the representational format of cognitive processes underlying this unique human ability. Traditional views hypothesizing that language comprehension operates on an abstract amodal code have been increasingly challenged by embodied accounts favoring concrete modal representational formats. The latter ones typically propose that we establish meaning through reactivating and combining sensorimotor experiences associated with the content of utterances that are processed – a mechanism which is also referred to as mental simulation. Even though an enormous amount of studies originated from this framework, important questions are far from being answered or have barely been investigated in an adequate manner. In this cumulative dissertation, some of these open issues are tackled. This includes methodological aspects, the nature of mental simulations created in response to both isolated words and complete sentences, and the embodiment of the abstract linguistic property of sentence polarity. First of all, I introduced and evaluated a new method that can replace actual vertical response movements. This provides researchers with a web-suited and easy to implement approach to investigating word-based reactivations of spatial experiences and other types of spatial associations. In further experiments, the focus was on contributing new insights into issues that are particularly relevant in the context of assessing to what extent human language comprehension might indeed rely on creating mental simulations. A series of studies revealed that implicit location words denoting entities typically located in lower or upper space (e.g., “worm” vs. “comet”) appear to evoke spatial simulations only if the experimental task involves the sensorimotor system in an adequate manner. My research also showed that language comprehenders tend to create mental simulations of sentential meaning merely at the end of sentences – possibly as a sort of sentential wrap-up effect after amodal meaning composition has taken place. These findings might suggest that language processing is not exclusively based on mental simulations, thus supporting the idea of so-called weak views of embodiment and hybrid models of cognition that acknowledge the role of both modal and amodal representational formats. It will be one core topic of future research to examine under which boundary conditions mental simulations are created and whether these are functionally relevant for language comprehension. Finally, I

conducted experiments to explore how the abstract linguistic property of sentence polarity could be represented in terms of sensorimotor experiences. The idea was that specific visual experiences related to negation or affirmation have evolved into non-verbal markers of sentence polarity carrying semantic meaning. Very first results from this promising new avenue of research indicated that this might not hold true for the so-called “not face” or red/green color cues. Since linguistic negation and affirmation regularly co-occur with various sensorimotor states induced by the use of non-verbal means of communication (e.g., head shake vs. head nod), future studies should test in a more systematic manner whether and in which language comprehension scenarios such experiences can contribute to the embodiment of sentence polarity.

Abstract (German)

In der Kognitionswissenschaft ist die Frage, wie Menschen die Bedeutung von Sprache erfassen, sehr eng mit der Untersuchung des Repräsentationsformats der kognitiven Prozesse verknüpft, die dieser einzigartigen menschlichen Fähigkeit zugrunde liegen. Die traditionelle Sichtweise, wonach Sprachverstehen auf einem abstrakten amodalen Code basiert, wird dabei in zunehmender Weise von Embodiment-Ansätzen herausgefordert, die konkrete modale Repräsentationsformate favorisieren. Typischerweise gehen diese davon aus, dass Bedeutung hergestellt wird, indem wir sensomotorische Erfahrungen reaktivieren und kombinieren, die mit dem Inhalt der zu verarbeitenden sprachlichen Stimuli assoziiert sind – ein Prozess, der auch als mentale Simulation bezeichnet wird. Zwar bildeten Embodiment-Ansätze den Ausgangspunkt für unzählige empirische Studien, allerdings sind wichtige Fragen weit davon entfernt, beantwortet zu sein oder wurden bisher kaum in angemessener Weise untersucht. Die vorliegende kumulative Dissertation stellt einen Versuch dar, einige dieser Punkte anzugehen. Das betrifft methodische Aspekte, Charakteristika von wort- und satzbasierten mentalen Simulationen sowie Ansätze zur mentalen Repräsentation des abstrakten sprachlichen Konzepts der Satzpolarität auf Grundlage von sensomotorischen Erfahrungen. Ich entwarf und evaluierte zunächst ein neues Paradigma, das es ermöglicht, in Studien Reaktionen in Form von vertikalen Bewegungen zu ersetzen. So steht nun ein leicht implementierbarer und für die web-basierte Datenerhebung geeigneter Ansatz zur Verfügung, der zur Untersuchung wort-basierter Reaktivierungen räumlicher Erfahrungen und anderer Arten von räumlichen Assoziationen eingesetzt werden kann. In weiteren Experimenten lag der Fokus dann darauf, neue Befunde zu generieren, die insbesondere für die Beantwortung der Frage relevant sind, inwieweit menschliches Sprachverstehen tatsächlich auf der Bildung mentaler Simulationen beruhen könnte. Hierbei zeigte sich, dass solche Wörter, die Entitäten beschreiben, die typischerweise im unteren oder oberen Bereich des Raumes zu finden sind (z.B. „Wurm“ vs. „Komet“), lediglich dann räumliche Simulationen zu evozieren scheinen, wenn der Aufgabenkontext das sensomotorische System in adäquater Weise einbezieht. Außerdem gelang es, aufzuzeigen, dass mentale Simulationen kompositorischer Art tendenziell erst zum Satzende gebildet werden – womöglich als Widerspiegelung des Inhalts („Wrap-Up-Effekt“) im Anschluss an eine amodale Art von Bedeutungsbildung. Eine Verarbeitung

von Sprache allein auf der Basis von mentalen Simulationen ist mit solchen Befunden schwer zu vereinbaren. Vielmehr liefern diese Anhaltspunkte zugunsten sogenannter schwacher Embodiment-Ansätze und hybrider Modelle, die gleichermaßen die Rolle modaler und amodaler Repräsentationsformate anerkennen. Eine der zentralen Aufgaben für zukünftige Forschung wird darin bestehen, genauer zu untersuchen, unter welchen Randbedingungen mentale Simulationen gebildet werden und ob diese eine funktionale Relevanz für das Sprachverstehens besitzen. Abschließend beschäftigte ich mich zudem mit der Frage, wie das abstrakte sprachliche Konzept der Satzpolarität in Form sensomotorischer Erfahrungen mental repräsentiert werden könnte. Die Idee bestand darin, dass sich mit Negation und Affirmation assoziierte konkrete visuelle Erfahrungen zu non-verbalen Markern der Satzpolarität entwickeln und entsprechend semantische Bedeutung tragen. Wie erste im Kontext dieses vielversprechenden neuen Forschungsansatzes erhaltene Resultate nahelegen, scheint dies möglicherweise nicht für das sogenannte „not face“ oder rote und grüne Farbreize zu gelten. Negation und Affirmation gehen allerdings regelmäßig mit weiteren sensomotorischen Erfahrungen einher. Diese werden beispielsweise durch non-verbale Formen der Kommunikation induziert (u.a. Kopfschütteln vs. Kopfnicken). Zukünftige Studien sollten systematisch prüfen, ob und in welchen spezifischen Settings derartige Erfahrungen tatsächlich zur mentalen Repräsentation von Satzpolarität herangezogen werden.

Overview of articles

This is a cumulative PhD thesis, which is based on four original research articles. For this reason, there are necessarily significant structural and content-related similarities and overlaps between the text of the articles and the thesis. Further information on the articles can be found below, including statements on the contributions of all co-authors (§ 6 Abs. 2 Satz 3 PromO). All articles are attached to the thesis.

Article No. 1 (Appendix A)

Schütt, E., Mackenzie, I. G., Kaup, B., & Dudschig, C. (2023). Replacing vertical actions by mouse movements: A web-suited paradigm for investigating vertical spatial associations. *Psychological Research*, *87*(1), 194-209. https://doi.org/10.1007/s00426-022-01650-6

Author Contributions (in %)

Author	Scientific Ideas	Data Generation	Analysis & Interpretation	Paper Writing
Schütt, E.	33	30	60	70
Mackenzie, I. G.	0	50	5	5
Kaup, B.	33	5	15	10
Dudschig, C.	33	15	20	15

Status in publication process: Published.

Article No. 2 (Appendix B)

Schütt, E., Kaup, B., & Dudschig, C. (2022). Investigating vertical language-space associations: Can visual action effects induce congruency effects? [Unpublished manuscript]. Department of Psychology, University of Tübingen.

Author Contributions (in %)

Author	Scientific Ideas	Data Generation	Analysis & Interpretation	Paper Writing
Schütt, E.	50	85	60	80
Kaup, B.	30	10	20	5
Dudschig, C.	20	5	20	15

Status in publication process: Unpublished manuscript.

Article No. 3 (Appendix C)

Schütt, E., Dudschig, C., Bergen, B. K., & Kaup, B. (2023). Sentence-based mental simulations: Evidence from behavioral experiments using garden-path sentences. *Memory & Cognition*, *51*(4), 952-965. https://doi.org/10.3758/s13421-022-01367-2

Author Contributions (in %)

Author	Scientific Ideas	Data Generation	Analysis & Interpretation	Paper Writing
Schütt, E.	30	65	50	70
Dudschig, C.	5	5	5	5
Bergen, B. K.	5	5	10	10
Kaup, B.	60	25	35	15

Status in publication process: Published.

Article No. 4 (Appendix D)

Schütt, E., Weicker, M., & Dudschig, C. (2024). Multimodal aspects of sentence comprehension: Do facial and color cues interact with processing negated and affirmative sentences? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *50*(6), 957-966. https://doi.org/10.1037/xlm0001302

Author Contributions (in %)

Author	Scientific Ideas	Data Generation	Analysis & Interpretation	Paper Writing
Schütt, E.	45	90	65	75
Weicker, M.	10	0	5	5
Dudschig, C.	45	10	30	20

Status in publication process: Published.

1 Introduction

There are various alternatives to approach the question of what exactly makes us human. Strategies and methods might highly depend on the research discipline. However, an answer to the question should most likely include referring to the complexity and the exceptional properties of the human language system – such as compositional language use and advanced vocal control – which cannot be found to the same extent in non-human species (for an overview, see Zuberbühler, 2015). When trying to explore the characteristics of human language comprehension, an important line of research in cognitive science has particularly focused on the format of mental representations that are created in response to linguistic input (e.g., written and spoken words, phrase, or sentences). On the one hand, there is the traditional view of cognition, which suggests that processes related to higher cognitive functions like language comprehension operate on abstract, amodal representations (e.g., Anderson, 1983; Chomsky, 1980; Fodor, 1975; Pylyshyn, 1984). On the other hand, the embodied or grounded view of cognition assumes that higher cognitive processes are based on representations that are rather concrete and modal (e.g., Barsalou, 1999; Bergen, 2012; Glenberg, 1997; Lakoff & Johnson, 1999; Zwaan & Madden, 2005). The latter view is sometimes referred to as neo-empiricism (see, for instance, Machery, 2006, 2007, 2016), pointing towards the fact that the controversy on the nature of mental representations has also become apparent in the diverging approaches to cognition discussed by philosophers of different epistemological views. As outlined by Kiefer and Pulvermüller (2012), empiricists such as Aristotle, Locke, and Hume believed that mental representations rely on sensorimotor (i.e., concrete and modal) impressions experienced during the interaction with the environment, whereas rationalists like Platon, Leibniz, and Kant preferred the idea that cognition works separated from any sensorimotor input. Nowadays, a growing number of researchers concerned with the human cognitive system favors so-called hybrid views, which acknowledge the role of both abstract, amodal and concrete, modal representations (e.g., Dove, 2009; Michel, 2021; Zwaan, 2014).

1.1 The traditional view of cognition and language comprehension

As a result of the cognitive revolution in the middle of the 20th century and influenced by new trends including computer science and artificial intelligence, early modern cognitive scientists started to develop ideas on the human representational system that build the core of what is usually called the traditional view of cognition (for reviews and comparable types of articles consulted to describe the traditional view of cognition within this paragraph, see Barsalou, 1999; Foglia & Wilson, 2013; Kiefer & Pulvermüller, 2012; Wilson, 2002). This approach to human cognition, which was decisively promoted by contributions of Jerry Fodor (1975) and Zenon Pylyshyn (1984), generally endorses the existence of distinct representational systems for perception, cognition, and action. The perceptual system serves to gather sensory experiences and the action system to perform behavioral actions. Crucially, the perceptual and the action system are expected to be functionally irrelevant in the context of realizing higher cognitive functions such as thinking, reasoning, or language comprehension. The cognitive system is often described by drawing comparisons with a computer. Specifically, it is assumed that cognitive processes operate on a language-like representational code, involving the manipulation and combination of abstract, amodal symbols. Sensory impressions captured by the perceptual system must be converted into this abstract representational code to be available for the cognitive system. This procedure inevitably causes a loss of modality-specific information associated with the original sensory input and its representation by the perceptual system. Accordingly, physical characteristics of an entity encountered in the environment (e.g., the coat color of a cat) should not be reflected in the representation of this entity in the cognitive system, resulting in a structurally arbitrary relationship between the sensory input and the corresponding representation created to perform cognitive processes. This also suggests that higher cognitive functions should be realized in highly specialized brain areas that are independent from those brain areas dealing with sensorimotor processes.

The traditional view of cognition critically shaped the way how cognitive scientists conceived the human cognitive system throughout the second half of the 20th century. Apart from influences by the rise of new technologies such as the computer, this might be related to specific beneficial features of such an approach to cognition. Michel (2021) proposed that there are two major advantages of a representational code that is based on abstract, amodal symbols. First, the free combination of such symbols allows to capture the systematicity and productivity of the human mind. Second, these symbols appear to be particularly suitable to represent abstract concepts (e.g., justice, democracy, hate, or love) that do not possess a perceivable physical referent in the world. However, the situation concerning empirical evidence in favor of the existence of abstract, amodal symbols is difficult. For instance, Barsalou (1999) strongly insisted that there is virtually no direct proof. Dove (2009), in turn, countered that evidence in cognitive science generally tends to be indirect and therefore recommended that we should focus on the question whether there is any evidence pointing towards the existence of abstract, amodal symbols. In this context, he reviewed findings of several behavioral and neuroscientific studies on number approximation that can be interpreted as indicating that at least this specific cognitive function might rely on an abstract, amodal representational code (see also Machery, 2007, 2016). To mention just one of the many quoted examples, there is the work by Barth et al. (2006) investigating arithmetic operations on non-symbolic numerosities. In one of their experiments, the authors asked participants to carry out numerical comparisons and addition tasks on sets of elements. The sets of elements were presented in both a unimodal condition (only visual arrays of dots) and a crossmodal condition (visual arrays of dots and sequences of tones). In the comparison task, participants judged which of two sets of elements (unimodal condition: a visual array of dots followed by another visual array of dots; crossmodal condition: a visual array of dots preceded or followed by a sequence of tones) included more elements. For the addition task, three sets of elements were displayed (unimodal condition: three visual arrays of dots; crossmodal condition: a visual array of dots and a sequence of tones followed by a visual array of dots or a sequence of tones). Participants were instructed to decide whether the third set of elements comprised more elements than the first two sets of elements taken together. Interestingly, the success in the unimodal and crossmodal conditions was comparable; thus, no across-modality performance costs were observed, suggesting that numerosities were represented via a common abstract, amodal representational platform. Further evidence in favor of the existence of abstract, amodal symbols was produced in studies with patients suffering from neurodegenerative diseases. For instance, as summarized by Kiefer and Pulvermüller (2012), semantic dementia – a condition associated with well-circumscribed lesions in the temporal lobe – causes general deficits in conceptual knowledge, including information from all modalities. Taken together, there is at least some empirical evidence supporting the idea that the human cognitive system or specific cognitive functions such as number approximation could rely on abstract, amodal symbols.

As in cognitive science in general, the traditional view of cognition also played a crucial role in theories and research on language comprehension. This is particularly exemplified by the proposal that the meaning of linguistic structures (e.g., sentences or texts) is mentally represented in terms of a set of amodal propositions (for a brief overview, see Kaup & Dudschig, 2017). According to Kintsch and van Dijk (1978), two of the most prominent exponents of such an approach to language comprehension, these propositions are built of concepts and need to comprise a predicate or relational concept and at least one argument. Arguments can be concepts or other propositions and serve to implement semantic functions (e.g., as agent, object, or goal). Predicates typically correspond to verbs, adjectives, adverbs, or sentence connectives. As suggested by the authors, one way to illustrate propositional representations are sorted lists that are constructed in line with the text sequence, which is realized for an example sentence in the following (capital letters are used to demonstrate that entries in the proposition list indicate concepts):

“Tom is eating broccoli, although he hates vegetables.”
Proposition N1: EAT (TOM, BROCCOLI)
Proposition N2: HATE (TOM, VEGETABLES)
Proposition N3: ALTHOUGH (N2, N1)

When referring to empirical evidence endorsing the creation of amodal propositional representations during language comprehension, the findings from a study by Kintsch and Keenan (1973) investigating the effect of the number of propositions on reading times are typically quoted. These authors presented participants with sentences that included an identical number of words but were constructed from different numbers of propositions. Indeed, reading times increased when additional propositions had to be processed, thus suggesting that the linguistic input was mentally represented in terms of amodal propositions. By contrast, Zwaan and Madden (2005) criticized that amodal propositions are barely suitable to capture physical properties of entities described in a sentence. To demonstrate this issue, they gave the following two sentences and respective propositional representations as examples:

“John pounded the nail into the wall.”
Proposition N1: POUND (JOHN, NAIL)
Proposition N2: IN (NAIL, WALL)

“John pounded the nail into the floor.”
Proposition N1: POUND (JOHN, NAIL)
Proposition N2: IN (NAIL, FLOOR)

These sentences basically differ with respect to the implied orientation of the nail (horizontal vs. vertical), but there is nothing in the propositional representation to account for this aspect. This strongly questions whether language comprehension can indeed exclusively operate on an amodal propositional representational code.

Generally, the traditional view of cognition has been associated with various problems and shortcomings. For instance – in addition to a fairly limited and shaky state of empirical evidence – there are neither accounts explaining adequately how impressions gathered by the sensory system could be transduced into abstract, amodal symbols nor findings indicating that such a process takes place at all (see, for instance, Barsalou, 1999). However, the probably most prominent and pressing issue is the symbol grounding problem (e.g., Harnad, 1990; Searle, 1980). This refers to the question of how abstract, amodal symbols with an arbitrary relationship to their physical referents in the world should be able to convey meaning. A straightforward solution to overcome the symbol grounding problem is to endorse the embodied view of cognition, which postulates the existence of concrete, modal symbols that show at least some similarities to the physical structure of their referents. Since the end of the 20th century, an ever-growing amount of behavioral and neuroscientific studies – also from the area of language comprehension research – has accumulated substantial empirical evidence supporting the idea that cognitive processes might indeed involve a representational code that is based on concrete, modal symbols (for overviews reviewing portions of this research, see, for instance, Barsalou, 2008; Barsalou et al., 2003; Dove, 2009; Fischer & Zwaan, 2008; Kaup et al., 2016; Kiefer & Pulvermüller, 2012; Meteyard et al., 2012), thus challenging the traditional view of cognition.

1.2 The embodied view of cognition and language comprehension

The embodied view of cognition has gained increasing attention in virtually all areas of cognitive science over the course of the last two to three decades (for reviews consulted to describe the embodied view of cognition within this paragraph, see Barsalou, 1999, 2008; Foglia & Wilson, 2013; Kiefer & Pulvermüller, 2012). In contrast to the traditional view of cognition, this approach favors the existence of a joint representational platform for perception, cognition, and action. Commonly, it is thought that higher cognitive functions such as thinking, problem solving, or language comprehension are realized by reactivating sensorimotor experiences stemming from interactions with the environment – a process, which is also referred to as mental simulation. Extracts of sensorimotor representations enriched by information from the different modalities thus build the functional foundation of human cognition. These representations are expected to show at least some structural similarities to the sensorimotor states that are evoked by interacting with their referents. On the neuroscientific level, higher cognitive functions should be closely linked to modal brain areas that are responsible for sensorimotor processing. The idea is that mentally simulating contents required to perform a cognitive task causes a partial re-enactment of brain activation patterns associated with experiencing relevant referents in the environment.

Embodied cognition accounts of human language comprehension focusing on the functional role of mental simulations (e.g., Barsalou, 1999; Bergen, 2012; Glenberg & Kaschak, 2002; Zwaan & Madden, 2005) typically propose that language comprehenders understand the meaning of a word by reactivating sensorimotor experiences that are associated with the word's referent. For instance, in the case of the word "chocolate", this should involve excerpts of sensorimotor states experienced by the comprehender while encountering different types of chocolate in the past – such as visual, olfactory, and gustatory impressions provoked by specific exemplars of chocolate. The reactivation of these sensorimotor experiences establish a mental simulation of the word's referent, which is thought to underly the comprehension process. The associations between single words and sensorimotor experiences related to their referents are assumed to originate from instances of co-occurrence. This idea is based on the observation that words are regularly used in the presence of their referents (e.g., when a product is labeled with its name or when a person located at a station points at a train while saying to another person "This is the train going to Berlin"). Such processes could be particularly important in the context of language acquisition during early childhood (e.g., a mother might hand her son a teddy bear and say "Here, your teddy bear"). Beyond the word level, language comprehenders are expected to combine word-based mental simulations to create mental simulations that are consistent with the meaning of larger linguistic units such as phrases, sentences, or paragraphs.

The number of behavioral and neuroscientific studies motivated by embodied cognition accounts of human language comprehension has grown tremendously during the last two decades, producing a vast amount of empirical findings. It is clearly not the aspiration and beyond the scope of this thesis to provide an extensive review. However, portions of results relevant to the articles that build the foundation of the thesis will be discussed in the following chapters, without making claims of completeness. The outlined research will therefore relate to the range of topics addressed by the experiments reported in this thesis. On the one hand, this includes the nature of mental simulations resulting from both isolated words and complete sentences – thus, linguistic units associated with fairly diverging affordances concerning meaning composition. On the other hand, this comprises potential grounding domains of sentence polarity (affirmation and negation), hence addressing the question of how more abstract linguistic properties and operators could be captured via sensorimotor experiences.

2 Issue 1: Word-based simulations of location information

One objective of my research was to develop and evaluate a paradigm that can replace vertical response movements in web-based experiments on simulation effects evoked by word-based stimuli conveying spatial meaning. In a further step, I asked whether such effects could also emerge from the pure anticipation of visual action effects related to lower or upper space. This chapter serves to review prior research, to describe the motivation for the experiments I performed, and to summarize findings. Similar (but more extensive) elaborations can be found in the corresponding articles:

Schütt, E., Mackenzie, I. G., Kaup, B., & Dudschig, C. (2023). Replacing vertical actions by mouse movements: A web-suited paradigm for investigating vertical spatial associations. *Psychological Research*, *87*(1), 194-209. https://doi.org/10.1007/s00426-022-01650-6

Schütt, E., Kaup, B., & Dudschig, C. (2022). Investigating vertical language-space associations: Can visual action effects induce congruency effects? [Unpublished manuscript]. Department of Psychology, University of Tübingen.

2.1 Empirical background, motivation, and research objectives

There are numerous studies suggesting that language comprehenders indeed tend to create word-based mental simulations. The investigation of vertical language-space associations has turned out to be a particularly fertile testbed in this context. An important line of research in this area deals with the question whether language comprehenders mentally simulate spatial experiences while encountering words denoting entities that are physically associated with lower or upper vertical space (*implicit location words*). Typically, experimental tasks from the visual domain are prone to produce spatial congruency effects in terms of both facilitation and interference (for a similar debate of this issue, see Pecher et al., 2010). For instance, Estes et al. (2008) presented participants with nouns whose referents are typically located in lower or upper vertical space (e.g., “foot” vs. “head”). These were followed by unrelated visual targets (the letters “X” and “O”) displayed at the bottom or top of the screen. Interestingly, the results revealed that there was an interference effect: Target discrimination was hampered when the spatial feature conveyed by the noun and the location of the visual target matched (e.g., when a word such as “bird” was followed by a visual target that appeared at the top of the screen). The authors suggested that processing implicit location words caused a shift of attention and evoked perceptual simulations in the vertical location their referents can usually be encountered. As the simulated entities and the targets highly differed in their visual properties, they argued that inhibiting the perceptual simulation was necessary to identify targets in the typical vertical location of the entity in question, hence producing longer response times (for further studies reporting similar interference effects, see Kaschak et al., 2005; Richardson et al., 2003). By contrast, Zwaan and Yaxley (2003) observed facilitation in the context of a semantic relatedness task. In each experimental trial, two object words describing entities with a canonical vertical relation (e.g., "basement" and "attic") were displayed concurrently one above the other on the screen. Judgments on the semantic relatedness of the words were significantly faster when the vertical arrangement of the words matched compared to mismatched the typical spatial relation of their referents (e.g., "attic" above "basement"). Similarly, in a semantic judgment task, participants responded faster to words denoting entities that are typically located in upper vertical space (e.g., "eagle"; "moon"; "helicopter") when they appeared in the upper part of the screen and faster to words denoting entities that are typically located in lower vertical space (e.g., "snail"; "carpet"; "abyss") when they appeared in the lower part of the screen (Šetić & Domijan, 2007). In addition, Dudschig et al. (2012a) could demonstrate that using implicit location words (e.g., "shoe" vs. "cloud") as task-irrelevant verbal cues can facilitate the mere detection of simple visual targets (i.e., boxes filled with white color) displayed in compatible vertical locations. Other studies, however, did not produce any spatial congruency effects. For instance, Pecher et al. (2010) presented words referring to entities usually encountered in the ocean (e.g., "shark") or the sky (e.g., "balloon") at the bottom or top of the screen. The task was to perform an ocean or a sky decision task. Crucially, the authors did not observe an interaction of spatial word meaning and word position, even though ocean decisions were faster for words presented at the bottom of the screen and sky decisions for words presented at the top of the screen. In sum, there is much variation in spatial congruency effects resulting from visual tasks. This is most probably due to varying task affordances (see also Pecher et al., 2010).

The situation concerning studies employing tasks from the motor domain is much clearer, as these have regularly produced spatial congruency effects in terms of facilitation. Much of this research is based on the vertical Stroop task. For example, in the study by Lachmair et al. (2011), implicit location words such as “submarine” or “comet” appeared in different font colors centered on the screen. Depending on the font color, participants performed a downward or upward arm movement on a vertical response device mounted in front of them. Thus – just as in the original Stroop task (Stroop, 1935) – access to the word meaning was not required. Nevertheless, the authors observed a spatial congruency effect: Response times were faster when the response direction was in line with the typical vertical location of the entity the presented implicit location word referred to. This strongly suggests that implicit location words are able to automatically reactivate spatial experiences associated with encountering their referents in the environment. Other studies using the identical or a highly similar vertical motor response task reliably replicated this spatial congruency effect (e.g., Dudschig et al., 2014, 2015; Öttl et al., 2017; Thornton et al., 2013; Vogt et al., 2019). In addition, the same pattern of results was obtained for further groups of words carrying spatial meaning, such as valence words denoting specific emotional states that are associated with a crouched or upright bodily posture (e.g., “depressed” vs. “excited”; Dudschig et al., 2015), direction-associated motion verbs (e.g., “to sink” vs. “to jump”; Dudschig et al., 2012b), or spatial prepositions (e.g., “below” vs. “above”; Ahlberg et al., 2018). Another line of research relying on tasks from the motor domain drew on saccadic eye movements. Dudschig et al. (2013) asked participants to perform lexical decisions on implicit location words (e.g., “sun” vs. “worm”) and non-words displayed centered on the screen. Responses were provided through eye movements towards the bottom or top of the screen. In accordance with the findings resulting from the Stroop-like motor response task involving vertical arm movements, a spatial congruency effect in terms of facilitation emerged. Saccadic eye movements in response to implicit location words were faster when the typical vertical location of the word’s referent and the direction of the eye movements matched (for comparable results, see Dunn et al., 2014). In conclusion, encountering words denoting entities or concepts that are physically associated with lower or upper vertical space can influence subsequent motor responses on the vertical axis in a reliable manner, thus indicating that these words indeed provoke mental simulations of spatial experiences.

A first article served to present a web-suited counterpart to the vertical Stroop task involving actual arm movements, which has been extensively used in lab-based research on vertical language-space associations (e.g., Ahlberg et al., 2018; Brookshire et al., 2010; Dudschig et al., 2014, 2015; Günther et al., 2018; Lachmair et al., 2011; Öttl et al., 2017; Thornton et al., 2013; Vogt et al., 2019). This required replacing vertical response movements by an alternative response mode that can be realized without specific response devices. Implicit location words and emotional valence words that are typically associated with vertical body postures served as a test bed. Critically, introducing such a web-suited paradigm appears to be useful due to several reasons. First, conducting cognitive research via the Internet has gained increasing importance over the course of the last years (e.g., Gosling & Mason, 2015; Stewart et al., 2017; Woods et al., 2015). This development was additionally accelerated by the recent coronavirus pandemic. Second, there is no existing straightforward way to convert the vertical Stroop task into a web-suited paradigm. Third and finally – aside from research on human language comprehension – many other fields in cognitive science such as spatial cognition, numerical cognition, or social cognition also have an interest in vertical spatial associations (e.g., Ito & Hatta, 2004; Koch et al., 2011; Schneider, 2020; Schubert, 2005; Schwarz & Keus, 2004) and could benefit from a reliable and easy to implement web-suited paradigm that can replace actual vertical response movements.

In a second article, I aimed to investigate whether spatial congruency effects typically observed with implication location words can emerge from the mere anticipation of visual action effects located in lower or upper vertical space. This research question was decisively motivated by the observation that many studies on word-based vertical language-space associations using tasks from the motor domain (e.g., the vertical Stroop task involving actual downward and upward arm movements) tend to show a confound that has apparently been neglected so far: Performing vertical response movements usually induces several sensory events related to vertical space. For instance, this includes seeing the arms moving downwards and upwards or hearing a click sound coming from lower or upper vertical space due to pressing the lower or upper response button. Interestingly, the relevance of such type of events has been highlighted by ideomotor theory (for reviews, see Badets et al., 2016; Hommel et al., 2001; Shin et al., 2010), which suggests that anticipating action effects is crucial to action planning. It is therefore a still to be resolved issue whether spatial congruency effects in the context of word-based vertical language-space associations might also originate from the pure anticipation of sensory action consequences located in lower or upper space (for a similar idea on the space-time congruency effect, which associates the mental represenation of time with horizontal space, see Janczyk & Ulrich, 2019).

2.2 Summary of the research articles

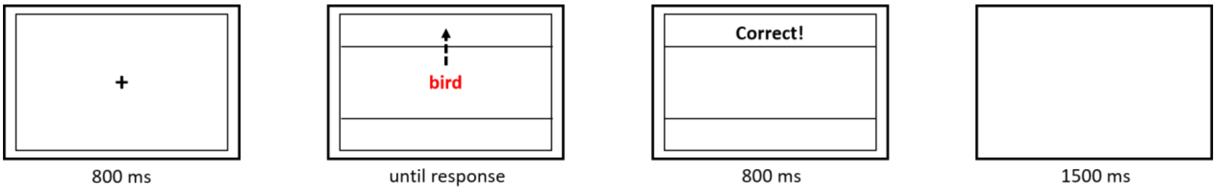
Schütt, E., Mackenzie, I. G., Kaup, B., & Dudschig, C. (2023). Replacing vertical actions by mouse movements: A web-suited paradigm for investigating vertical spatial associations. *Psychological Research*, *87*(1), 194-209. https://doi.org/10.1007/s00426-022-01650-6

This article introduced and empirically tested a web-suited counterpart to the manual vertical Stroop task, which has been widely used in lab-based research on vertical language-space associations (e.g., Ahlberg et al., 2018; Brookshire et al., 2010; Dudschig et al., 2014, 2015; Günther et al., 2018; Lachmair et al., 2011; Öttl et al., 2017; Thornton et al., 2013; Vogt et al., 2019). I replaced actual vertical response movements by mouse movements on the horizontal plane that induced vertical stimulus movements. Specifically, participants used their computer mouse to drag word stimuli appearing in the center of the screen to a lower or upper target area. The dragging direction was determined by the font color of the words (see Figure 1 for a more thorough illustration of the trial procedure). In accordance with prior lab-based research using the manual vertical Stroop task (e.g., Dudschig et al., 2015; Lachmair et al., 2011; Thornton et al., 2013), the most important dependent variable was an indicator of response selection and planning and defined as the time period from the appearance of the word stimulus on the screen until the initial stimulus movement occurred (response time) – this included clicking on the word and starting the mouse movement. Moreover, I performed additional analyses on the time period from the initial stimulus movement until the word entered one of the target areas (movement time). Participants were presented with implicit location words denoting entities usually located in lower or upper space (e.g., “submarine” vs. “ceiling”; Experiments 1 and 2) or emotional valence words related to a crouched or upright bodily posture (e.g., “joyful” vs. “depressed”; Experiment 3). Figure 2 gives an overview of the results. Most importantly, across all experiments, response times were significantly faster when the vertical association of the entity or the emotional state described by the word (up vs. down) matched compared to mismatched the dragging direction (up vs. down). Similarly, this effect was marginally significant (Experiment 1) or significant (Experiments 2 and 3) for movement times.

The web-suited paradigm produced spatial congruency effects comparable to those obtained with lab-based paradigms requiring participants to perform actual vertical response movements on a specific response device. I thus concluded that a direct real-time coupling of mouse movements on the horizontal plane and vertical stimulus movements on the screen is a reliable method to examine vertical spatial associations. Consequently, the mouse-based paradigm provides researchers from the different branches of cognitive science interested in spatial associations (e.g., language comprehension; numerical cognition; social cognition; spatial cognition) with an approach that converts vertical response movements into an easy to implement response mode, thus making web-based data collection feasible and specific input devices superfluous.

Figure 1

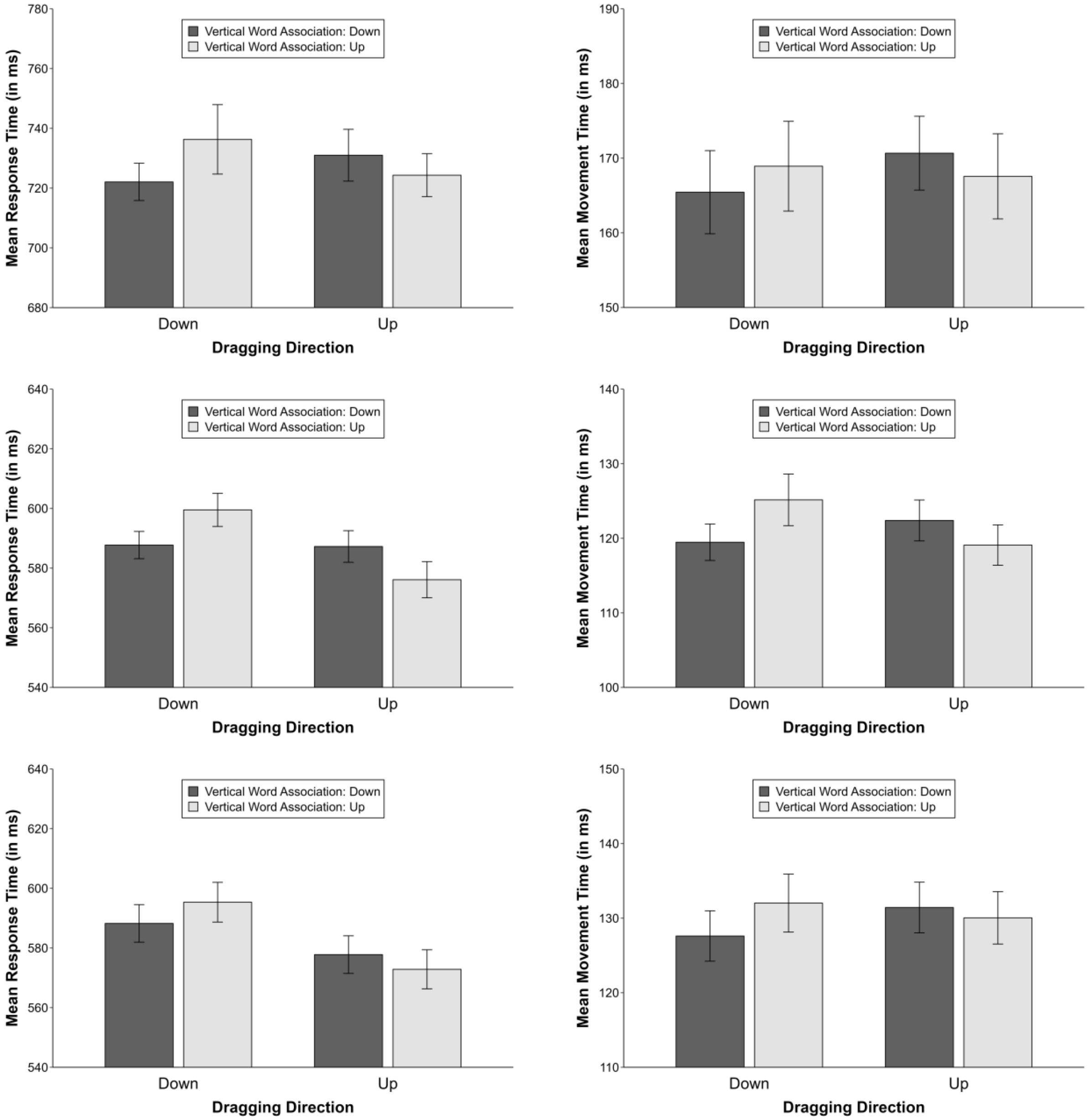
Trial procedure in Experiments 1 to 3



Note. Each trial started with the presentation of the fixation cross. Concurrently, a frame appeared that defined the experimentally relevant part of the screen. Then, a lower and an upper target area was introduced by displaying a borderline at the lower and upper end of the framed part of the screen. Moreover, the word stimulus replaced the fixation cross. The participants used their mouse to respond to the font color of the word stimulus: They dragged the stimulus either to the lower or the upper target area. Once the stimulus was completely located in one of the target areas, it was replaced by the feedback on response accuracy. Finally, the intertrial interval followed before the next trial started. This figure was reproduced with changes from “Replacing vertical actions by mouse movements: A web-suited paradigm for investigating vertical spatial associations”, by E. Schütt, I. G. Mackenzie, B. Kaup, and C. Dudschig, 2023, *Psychological Research*, *87*(1), p. 198. Copyright 2022 by the authors. The article, including all images, is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaption, distribution, and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if any changes were made. To view a copy of the license, please visit the website http://creativecommons.org/licenses/by/4.0/.

Figure 2

Response and movement times as a function of dragging direction and vertical word association in Experiment 1 (upper panel), Experiment 2 (middle panel), and Experiment 3 (lower panel)



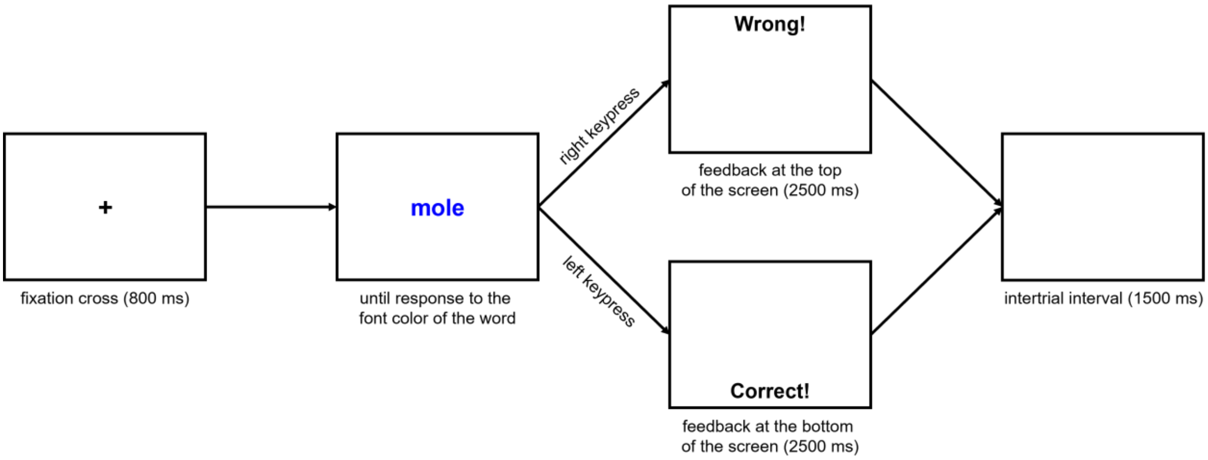
Note. Error bars denote 95% within-subjects confidence intervals calculated as recommended by Morey (2008). This figure was reproduced with changes from "Replacing vertical actions by mouse movements: A web-suited paradigm for investigating vertical spatial associations", by E. Schütt, I. G. Mackenzie, B. Kaup, and C. Dudschig, 2023, *Psychological Research*, *87*(1), pp. 199-203. Copyright 2022 by the authors. The article, including all images, is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaption, distribution, and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if any changes were made. To view a copy of the license, please visit the website http://creativecommons.org/licenses/by/4.0/.

Schütt, E., Kaup, B., & Dudschig, C. (2022). Investigating vertical language-space associations: Can visual action effects induce congruency effects? [Unpublished manuscript]. Department of Psychology, University of Tübingen.

In this research, I aimed to explore whether spatial congruency effects typically observed in the context of implicit location words (e.g., “grave” vs. “balloon”) can emerge from the mere anticipation of visual action effects related to lower or upper vertical space. Experiment 1 replicated the prior finding that the direct real-time coupling of mouse movements on the horizontal plane and vertical movements of implicit location words on the computer screen produces spatial congruency effects (see also Schütt et al., 2023). It might be that integrating anticipated visual word movements towards the lower or upper screen location into action planning lies at the heart of these effects. However, it is also possible that participants just re-coded and processed the mouse movements on the horizontal plane in terms of vertical response movements. Thus, to reveal whether and under which conditions visual action effects in vertical space could be able to evoke spatial congruency effects associated with implicit location words, I conducted further experiments. Participants responded to the font color of implicit location words displayed in the center of the screen by means of simple stationary keypresses, which caused visual action effects related to lower or upper vertical space. In Experiment 2, the response keys were located on the left-right axis of the keyboard. Visual feedback on response accuracy (“Correct!” vs. “Wrong!”) appeared always at the bottom of the screen when the left response key was pressed and at the top of the screen when the right response key was pressed (see Figure 3 for an overview of the trial procedure). There was no processing time advantage for congruent trials with a match of referent location and feedback location (e.g., when participants responded to words denoting entities typically encountered in an upper location by pressing the right response key that produced visual feedback on response accuracy at the top of the screen). The question arose whether the salience of the spatial dimension could critically affect the impact of anticipated visual action effects on the occurrence of the spatial congruency effects under investigation. Thus, in three follow-up experiments, contextual task conditions were adapted. I introduced response keys located on the front-back axis of the keyboard, as I hypothesized that the front-back axis should be more closely related to vertical space than the left-right axis. Visual feedback on response accuracy appeared always at the bottom of the screen when the response key on the front of the keyboard was pressed and at the top of the screen when the response key on the back of the keyboard was pressed. In Experiment 3a, the experimental setup produced a spatial congruency effect. Response times were faster when referent location and feedback location matched. However, this finding could not be replicated in Experiment 3b, suggesting that emphasizing the vertical spatial dimension in terms of introducing response keys located on the front-back axis of the keyboard does not suffice to evoke spatial congruency effects in a stable manner. In Experiment 4, I decided to replace the accuracy feedback by vertical stimulus movements, probably making the spatial dimension even more salient. The implicit location words moved in a stepwise manner to the bottom of the screen when the response key on the front of the keyboard was pressed and in a stepwise manner to the top of the screen when the response key on the back of the keyboard was pressed. Again, no spatial congruency effect emerged. These results demonstrate that – in contrast to the direct real-time coupling of mouse movements on the horizontal plane and vertical stimulus movements – associating non-vertical keypress responses with visual action effects in vertical space cannot reliably induce spatial congruency effects that are typically obtained for implicit location words. This indicates that these effects do not originate from the pure integration of anticipated spatially matching or mismatching action consequences into action planning. Rather, a more direct and distinct involvement of the motor system appears to be required, thus supporting the idea that processing implicit location words is deeply grounded in the sensorimotor system.

Figure 3

Trial procedure in Experiment 2



3 Issue 2: Sentence-based simulations of entity shapes

Going one step further, I also aimed to tackle the question whether language comprehenders create compositional mental simulations of sentential meaning and if so, whether these are built incrementally over the course of sentences or globally at the end of sentences. This chapter serves to review previous studies, to describe the motivation for the research, and to summarize findings. Very similar (but more extensive) elaborations can be found in the corresponding article:

Schütt, E., Dudschig, C., Bergen, B. K., & Kaup, B. (2023). Sentence-based mental simulations: Evidence from behavioral experiments using garden-path sentences. *Memory & Cognition*, *51*(4), 952-965. https://doi.org/10.3758/s13421-022-01367-2

3.1 Empirical background, motivation, and research objectives

Words rarely occur in an isolated manner. Rather, human language users combine single words to form phrases and sentences. Thus, if concrete, modal representations indeed play a functional role in language comprehension, comprehenders should definitely create mental simulations reflecting sentential meaning – even though their existence would not automatically imply their functional relevance. In contrast to the situation concerning word-based mental simulations, empirical findings in favor of sentence-based simulations are still sparse. Some evidence comes from a study by Bergen et al. (2007), which used sentences describing events that are literally or metaphorically related to lower or upper space (e.g., “The patient rose” vs. “The amount rose”). While listening to these sentences, participants categorized shapes presented in the lower or upper part of the screen as squares or circles. Interestingly, spatial congruency effects in terms of interference emerged for literal but not for metaphorical language. The mere appearance of a single word associated with lower or upper space (e.g., the word “rose” in the sentence “The amount rose”) did not evoke an interference effect, suggesting that mental simulations of vertical space were modulated by sentential meaning aspects beyond the word level. Furthermore, some studies on the action-sentence compatibility effect (ACE) indicate that modifying pure linguistic properties of sentences can influence simulation effects. For example, Bergen and Wheeler (2010) showed that sensibility judgments on progressive sentences referring to concrete sensorimotor events (e.g., “Ashley is stretching her arms”) were faster when the direction of the response movement was in accordance with the direction of the movement described in the sentence (e.g., response movement away from the body for a sentence such as “Ashley is stretching her arms”), whereas such a spatial congruency effect did not occur for perfect sentences (e.g., “Ashley has stretched her arms”). In another study (Taylor & Zwaan, 2008) on the rotational ACE, participants turned a knob device clockwise or counterclockwise to read sentences describing clockwise or counterclockwise manual rotations frame by frame in a self-paced manner. The ACE persisted when a postverbal adverb kept the linguistic focus on the action (e.g., “The cook/walked/over to/the oven/which he/turned down/*slowly*”) but stopped to emerge when a postverbal adverb shifted the linguistic focus to the agent of the sentence (e.g., “The cook/walked/ over to/the oven/which he/turned down/*happily*”). In sum, this suggests that language comprehenders are able to use information on linguistic properties to adapt mental simulations in such a way that these reflect meaning aspects beyond the word level.

There are many further studies that used sentential materials and provided empirical findings pointing towards the creation of mental simulations during language comprehension. However, most commonly, it is barely possible to judge whether reported effects emerged from sentence-based or word-based simulations. For instance, in a recent study by Hauf et al. (2020), children and adults listened to sentences implying downward or upward motion (e.g., “The rocket takes off into space” vs. “The treasure chest sinks to the seabed”). Each sentence was followed by a picture of an object that appeared at the top or bottom of the screen and moved downwards or upwards to the center of the screen. Participants performed a sentence-picture verification task: They decided whether the object shown on the picture had been mentioned in the previous sentence or not. The results revealed that both children and adults responded faster when the direction of the motion implied by the sentence matched the direction of the picture movement. Nevertheless, it is not clear whether these effects resulted from mental simulations of compositional meaning beyond the word level. It is equally possible that mere associations with single words referring to entities or concepts related to upward or downward motion (e.g., the noun “rocket” in the sentence “The rocket takes off into space” or the verb “to sink” in the sentence “The treasure chest sinks to the seabed”) produced the simulation effects. The same issue occurs for the highly cited study by Zwaan et al. (2002) that investigated the question whether language comprehenders mentally simulate the shape of entities described in sentences.

In the context of a sentence-picture verification task, participants were presented with sentences about concrete entities. Critically, the shape of these entities varied depending on their location. For instance, the sentence "The ranger saw the eagle in the sky" indicated that the wings of the eagle were outstretched, whereas the sentence "The ranger saw the eagle in its nest" indicated that the wings of the eagle were folded. Responses times to picture probes were faster when the depicted entity shape matched the shape implied by the sentential meaning. However, these simulation effects could again result from lexical associations. Encountering the words "eagle" and "sky" could have prompted participants to mentally simulate an eagle with outstretched wings. Similarly, the words "eagle" and "nest" might have provoked mental simulations of an eagle with folded wings. This type of confound also applies to many other studies on embodied language comprehension, including the seminal experiments of Glenberg and Kaschak (2002) on the ACE and well-known research on the mental simulation of specific manual actions during sentence processing (e.g., Bub & Masson, 2010; Masson et al., 2013; Masson, Bub, & Newton-Taylor, 2008; Masson, Bub, & Warren, 2008). In a considerable number of cases, simulation effects in response to sentential materials cannot be ascribed unequivocally to compositional mental simulations of sentential meaning – which is the reason why robust evidence in favor of sentence-based mental simulations is still fairly limited. I therefore aimed to conduct experiments that are more appropriate to reveal whether language comprehenders indeed create mental simulations beyond the word level. For this purpose, I adapted the common sentence-picture verification framework (e.g., Zwaan et al., 2002) and designed sentential materials that allowed disentangling whether simulation effects resulted from independent word associations or sentence-based mental simulations.

If language comprehenders indeed create compositional mental simulations of sentential meaning, the question of time course arises. On the one hand, it is conceivable that such simulations evolve in an incremental manner and are thus updated as soon as further relevant linguistic input is encountered over the course of a sentence. On the other hand, it might be that simulations are created globally at the end of sentences after amodal meaning composition has taken place, hence reflecting a sort of sentential wrap-up effect. Studies exploring the temporal dynamics of mental simulations during sentence comprehension are really scarce. For instance, it was shown that the rotational ACE occurs as a function of linguistic input. When participants turned a knob device clockwise or counterclockwise to read sentences on clockwise or counterclockwise manual rotations frame by frame in a self-paced manner (e.g., “Craving/a juicy/pickle/he took/the jar/off the/shelf and/opened/the/jar”), motor resonance effects emerged only for the verb region that specified the rotation direction (e.g., “opened”; Zwaan & Taylor, 2006). However, these effects extended to a postverbal adverb if it served to modify the manual rotation (such as “quickly” in the sentence “A fan/handed him/a bottle/of cold/water/which he/opened/quickly”; Taylor & Zwaan, 2008). This indicates that motor simulations are dynamically adapted in accordance with incoming linguistic information to represent the current state of affairs. Another study by Sato et al. (2013) was interested in exploring whether language comprehenders activate specific object shape representations early during sentence processing. The authors used the typical verb-final word order of the Japanese language and hypothesized that native speakers might have concrete expectations about the shape of an object mentioned in a sentence even before processing the verb at the end of the sentence. For instance, an item paraphrased as “Mother put the shirt neatly in the drawer” was arranged in the specific Japanese word order: “Mother-NOM shirt-ACC drawer-LOC neatly put”. Importantly, reading the preverbal arguments of this sentence (i.e., “mother”-NOM, “shirt”-ACC, and “drawer”-LOC) should be sufficient to conclude that the shape of the shirt was folded. Participants read the sentences word by word in a self-paced manner. Prior to the verb, they were presented with a picture probe. The task was to decide whether the depicted object had been mentioned in the sentence or not. Picture verification times were faster when the specific object shape suggested by the preverbal arguments matched the object shape shown on the picture. This might imply that language comprehenders generated detailed mental simulations of object shapes early during sentence comprehension. However, it is also possible that the picture probes provoked such simulations, enabling participants to perform the task – a potential limitation that was acknowledged by the authors themselves. In sum, research on the time course of mental simulations during sentence comprehension is barely available and appears not to permit deciding whether language comprehenders update mental simulations in an incremental manner. I therefore also presented transitionally ambiguous garden-path sentences in the context of the sentence-picture verification paradigm, which should prompt participants to form an initial and a sentence-based entity shape interpretation. This made it possible to explore whether comprehenders tend to create and retain mental simulations of entity shape interpretations activated early during sentence comprehension.

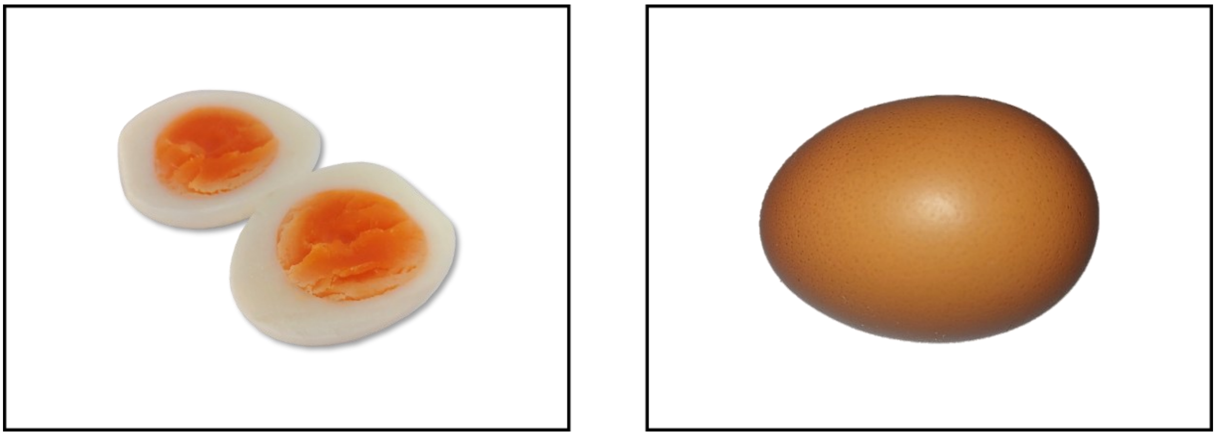
3.2 Summary of the research article

Schütt, E., Dudschig, C., Bergen, B. K., & Kaup, B. (2023). Sentence-based mental simulations: Evidence from behavioral experiments using garden-path sentences. *Memory & Cognition*, *51*(4), 952-965. https://doi.org/10.3758/s13421-022-01367-2

This research attempted to tackle the question whether mental simulations generated in response to sentences result from independent word associations or compositional processes beyond the word level. Additionally, I also investigated the time course of mental simulations: Are they created incrementally over the course of a sentence or globally at the end of the sentence as a sort of sentential wrap-up effect? In two experiments, participants were presented with a sentence-picture verification task. Thus, in each trial, they read a sentence before a picture probe appeared. The task was to decide as fast and as accurately as possible whether the entity shown on the picture had been mentioned in the sentence or not. Most importantly, two different types of sentences were shown. There were unambiguous sentences including words that are related to different shapes of the target entity described in the sentence. For instance, the sentence “The egg was in the fridge as Mary ate” comprised the word “egg” denoting the target entity and the words “fridge” and “ate” that are associated with different shapes an egg can take on: “fridge” with an intact egg in its shell and “ate” with a peeled egg that is ready for consumption (see Figure 4 for corresponding picture probes). However, only one of these shapes – the intact egg in its shell – complied with the sentential meaning. If language comprehenders indeed build compositional mental simulations of sentential meaning, responses to picture probes compatible with the sentence-based entity interpretation (here: the intact egg in its shell) should be faster than responses to picture probes incompatible with the sentence-based entity interpretation (here: the peeled egg). As the sentences always included the same number of words consistent with each of the depicted entity shape interpretations, such an effect should not occur if mental simulations result from word-based associations alone. Furthermore, there were transitionally ambiguous garden-path sentences, which had the same meaning and words as the respective unambiguous sentences (e.g., “As Mary ate the egg was in the fridge"). In an early step of the comprehension process, the first verb of these garden-path sentences is typically read as being transitive, provoking an initial interpretation of the entity shape (here: the peeled egg). When arriving at the second verb, language comprehenders realize that the first verb must be read as being intransitive. This leads to reanalyzing the sentence and establishing the final sentence-based interpretation of the entity shape (here: the intact egg in its shell). Previous research demonstrated that the initial entity interpretation generated during garden-path processing tends to linger even after the sentence has been reanalyzed (e.g., Christianson et al., 2001; Patson et al., 2009; Slattery et al., 2013). Consequently, at the end of garden-path sentences, representations of both the initial and the sentence-based entity interpretation should be available. By contrast, the situation was different for unambiguous sentences. These were exclusively associated with the sentence-based entity interpretation (here: the intact egg in its shell). Thus, if sentence-level simulations evolve in an incremental manner, the sentence-picture compatibility effect was expected to be smaller for garden-path sentences than for unambiguous sentences. Figure 5 summarizes the findings. In Experiment 1, sentence-picture verification times were faster when the picture probe was compatible with the sentence-based entity interpretation, indicating that participants created compositional mental simulations of sentential meaning. This effect was not modulated by sentence type, hence suggesting that mental simulations were built globally at the end of sentences. Experiment 2 replicated this pattern of results, even though unambiguous sentences were modified in such a way that the sentence-based entity interpretation was always associated with the opposite shape and picture. For instance, the sentence "The egg was in the fridge as Mary ate" (sentence-based entity interpretation: the intact egg in its shell) was replaced by the sentence "As Mary ate the egg the butter was in the fridge" (sentence-based entity interpretation: the peeled egg). This allowed ruling out that the results generated in Experiment 1 were due to participants preferring the picture probes showing the sentence-based entity interpretation or specific shapes being more consistent with the aggregate lexical associations of the sentences. In sum, participants tended to produce mental simulations that could not originate from word-based associations alone. I therefore concluded that language comprehenders combine reactivated sensorimotor experiences to build complex mental simulations of sentential meaning. However, these appear to operate over the sentence as a whole – there was not any evidence pointing towards the creation of incremental simulations during garden-path processing, although previous research revealed that encountering such sentences induces an intermediate entity interpretation (e.g., Christianson et al., 2001; Patson et al., 2009; Slattery et al., 2013). This could imply that mental simulations are not a necessary prerequisite of sentence comprehension. Rather, they might be created in the context of a sentential wrap-up effect, possibly after amodal meaning composition has taken place. Further research using methods permitting a direct look at on-line processes is needed to validate these findings and exclude the possibility that mental simulations of early entity interpretations were built but not sufficiently retained until the critical picture probe appeared.

Figure 4

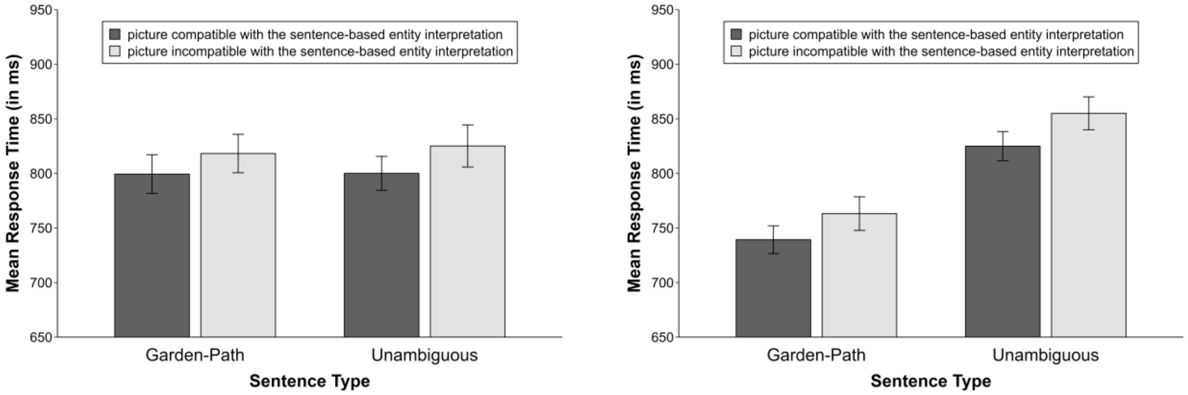
Example of picture probes used in Experiments 1 and 2



Note. These picture probes referred to the sentence pair including the garden-path sentence “As Mary ate the egg was in the fridge” and the unambiguous sentence “The egg was in the fridge as Mary ate” (Experiment 1) or “As Mary ate the egg the butter was in the fridge” (Experiment 2), respectively. The peeled egg (picture probe on the left) should show the initial entity interpretation during garden-path processing, whereas the intact egg in its shell (picture probe on the right) illustrated the final sentence-based entity interpretation. For unambiguous sentences, I expected participants to create only a single entity interpretation corresponding to the sentence-based entity interpretation (Experiment 1: intact egg in its shell; Experiment 2: peeled egg in its shell). This figure was reproduced with changes from “Sentence-based mental simulations: Evidence from behavioral experiments using garden-path sentences”, by E. Schütt, C. Dudschig, B. K. Bergen, and B. Kaup, 2023, *Memory & Cognition*, *51*(4), p. 957. Copyright 2022 by the authors. The article, including all images, is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaption, distribution, and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if any changes were made. To view a copy of the license, please visit the website http://creativecommons.org/licenses/by/4.0/.

Figure 5

Mean response times for the sentence-picture verification task as a function of sentence type and picture type in Experiment 1 (on the left) and Experiment 2 (on the right)



Note. Error bars denote 95% within-subjects confidence intervals calculated as recommended by Morey (2008). This figure was reproduced with changes from "Sentence-based mental simulations: Evidence from behavioral experiments using garden-path sentences", by E. Schütt, C. Dudschig, B. K. Bergen, and B. Kaup, 2023, *Memory & Cognition*, *51*(4), pp. 958-960. Copyright 2022 by the authors. The article, including all images, is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaption, distribution, and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if any changes were made. To view a copy of the license, please visit the website http://creativecommons.org/licenses/by/4.0/.

4 Issue 3: Embodiment of sentence polarity

Finally, I was interested in approaching and tackling the question of how sentence polarity – an abstract linguistic property without a straightforward physical referent in the world – could be represented through sensorimotor experiences. Specifically, I explored whether certain facial expressions and colors facilitate the comprehension of negated and affirmative sentences and might thus contribute to the embodied meaning of sentential negation and affirmation. This chapter serves to review previous studies, to describe the motivation for my research, and to summarize findings. Related elaborations can be found in the corresponding article:

Schütt, E., Weicker, M., & Dudschig, C. (2024). Multimodal aspects of sentence comprehension: Do facial and color cues interact with processing negated and affirmative sentences? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *50*(6), 957-966. https://doi.org/10.1037/xlm0001302

4.1 Empirical background, motivation, and research objectives

Research on embodied language comprehension has focused on language describing entities, situations, and events that can be experienced directly while interacting with the environment. In these cases, it is indeed clearly conceivable that language comprehenders might be able to draw on sensorimotor traces to create mental simulations reflecting the meaning of words, phrases, and sentences. The situation becomes much more challenging for abstract language referring to concepts, contents, and ideas that are not associated with concrete physical referents in the world (e.g., time; justice; love). Nevertheless, there have been several approaches to tackle this issue within the embodied cognition framework (for a comprehensive review, see Borghi et al., 2017). For instance, it was suggested that abstract language could be embodied via metaphorical mappings onto experiential domains such as space (e.g., Lakoff & Johnson, 1980, 1999), social, event, and introspective aspects of situations (e.g., Barsalou & Wiemer-Hastings, 2005), or affective experiences (e.g., Kousta et al., 2011; Vigliocco et al., 2014).

By contrast, there is barely any research on the question of how abstract linguistic properties might be captured by mental simulations of sensorimotor experiences. This also applies to sentence polarity – particularly, negation has usually been investigated in terms of an abstract verbal operator changing the truth-value of a proposition.

Interestingly, a recent attempt to gain insight into the embodiment of sentential negation produced initial evidence indicating that sentential negation processing reuses neurophysiological mechanisms of inhibition and cognitive control (e.g., Beltrán et al., 2018, 2019, 2021; de Vega et al., 2016; Liu et al., 2020). I aimed to identify further potential grounding domains of sentence polarity. This was addressed by referring to non-verbal markers of negation and affirmation that can be experienced through the sensorimotor system: the so-called "not" face and color information (red vs. green).

Apart from using verbal linguistic operators such as "not" or "no", humans can convey negation in a non-verbal manner. In a recent study, Benitez-Quiroz et al. (2016) explored whether articulating negation is associated with a distinct subset of facial expressions. They instructed participants of different cultural and ethnic backgrounds to show facial expressions of negation. This revealed that facial signs of negative moral judgment, including anger, disgust, and contempt, were typically produced. The authors concluded that these facial expressions have been combined to serve as the "not" face – a universal and unique marker of negation in human communication. In a further step, they provided additional evidence in favor of this hypothesis: Participants were much more likely to express the "not" face during reproducing negated sentences than during reproducing affirmative sentences, demonstrating that the "not" face is specifically co-articulated in the context of negated utterances. I wondered whether the "not" face might be part of the embodied meaning of sentential negation. Although this is speculative, the "not" face could constitute a sensorimotor precursor of higher-level verbal negation. For instance, it is well-known that children are able to show rejection by means of specific facial expressions even before they are familiar with the usage of verbal linguistic operators such as "not", "nor", or "none" (Dimroth, 2010). Moreover – given that the "not" face is indeed routinely produced during negated utterances – we should experience this non-verbal marker of negation regularly in co-occurrence with verbal negation, potentially establishing a close relationship between the "not" face and the mental representation of negation. In my research, I therefore investigated whether visual primes in terms of the "not" face are integrated during sentential negation comprehension.

Another potential grounding domain of sentence polarity appears to be visual color information. On the one hand, numerous experiences of red color are associated with denial and action inhibition. For instance, traffic signals flashing red suggest not to cross an intersection and thus to refrain from moving the car. On the other hand, several experiences of green color are closely related to approval and action initiation. For example, in some trains, green-colored signals imply that passengers can press the button that opens the doors. It is thus conceivable that red and green color evolved into non-verbal markers of negation and affirmation. Interestingly, a recent study by Dudschig et al. (2023) investigated the influence of red and green color primes on performing "yes" and "no" responses. In a lexical decision task, participants were presented with words and non-words (e.g., "garage" vs. "knamet"). The response key assignment was conveyed employing two response button symbols displayed next to each other on the screen. In each trial, one of the symbols was green-colored, whereas the other one was red-colored. The mapping of response labels ("yes" vs. "no") and colors onto response button symbols was randomized trial by trial. Importantly, "yes" responses were faster when the respective response button symbol was green-colored and "no" responses were faster when the respective response button symbol was red-colored. This indicates that giving "yes" responses is related to green color and giving "no" responses to red color. In a series of studies, I tested whether this pattern of results extends to the comprehension of negated and affirmative sentences, thus clarifying the potential role of red and green color in the embodied meaning representation of sentential negation and affirmation.

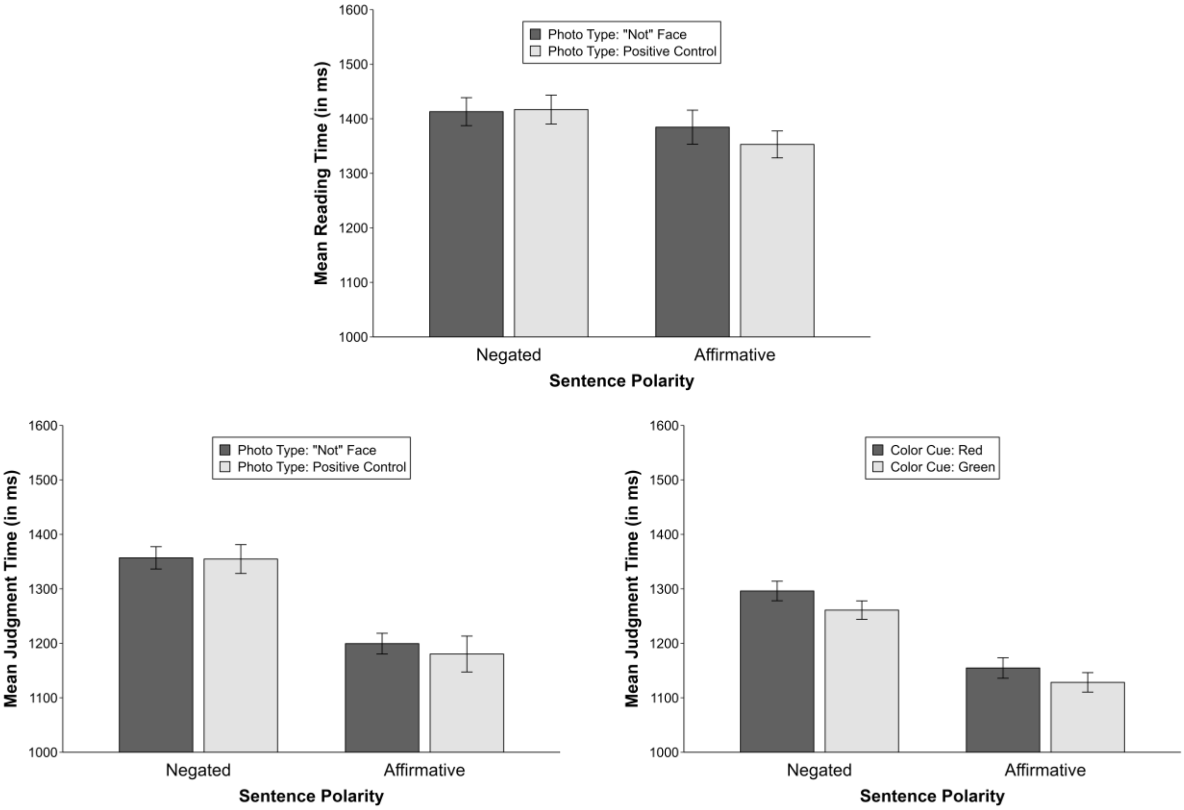
4.2 Summary of the research article

Schütt, E., Weicker, M., & Dudschig, C. (2024). Multimodal aspects of sentence comprehension: Do facial and color cues interact with processing negated and affirmative sentences? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *50*(6), 957-966. https://doi.org/10.1037/xlm0001302

In three pre-registered experiments, I explored whether specific non-verbal markers of negation and affirmation might contribute to the embodiment of sentential negation and affirmation. I referred to the "not" face, which was identified to be a regular co-articulator of verbal negation (Benitez-Quiroz et al., 2016), and to red and green color information, which is usually used to indicate denial and approval and was found to facilitate "yes" and "no" responses (Dudschig et al., 2023). If visual cues in terms of the "not" face or red and green color information are indeed integrated during sentential negation and affirmation comprehension, priming effects are expected to emerge. In Experiment 1, participants were presented with instances of the "not" face and positive controls. Shortly after photo onset, a negated or affirmative sentence (e.g., "No, I do not want to sing" vs. "Yes, I would like to buy a sofa") appeared right below the photo. I asked participants to press the space bar as soon as they had read and understood the sentence properly. Moreover, one fourth of the sentences was followed by a comprehension question (e.g., "Does the person want to take photos?" or "Is the statement of the person about hiking?"). Figure 6 provides an overview of the results. Both frequentist statistics and Bayes factors from linear mixed effects analyses revealed that there was no significant interaction: Reading times for negated and affirmative sentences were not differently modulated as a function of photo type. However, participants were not required to give a response related to the sentential meaning when they pressed the space bar to indicate that they had finished reading. It therefore remained uncertain whether sentence comprehension was already completed, which is why the self-paced reading task was replaced with a sensibility judgment task in follow-up experiments. Participants were primed with instances of the "not" face and positive controls (Experiment 2) or red and green color patches (Experiment 3). In addition to sensible negated and affirmative sentences, there were non-sensible negated and affirmative filler sentences showing semantic violations (e.g., "Yes, I want to solace a light bulb"; "Yes, I want to be peeled now"; "No, I would not like to teach a wall socket"; "No, I do not want to juggle with arteries"). The task was to judge as fast and as accurately as possible whether the displayed sentence was sensible or not. In line with Experiment 1, no significant interaction effects emerged. Thus, neither the "not" face nor red and green color cues were integrated during sentential negation and affirmation comprehension, although previous research using similar presentation modes demonstrated that extralinguistic information – for instance, on the gender identity of speakers (Rück et al., 2017) or real-world surface materials (Dudschig et al., 2021) – can be incorporated quickly during sentence comprehension. This indicates that the influence of the investigated non-verbal markers of negation and affirmation does not extend to the embodied meaning representation of sentential negation and affirmation. It is even conceivable that relevant cognitive processes generally operate in a more abstract, amodal representational format. However, contextual task affordances could also play a crucial role. For instance, the “not” face might accelerate sentential negation comprehension in situations where language comprehenders are used to integrate facial cues (e.g., during a face-to-face conversation). As the “not” face involves facial expressions of contempt, anger, and disgust, effects could also particularly occur for negated sentences implying such emotions (e.g., “I do not like eating broccoli”). Similarly, red and green color cues might modulate the comprehension of sentences denying or approving concrete motor actions (e.g., “Do not touch my car!” or “Enter the room, please!”). Further research is needed to shed some light on these issues and to evaluate the role of other non-verbal makers of negation and affirmation (e.g., head shake vs. head nod).

Figure 6

Reading times as a function of sentence polarity and photo type in Experiment 1 (upper panel), sensibility judgment times as a function of sentence polarity and photo type in Experiment 2 (on the left of the lower panel), and sensibility judgment times as a function of sentence polarity and color cue in Experiment 3 (on the right of the lower panel)



Note. Error bars denote 95% within-subjects confidence intervals calculated as recommended by Morey (2008).

5 General discussion

The research outlined in this dissertation was dedicated to investigating issues related to the question to what extent language comprehenders create embodied mental simulations associated with the content of linguistic structures and properties. This included behavioral experiments on the emergence and nature of embodied mental simulations in response to linguistic structures of highly different complexity – single words and full sentences. Both of them described concrete entities that can be experienced directly by means of the sensorimotor system while interacting with the environment. At the word level, I largely focused on vertical language-space associations in the context of implicit location words whose referents are typically encountered in lower or upper vertical space (e.g., “moon” vs. “mole”). Language comprehenders processing such words are expected to mentally simulate spatial meaning aspects, which should be reflected in terms of spatial congruency effects. A considerable amount of studies in this research area has been based on specific versions of the Stroop task that require participants to perform vertical arm movements on custom-built response devices (e.g., Dudschig et al., 2014, 2015; Lachmair et al., 2011; Öttl et al., 2017; Thornton et al., 2013; Vogt et al., 2019). I developed and evaluated an easy to implement, web-suited counterpart to this type of task by replacing actual vertical response movements with a direct real-time coupling of mouse movements on the horizontal plane and vertical stimulus movements on the screen. Furthermore, I conducted a series of several experiments to explore whether spatial congruency effects usually observed with implicit location words can emerge from the mere anticipation of spatially matching and mismatching visual action effects. The results revealed that associating non-vertical stationary keypress responses with visual action effects related to lower or upper space does not reliably produce spatial congruency effects – this was also true when the spatial dimension was emphasized by modifying contextual task conditions. At the sentence level, it is still a matter of debate whether language comprehenders create compositional mental simulations of sentential meaning. Adapting the standard sentence-picture verification task (e.g., Zwaan et al., 2002), I presented participants with sentential materials that allowed me to decide whether simulation effects indeed resulted from sentence-based mental simulations and if so, whether these are built incrementally over the course of the sentence. I found that language comprehenders produced mental simulations of sentential meaning that can hardly be attributed to independent word associations alone. However, these appear to be established only globally at the end of sentences. A last series of experiments was designed to tackle the challenging question how the abstract linguistic property of sentence polarity might be grounded in sensorimotor experiences. I tested whether primes in terms of visual information that has been associated with negation or affirmation – namely the “not face” (Benitez-Quiroz et al., 2016) and red/green color cues (Dudschig et al., 2023) – can facilitate the processing of negated and affirmative sentences. No effects emerged, indicating that these specific visual markers of negation and affirmation do not contribute to the embodied meaning of sentence polarity. Crucially, the findings presented in this dissertation relate to important topics and issues debated in the research on embodied language comprehension. This will be discussed in detail in the following sections.

5.1 Context-dependence of embodiment effects

Embodied cognition accounts of language comprehension (e.g., Barsalou, 1999; Bergen, 2012; Glenberg & Kaschak, 2002; Zwaan & Madden, 2005) assume that we understand the meaning of linguistic structures by mentally simulating the entities, situations, and events they describe. For instance, language comprehenders are expected to activate sensorimotor features associated with experiencing the referent of a word. This might include the typical color, shape or spatial location of the denoted entity. Much research on embodied language comprehension is based on the reasoning that the activation of sensorimotor features as part of a mental simulation should affect the performance in tasks involving corresponding perceptual or motor aspects and thus lead to embodied congruency effects. However, there is increasing empirical evidence and consensus that the emergence and characteristics of such effects highly depend on the experimental context (e.g., Areshenkoff et al., 2017; Brookshire et al., 2010; Huettig et al., 2020; Ibáñez et al., 2023; Lebois et al., 2015; Miller & Kaup, 2020; Tsaregorodtseva et al., 2023; van Dam et al., 2014; Willems & Casasanto, 2011). This is also reflected in the research on language-space associations originating from implicit location words (see section 2.1). As already described, tasks from the visual domain have been prone to evoke spatial congruency effects in terms of both facilitation and interference – and some studies could not find any effects at all. Details such as the timing of the processes of mental simulation and the presentation of visual target stimuli have been suggested to be potential factors of influence (Pecher et al., 2010). In addition, although tasks from the motor domain have regularly induced spatial congruency effects in terms of facilitation, their occurrence also seems to be modulated by task conditions (e.g., Dudschig & Kaup, 2017). In my research, I could reveal further constraints for the emergence of spatial congruency effects in the context of implicit location words: These can be realized by means of a direct real-time coupling of mouse movements on the horizontal plane and vertical stimulus movements but not through associating simple stationary keypress responses with visual action effects related to lower or upper space. Hence, the mere anticipation of spatially matching or mismatching action consequences does not suffice to produce spatial congruency effects. A more distinct and suitable involvement of the sensorimotor system – just as in terms of bringing together mouse movements and vertical stimulus movements – is apparently needed. These findings are in line with the suggestion by Lebois et al. (2015) that words do not have conceptual cores that are routinely activated irrespective of the context. The authors emphasize that this even applies to the most significant components of a word's meaning. For this reason, embodied congruency effects are expected to vary as a function of the experimental context – and their absence under certain conditions should not be interpreted as implying that the activation of sensorimotor experiences does not contribute to language comprehension. Ostarek and Huettig (2019) classified the systematic exploration of the context-dependence of embodiment effects as one of the most important challenges for future research on embodied language comprehension. Crucially, they also proposed to draw on more natural settings resembling everyday communication instead of focusing overly on experimental tasks that possibly prompt participants to rely on sensorimotor processes – this might help to understand to what extent mental simulations are indeed routinely created during language comprehension.

5.2 Sentence-based mental simulations

Human language use is characterized by the combination of words and phrases to form sentences conveying meaning beyond the word level. Hence, any comprehensive theoretical account of language comprehension needs to be able to explain how mental representations of sentential meaning might be established. Embodied cognition views of language comprehension (e.g., Barsalou, 1999; Bergen, 2012; Glenberg & Kaschak, 2002; Zwaan & Madden, 2005) usually suggest that sensorimotor experiences reactivated by words are combined to create compositional mental simulations of sentential meaning. Critically, the existence of such type of mental representation is indispensable if mental simulations should indeed play a functional role for language comprehension processes. Providing robust empirical evidence in favor of sentence-based mental simulations is therefore of utmost theoretical interest. However, many previous studies using sentential materials do not allow to judge whether reported simulation effects originated from independent word associations or meaning composition beyond the word level (see section 3.1). Findings on the temporal dynamics of mental simulations during sentence processing are rare as well (for exceptions, see Sato et al., 2013; Taylor & Zwaan, 2008; Zwaan & Taylor, 2006). Specifically – if language comprehenders build mental simulations of sentential meaning – the question arises whether these are created in an incremental manner over the course of sentences or only globally at the end of sentences, as a sort of wrap-up effect. In two experiments, I adapted the standard sentence-picture verification framework (e.g., Zwaan et al., 2002) and presented participants with specific materials that allowed me to provide insights into the nature and time course of mental simulations generated in response to sentences. The results suggested that language comprehenders tend to create compositional mental simulations of sentential meaning. However, there was no evidence that entity shape interpretations induced early during sentence processing were simulated, indicating that mental simulations of sentential content merely operated over the sentence as a whole. Given the theoretical importance of sentence-based mental simulations for embodied accounts of language comprehension and the scarce availability of studies producing simulation effects that can be attributed unequivocally to meaning composition processes beyond the word level, further research is needed to investigate the stability and boundary conditions of these findings. This might particularly include using other methods than the picture-verification framework, as it has been questioned whether matching effects in such tasks indeed rely on perceptual simulations (e.g., Ostarek et al., 2019). Finally, to be able to make more decisive conclusions on the time course of mental simulations during sentence comprehension, it will be necessary to draw on methods that allow a more direct look at on-line processes.

5.3 Functional relevance of mental simulations

Even if language comprehenders should routinely establish mental simulations reflecting the meaning of linguistic structures they encounter, this does not necessarily imply that these are indeed functionally relevant. Mental simulations could be mere by-products of amodal meaning composition processes that are not causally involved in language comprehension. Likewise, it is well conceivable that the influence of mental simulations highly depends on the specific task affordances at hand. Surprisingly, previous research has mostly focused on the pure activation of sensorimotor meaning aspects – and the very few behavioral studies addressing the question of functional relevance are usually limited to word-based stimuli and provided mixed results (see Pecher, 2013; Shebani & Pulvermüller, 2013; Strozyk et al., 2019; Yee et al., 2013). Even though the primary goal of my research on sentence-based mental simulations was not to investigate their causal role during language comprehension, the findings appear to provide some preliminary insights into this issue. This is particularly due to the fact that participants were presented with transitionally ambiguous garden-path sentences prompting them to mentally represent both an initial and a final sentence-based entity shape interpretation (Christianson et al., 2001; Patson et al., 2009; Slattery et al., 2013). If mental simulations are indeed functionally relevant, one would thus expect to see evidence in favor of mental simulations related to the initial entity shape interpretation. However, this was not the case: The results rather suggested that mental simulations were created only globally at the end of sentences, possibly as a sort of wrap-up effect following amodal meaning composition. As already described, studies varying the experimental context and involving measurements of on-line processes would help to probe these outcomes and arrive at robust and differentiated conclusions. Research on embodied cognition should generally be more ambitious and try harder to come up with designs that permit causal inferences with respect to the functional role of mental simulations during language processing – for instance, by using behavioral interference paradigms or applying transcranial magnetic stimulation (TMS) to sensorimotor brain areas (e.g., Ostarek & Bottini, 2020; Ostarek & Huettig, 2019). This is the actual litmus test and will give the opportunity to verify and refine or reject embodied views of language comprehension.

5.4 Embodiment of sentence polarity

The sophisticated human ability to reason and communicate about abstract concepts such as democracy or love is unique and must thus be addressed adequately by any theoretical framework of higher-order cognition (e.g., Borghi et al., 2017). This is particularly challenging to embodied views of cognition and language processing, as abstract concepts do not possess a straightforward physical referent in the world that can be captured by the sensorimotor system. There is no clear way to build associations between linguistic structures denoting abstract concepts and experiential states. Nonetheless, several attempts have been made to approach this issue (for comprehensive overviews, see Borghi et al., 2017; Kiefer & Harpaintner, 2020). The situation is much different when it comes to the vital question of how abstract linguistic properties might be grounded in mental simulations of sensorimotor experiences. This also applies to the case of sentence polarity, although there is some initial evidence indicating that sentential negation processing is embodied by means of reusing neurophysiological mechanisms of inhibition and cognitive control (e.g., Beltrán et al., 2018, 2019, 2021; de Vega et al., 2016; Liu et al., 2020). A complex phenomenon such as sentence polarity is likely to depend on various experiential meaning aspects. I therefore aimed to explore further potential grounding domains by investigating whether specific visual information related to negation or affirmation – namely the “not face” (Benitez-Quiroz et al., 2016) and red/green color cues (Dudschig et al., 2023) – can be integrated during the comprehension of negated and affirmative sentences. However, no priming effects could be detected, which might suggest that this type of perceptual experiences does not contribute to the embodied meaning representation of sentence polarity. These are first insights from a promising new avenue of research. Negation and affirmation tend to co-occur with several other sensorimotor states evoked by gestures or similar non-verbal ways of communication that could have gained semantic meaning. The task of future studies will be to examine in a more systematic manner whether some of these experiences are more than mere by-products of language use. If not, this might imply that mental processes enabling us to comprehend sentence polarity generally depend on a more abstract representational format.

5.5 Methodological aspects

Apart from providing new empirical findings related to theoretically important issues in the research on embodied language comprehension such as the creation of sentence-based mental simulations or the grounding of abstract linguistic properties, my work addressed aspects of practical relevance by introducing and testing an easy to implement, web-suited method for investigating vertical spatial associations. This particularly included replacing actual vertical response movements on specific input devices with a direct real-time coupling of horizontal mouse movements and vertical stimulus movements. Given the ever-increasing importance of data collection via the Internet (e.g., Gosling & Mason, 2015; Stewart et al., 2017; Woods et al., 2015), such a web-suited paradigm makes contemporary research on simulation effects induced by implicit location words much more feasible. Moreover, there are several other lines of research on embodied language comprehension and human cognition in general that are interested in investigating vertical spatial associations and might therefore benefit from a reliable method permitting researchers to refrain from gathering actual vertical response movements. For instance, an influential and extensively researched approach to explaining the mental representation of abstract concepts proposes that these are grounded in sensorimotor experiences by means of a metaphorical mapping onto domains that can be experienced in a direct manner (e.g., Lakoff & Johnson, 1980, 1999). Vertical space has turned out to be a particularly fertile experiential domain and was linked with the embodiment of numerous abstract concepts such as power (e.g., Jiang & Henley, 2012; Schubert, 2005; Wu et al., 2016; Zanolie et al., 2012), social status (e.g., Lu et al., 2014; Tower-Richardi et al., 2014; von Hecker et al., 2013), dominance (e.g., Moeller et al., 2008; Robinson et al., 2008), morality (e.g., Hill & Lapsley, 2009; Zhai et al., 2018), divinity (e.g., Chasteen et al., 2010; Meier et al., 2007), and emotional valence (e.g., Ansorge et al., 2013; Meier & Robinson, 2004; Santiago et al., 2012). Another area of application for the newly introduced paradigm might be numerical cognition, as number processing was shown to activate not only horizontal but also vertical spatial associations (Gevers et al., 2006; Ito & Hatta, 2004; Schwarz & Keus, 2004). However, it is also easily possible to convert the paradigm into a tool for investigating horizontal spatial associations. Apart from the case of numerical cognition, horizontal spatial associations play an important role in the research on the mental representation of time. It has been demonstrated that humans typically tend to map temporal information on a culturally salient horizontal mental timeline (von Sobbe et al., 2019). Likewise, testing the body-specificity hypothesis (Casasanto, 2009) is closely related to horizontal spatial associations. Originating from the assumption that cognitive processes rely on mental simulations of sensorimotor experiences, the basic idea is that humans with varying bodily characteristics make different experiences and should therefore create different mental representations. In particular, a lot of research has focused on the issue whether acting more fluently on the side of space that is associated with our dominant hand can modulate the mapping of positive and negative valence onto horizontal space (e.g., Casasanto & Henetz, 2012; Casasanto & Jasmin, 2010; de la Fuente et al., 2015; de la Vega et al., 2012; Li & Cao, 2019). Thus, to sum up, the newly introduced paradigm is extremely versatile and flexible and can help cognitive scientists from various areas of interest to exploit the full potential of web-based data collection.

5.6 Conclusions and outlook

Even though the number of publications dealing with embodied language comprehension has been growing tremendously during the last two decades, the current state of the art still appears to be quite primitive. Several aspects that must be tackled adequately by any serious approach to human language processing have been barely investigated in sufficient quantity and quality. For instance, this includes the question whether language comprehenders really create compositional mental simulations of sentential meaning and whether these are functionally relevant. Similarly, research has been inclined to report simulation effects in terms of both facilitation and interference, but there is – to the best of my knowledge – no theoretical account that integrates such findings and can successfully predict the direction of effects. The experiments outlined in this dissertation might contribute to a better understanding of some of these understudied and unresolved issues in the area of embodied language comprehension. For example, I presented a promising avenue of research that aims to ground sentence polarity in perceptual experiences that typically co-occur with negation or affirmation. Another highlight is certainly the finding that language comprehenders tend to create sentence-based mental simulations only globally at the end of sentences, possibly as a sort of wrap-up effect after amodal meaning composition processes have taken place. This could suggest that human language comprehension is not exclusively based on mental simulations, thus supporting the idea of so-called weak views of embodiment and hybrid models of cognition that acknowledge the role of both modal and amodal representational formats (e.g., Dove, 2009; Meteyard et al., 2012; Zwaan, 2014). Indeed, such approaches might be seen as a very first but important step on the indispensable way of revising and refining theories of embodied cognition and language processing.

6 References

Ahlberg, D. K., Bischoff, H., Kaup, B., Bryant, D., & Strozyk, J. V. (2018). Grounded cognition: Comparing language × space interactions in first language and second language. *Applied Psycholinguistics*, *39*(2), 437–459. https://doi.org/10.1017/S014271641700042X

Anderson, J. R. (1983). *The architecture of cognition*. Harvard University Press.

Ansorge, U., Khalid, S., & König, P. (2013). Space-valence priming with subliminal and supraliminal words. *Frontiers in Psychology*, *4*, Article 81. https://doi.org/10.3389/fpsyg.2013.00081

Areshenkoff, C. N., Bub, D. N., & Masson, M. E. J. (2017). Task-dependent motor representations evoked by spatial words: Implications for embodied accounts of word meaning. *Journal of Memory and Language, 92*, 158–169. https://doi.org/10.1016/j.jml.2016.06.006

Badets, A., Koch, I., & Philipp, A. M. (2016). A review of ideomotor approaches to perception, cognition, action, and language: Advancing a cultural recycling hypothesis. *Psychological Research*, *80*(1), 1–15. https://doi.org/10.1007/s00426-014-0643-8

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*(4), 577–660. https://doi.org/10.1017/S0140525X99002149

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

Barsalou, L. W., Simmons, W. K., Barbey, A. K., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences*, *7*(2), 84–91. https://doi.org/10.1016/S1364-6613(02)00029-3

Barsalou, L. W., & Wiemer-Hastings, K. (2005). Situating abstract concepts. In D. Pecher & R. A. Zwaan (Eds.), *Grounding cognition: The role of perception and action in memory, language, and thinking* (pp. 129–163). Cambridge University Press. https://doi.org/10.1017/CBO9780511499968.007

Barth, H., La Mont, K., Lipton, J., Dehaene, S., Kanwisher, N., & Spelke, E. (2006). Non-symbolic arithmetic in adults and young children. *Cognition*, *98*(3), 199–

222. https://doi.org/10.1016/j.cognition.2004.09.011

Beltrán, D., Liu, B., & de Vega, M. (2021). Inhibitory mechanisms in the processing of negations: A neural reuse hypothesis. *Journal of Psycholinguistic Research*, *50*(6), 1243–1260. https://doi.org/10.1007/s10936-021-09796-x

Beltrán, D., Morera, Y., García-Marco, E., & de Vega, M. (2019). Brain inhibitory mechanisms are involved in the processing of sentential negation, regardless of its content: Evidence from EEG theta and beta rhythms. *Frontiers in Psychology*, *10*, Article 1782. https://doi.org/10.3389/fpsyg.2019.01782

Beltrán, D., Muñetón-Ayala, M., & de Vega, M. (2018). Sentential negation modulates inhibition in a stop-signal task: Evidence from behavioral and ERP data. *Neuropsychologia*, *112*, 10–18. https://doi.org/10.1016/j.neuropsychologia.2018.03.004

Benitez-Quiroz, C. F., Wilbur, R. B., & Martinez, A. M. (2016). The not face: A grammaticalization of facial expressions of emotion. *Cognition*, *150*, 77–84. https://doi.org/10.1016/j.cognition.2016.02.004

Bergen, B. K. (2012). *Louder than words: The new science of how the mind makes meaning*. Basic Books.

Bergen, B. K., Lindsay, S., Matlock, T., & Narayanan, S. (2007). Spatial and linguistic aspects of visual imagery in sentence comprehension. *Cognitive Science*, *31*(5), 733–764. https://doi.org/10.1080/03640210701530748

Bergen, B. K., & Wheeler, K. (2010). Grammatical aspect and mental simulation. *Brain and Language*, *112*(3), 150–158. https://doi.org/10.1016/j.bandl.2009.07.002

Borghi, A. M., Binkofski, F., Castelfranchi, C., Cimatti, F., Scorolli, C., & Tummolini, L. (2017). The challenge of abstract concepts. *Psychological Bulletin*, *143*(3), 263–292. https://doi.org/10.1037/bul0000089

Brookshire, G., Ivry, R., & Casasanto, D. (2010). Modulation of motor-meaning congruity effects for valenced words. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *32*, 1940–1945. https://escholarship.org/uc/item/2ch97289

Bub, D. N., & Masson, M. E. J. (2010). On the nature of hand-action representations evoked during written sentence comprehension. *Cognition*, *116*(3), 394–408. https://doi.org/10.1016/j.cognition.2010.06.001

Casasanto, D. (2009). Embodiment of abstract concepts: Good and bad in right- and left-handers. *Journal of Experimental Psychology: General*, *138*(3), 351–367. https://doi.org/10.1037/a0015854

Casasanto, D., & Henetz, T. (2012). Handedness shapes children's abstract concepts. *Cognitive Science*, *36*(2), 359–372. https://doi.org/10.1111/j.1551-6709.2011.01199.x

Casasanto, D., & Jasmin, K. (2010). Good and bad in the hands of politicians: Spontaneous gestures during positive and negative speech. *PLoS ONE*, *5*(7), Article e11805. https://doi.org/10.1371/journal.pone.0011805

Chasteen, A. L., Burdzy, D. C., & Pratt, J. (2010). Thinking of God moves attention. *Neuropsychologia*, *48*(2), 627–630. https://doi.org/10.1016/j.neuropsychologia.2009.09.029

Chomsky, N. (1980). *Rules and representations*. Columbia University Press.

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, *42*(4), 368–407. https://doi.org/10.1006/cogp.2001.0752

de la Fuente, J., Casasanto, D., Román, A., & Santiago, J. (2015). Can culture influence body-specific associations between space and valence? *Cognitive Science*, *39*(4), 821–832. https://doi.org/10.1111/cogs.12177

de la Vega, I., de Filippis, M., Lachmair, M., Dudschig, C., & Kaup, B. (2012). Emotional valence and physical space: Limits of interaction. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(2), 375–385. https://doi.org/10.1037/a0024979

de Vega, M., Morera, Y., León, I., Beltrán, D., Casado, P., & Martín-Loeches, M. (2016). Sentential negation might share neurophysiological mechanisms with action inhibition: Evidence from frontal theta rhythm. *Journal of Neuroscience*, *36*(22), 6002–6010. https://doi.org/10.1523/JNEUROSCI.3736-15.2016

Dimroth, C. (2010). The acquisition of negation. In L. R. Horn (Ed.), *The expression of* *negation* (pp. 39–71). De Gruyter Mouton.

Dove, G. (2009). Beyond perceptual symbols: A call for representational pluralism. *Cognition*, *110*(3), 412–431. https://doi.org/10.1016/j.cognition.2008.11.016

Dudschig, C., de la Vega, I., & Kaup, B. (2014). Embodiment and second-language: Automatic activation of motor responses during processing spatially associated L2 words and emotion L2 words in a vertical Stroop paradigm. *Brain and Language*, *132*, 14–21. https://doi.org/10.1016/j.bandl.2014.02.002

Dudschig, C., de la Vega, I., & Kaup, B. (2015). What's up? Emotion-specific activation of vertical space during language processing. *Acta Psychologica*, *156*, 143–155. https://doi.org/10.1016/j.actpsy.2014.09.015

Dudschig, C., & Kaup, B. (2017). Is it all task-specific? The role of binary responses, verbal mediation, and saliency for eliciting language-space associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(2), 259–270. https://doi.org/10.1037/xlm0000297

Dudschig, C., Kaup, B., Leuthold, H., & Mackenzie, I. G. (2021). Conceptual representation of real-world surface material: Early integration with linguistic labels indicated in the N400 component. *Psychophysiology*, *58*(12), Article e13916. https://doi.org/10.1111/psyp.13916

Dudschig, C., Kaup, B., & Mackenzie, I. G. (2023). The grounding of logical operations: The role of color, shape, and emotional faces for "yes" or "no" decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *49*(3), 477–492. https://doi.org/10.1037/xlm0001181

Dudschig, C., Lachmair, M., de la Vega, I., De Filippis, M., & Kaup, B. (2012a). From top to bottom: Spatial shifts of attention caused by linguistic stimuli. *Cognitive Processing*, *13*(S1), 151–154. https://doi.org/10.1007/s10339-012-0480-x

Dudschig, C., Lachmair, M., de la Vega, I., De Filippis, M., & Kaup, B. (2012b). Do task-irrelevant direction-associated motion verbs affect action planning? Evidence from a Stroop paradigm. *Memory & Cognition*, *40*(7), 1081–1094. https://doi.org/10.3758/s13421-012-0201-9

Dudschig, C., Souman, J., Lachmair, M., de la Vega, I., & Kaup, B. (2013). Reading "sun" and looking up: The influence of language on saccadic eye movements in the vertical dimension. *PLoS ONE*, *8*(2), Article e56872. https://doi.org/10.1371/journal.pone.0056872

Dunn, B., Kamide, Y., & Scheepers, C. (2014). Hearing "moon" and looking up: Word-related spatial associations facilitate saccades to congruent locations. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *36*, 433–438. https://escholarship.org/uc/item/78z895zd

Estes, Z., Verges, M., & Barsalou, L. W. (2008). Head up, foot down: Object words orient attention to the objects' typical location. *Psychological Science*, *19*(2), 93–97. https://doi.org/10.1111/j.1467-9280.2008.02051.x

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

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

Foglia, L., & Wilson, R. A. (2013). Embodied cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*(3), 319–325. https://doi.org/10.1002/wcs.1226

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*(3), 221–233. https://doi.org/10.1016/j.actpsy.2005.11.004

Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences*, *20*(1), 1–55. https://doi.org/10.1017/S0140525X97000010

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, *9*(3), 558–565. https://doi.org/10.3758/BF03196313

Gosling, S. D., & Mason, W. (2015). Internet research in psychology. *Annual Review of Psychology*, *66*, 877–902. https://doi.org/10.1146/annurev-psych-010814-015321

Günther, F., Dudschig, C., & Kaup, B. (2018). Symbol grounding without direct experience: Do words inherit sensorimotor activation from purely linguistic context? *Cognitive Science*, *42*(S2), 336–374. https://doi.org/10.1111/cogs.12549

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, *42*(1–3), 335–346. https://doi.org/10.1016/0167-2789(90)90087-6

Hauf, J. E. K., Nieding, G., & Seger, B. T. (2020). The development of dynamic perceptual simulations during sentence comprehension. *Cognitive Processing, 21*(2), 197–208. https://doi.org/10.1007/s10339-020-00959-7

Hill, P. L., & Lapsley, D. K. (2009). The ups and downs of the moral personality: Why it's not so black and white. *Journal of Research in Personality*, *43*(3), 520–523. https://doi.org/10.1016/j.jrp.2009.01.014

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, *24*(5), 849–878. https://doi.org/10.1017/S0140525X01000103

Huettig, F., Guerra, E., & Helo, A. (2020). Towards understanding the task dependency of embodied language processing: The influence of colour during language-vision interactions. *Journal of Cognition*, *3*(1), Article 41. https://doi.org/10.5334/joc.135

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

Ito, Y., & Hatta, T. (2004). Spatial structure of quantitative representation of numbers: Evidence from the SNARC effect. *Memory & Cognition*, *32*(4), 662–673. https://doi.org/10.3758/BF03195857

Janczyk, M., & Ulrich, R. (2019). Action consequences affect the space-time congruency effect on reaction time. *Acta Psychologica*, *198*, Article 102850. https://doi.org/10.1016/j.actpsy.2019.05.002

Jiang, M., & Henley, T. B. (2012). Power and spatial relations. *Journal of Cognitive Psychology*, *24*(7), 829–835. https://doi.org/10.1080/20445911.2012.702749

Kaschak, M. P., Madden, C. J., Therriault, D. J., Yaxley, R. H., Aveyard, M., Blanchard, A. A., & Zwaan, R. A. (2005). Perception of motion affects language processing. *Cognition*, *94*(3), B79–B89. https://doi.org/10.1016/j.cognition.2004.06.005

Kaup, B., de la Vega, I., Strozyk, J., & Dudschig, C. (2016). The role of sensorimotor processes in meaning composition. In M. H. Fischer & Y. Coello (Eds.), *Foundations of embodied cognition: Conceptual and interactive embodiment* (pp. 46–70). Routledge. https://doi.org/10.4324/9781315751962

Kaup, B., & Dudschig, C. (2017). Sätze und Texte verstehen und produzieren. In J. Müsseler & M. Rieger (Eds.), *Allgemeine Psychologie* (3rd ed., pp. 467–530). Springer. https://doi.org/10.1007/978-3-642-53898-8_14

Kiefer, M., & Harpaintner, M. (2020). Varieties of abstract concepts and their grounding in perception or action. *Open Psychology*, *2*(1), 119–137. https://doi.org/10.1515/psych-2020-0104

Kiefer, M., & Pulvermüller, F. (2012). Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions. *Cortex*, *48*(7), 805–825. https://doi.org/10.1016/j.cortex.2011.04.006

Kintsch, W., & Keenan, J. (1973). Reading rate and retention as a function of the number of propositions in the base structure of sentences. *Cognitive Psychology*, *5*(3), 257–274. https://doi.org/10.1016/0010-0285(73)90036-4

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*(5), 363–394. https://doi.org/10.1037/0033-295X.85.5.363

Koch, S. C., Glawe, S., & Holt, D. V. (2011). Up and down, front and back: Movement and meaning in the vertical and sagittal axes. *Social Psychology*, *42*(3), 214–224. https://doi.org/10.1027/1864-9335/a000065

Kousta, S.-T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: Why emotion matters. *Journal of Experimental Psychology: General*, *140*(1), 14–34. https://doi.org/10.1037/a0021446

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*(6), 1180–1188. https://doi.org/10.3758/s13423-011-0158-x

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. University of Chicago Press.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to western thoughts*. Basic Books.

Lebois, L. A. M., Wilson-Mendenhall, C. D., & Barsalou, L. W. (2015). Are automatic conceptual cores the gold standard of semantic processing? The context-dependence of spatial meaning in grounded congruency effects. *Cognitive Science*, *39*(8), 1764–1801. https://doi.org/10.1111/cogs.12174

Li, H., & Cao, Y. (2019). The body in religion: The spatial mapping of valence in Tibetan practitioners of Bön. *Cognitive Science*, *43*(4), Article e12728. https://doi.org/10.1111/cogs.12728

Liu, B., Wang, H., Beltrán, D., Gu, B., Liang, T., Wang, X., & de Vega, M. (2020). The generalizability of inhibition-related processes in the comprehension of linguistic negation: ERP evidence from the Mandarin language. *Language, Cognition, and Neuroscience*, *35*(7), 885–895. https://doi.org/10.1080/23273798.2019.1662460

Lu, A., Zhang, H., He, G., Zheng, D., & Hodges, B. H. (2014). Looking up to others: Social status, Chinese honorifics, and spatial attention. *Canadian Journal of Experimental Psychology*, *68*(2), 77–83. https://doi.org/10.1037/cep0000008

Machery, E. (2006). Two dogmas of neo-empiricism. *Philosophy Compass*, *1*(4), 398–412. https://doi.org/10.1111/j.1747-9991.2006.00030.x

Machery, E. (2007). Concept empiricism: A methodological critique. *Cognition*, *104*(1), 19–46. https://doi.org/10.1016/j.cognition.2006.05.002

Machery, E. (2016). The amodal brain and the offloading hypothesis. *Psychonomic Bulletin & Review*, *23*(4), 1090–1095. https://doi.org/10.3758/s13423-015-0878-4

Masson, M. E. J., Bub, D. N., & Lavelle, H. (2013). Dynamic evocation of hand action representations during sentence comprehension. *Journal of Experimental Psychology: General*, *142*(3), 742–762. https://doi.org/10.1037/a0030161

Masson, M. E. J., Bub, D. N., & Newton-Taylor, M. (2008). Language-based access to gestural components of conceptual knowledge. *Quarterly Journal of Experimental Psychology*, *61*(6), 869–882. https://doi.org/10.1080/17470210701623829

Masson, M. E. J., Bub, D. N., & Warren, C. M. (2008). Kicking calculators: Contribution of embodied representations to sentence comprehension. *Journal of Memory and Language*, *59*(3), 256–265. https://doi.org/10.1016/j.jml.2008.05.003

Meier, B. P., Hauser, D. J., Robinson, M. D., Friesen, C. K., & Schjeldahl, K. (2007). What's "up" with God? Vertical space as a representation of the divine. *Journal of Personality and Social Psychology*, *93*(5), 699–710. https://doi.org/10.1037/0022-3514.93.5.699

Meier, B. P., & Robinson, M. D. (2004). Why the sunny side is up: Associations between affect and vertical position. *Psychological Science*, *15*(4), 243–247. https://doi.org/10.1111/j.0956-7976.2004.00659.x

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

Michel, C. (2021). Overcoming the modal/amodal dichotomy of concepts. *Phenomenology and the Cognitive Sciences*, *20*(4), 655–677. https://doi.org/10.1007/s11097-020-09678-y

Miller, J., & Kaup, B. (2020). Influences of task and attention on action verb congruence effects: How automatic are embodiment effects? *Acta Psychologica*, *210*, Article 103155. https://doi.org/10.1016/j.actpsy.2020.103155

Moeller, S. K., Robinson, M. D., & Zabelina, D. L. (2008). Personality dominance and preferential use of the vertical dimension of space: Evidence from spatial attention paradigms. *Psychological Science*, *19*(4), 355–361. https://doi.org/10.1111/j.1467-9280.2008.02093.x

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, *4*(2), 61–64. https://doi.org/10.20982/tqmp.04.2.p061

Ostarek, M., & Bottini, R. (2020). Towards strong inference in research on embodiment: Possibilities and limitations of causal paradigms. *Journal of Cognition*, *4*(1), Article 5. https://doi.org/10.5334/joc.139

Ostarek, M., & Huettig, F. (2019). Six challenges for embodiment research. *Current Directions in Psychological Science*, *28*(6), 593–599. https://doi.org/10.1177/0963721419866441

Ostarek, M., Joosen, D., Ishag, A., de Nijs, M., & Huettig, F. (2019). Are visual processes causally involved in "perceptual simulation" effects in the sentence-picture verification task? *Cognition*, *182*, 84–94. https://doi.org/10.1016/j.cognition.2018.08.017

Öttl, B., Dudschig, C., & Kaup, B. (2017). Forming associations between language and sensorimotor traces during novel word learning. *Language and Cognition*, *9*(1), 156–171. https://doi.org/10.1017/langcog.2016.5

Patson, N. D., Darowski, E. S., Moon, N., & Ferreira, F. (2009). Lingering misinterpretations in garden-path sentences: Evidence from a paraphrasing task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(1), 280–285. https://doi.org/10.1037/a0014276

Pecher, D. (2013). No role for motor affordances in visual working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(1), 2–13. https://doi.org/10.1037/a0028642

Pecher, D., Van Dantzig, S., Boot, I., Zanolie, K., & Huber, D. E. (2010). Congruency between word position and meaning is caused by task-induced spatial attention. *Frontiers in Psychology*, *1*, Article 30. https://doi.org/10.3389/fpsyg.2010.00030

Pylyshyn, Z. W. (1984). *Computation and cognition: Towards a foundation for cognitive science*. MIT Press.

Richardson, D. C., Spivey, M. J., Barsalou, L. W., & McRae, K. (2003). Spatial representations activated during real-time comprehension of verbs. *Cognitive Science*, *27*(5), 767–780. https://doi.org/10.1016/S0364-0213(03)00064-8

Robinson, M. D., Zabelina, D. L., Ode, S., & Moeller, S. K. (2008). The vertical nature of dominance-submission: Individual differences in vertical attention. *Journal of Research in Personality*, *42*(4), 933–948. https://doi.org/10.1016/j.jrp.2007.12.002

Rück, F., de la Vega, I., Leuthold, H., & Kaup, B. (2017). Integration of visual information about the speaker during sentence processing. In S. Featherston, R. Hörnig, R. Steinberg, B. Umbreit, & J. Wallis (Eds.), *Proceedings of Linguistic Evidence 2016: Empirical, theoretical, and computational perspectives*. University of Tübingen. https://doi.org/10.15496/publikation-19060

Santiago, J., Ouellet, M., Román, A., & Valenzuela, J. (2012). Attentional factors in conceptual congruency. *Cognitive Science*, *36*(6), 1051–1077. https://doi.org/10.1111/j.1551-6709.2012.01240.x

Sato, M., Schafer, A. J., & Bergen, B. K. (2013). One word at a time: Mental representations of object shape change incrementally during sentence processing. *Language and Cognition*, *5*(4), 345–373. https://doi.org/10.1515/langcog-2013-0022

Schneider, D. W. (2020). Alertness and cognitive control: Interactions in the spatial Stroop task. *Attention, Perception, and Psychophysics*, *82*(5), 2257–2270. https://doi.org/10.3758/s13414-020-01993-5

Schubert, T. W. (2005). Your highness: Vertical positions as perceptual symbols of power. *Journal of Personality and Social Psychology*, *89*(1), 1–21. https://doi.org/10.1037/0022-3514.89.1.1

Schütt, E., Mackenzie, I. G., Kaup, B., & Dudschig, C. (2023). Replacing vertical actions by mouse movements: A web-suited paradigm for investigating vertical spatial associations. *Psychological Research*, *87*(1), 194–209. https://doi.org/10.1007/s00426-022-01650-6

Schwarz, W., & Keus, I. M. (2004). Moving the eyes along the mental number line: Comparing SNARC effects with saccadic and manual responses. *Perception and Psychophysics*, *66*(4), 651–664. https://doi.org/10.3758/BF03194909

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, *3*(3), 417–457. https://doi.org/10.1017/S0140525X00005756

Šetić, M., & Domijan, D. (2007). The influence of vertical spatial orientation on property verification. *Language and Cognitive Processes*, *22*(2), 297–312. https://doi.org/10.1080/01690960600732430

Shebani, Z., & Pulvermüller, F. (2013). Moving the hands and feet specifically impairs working memory for arm- and leg-related action words. *Cortex*, *49*(1), 222–231. https://doi.org/10.1016/j.cortex.2011.10.005

Shin, Y. K., Proctor, R. W., & Capaldi, E. J. (2010). A review of contemporary ideomotor theory. *Psychological Bulletin*, *136*(6), 943–974. https://doi.org/10.1037/a0020541

Slattery, T. J., Sturt, P., Christianson, K., Yoshida, M., & Ferreira, F. (2013). Lingering misinterpretations of garden path sentences arise from competing syntactic representations. *Journal of Memory and Language*, *69*(2), 104–120. https://doi.org/10.1016/j.jml.2013.04.001

Stewart, N., Chandler, J., & Paolacci, G. (2017). Crowdsourcing samples in cognitive science. *Trends in Cognitive Sciences*, *21*(10), 736–748. https://doi.org/10.1016/j.tics.2017.06.007

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662. https://doi.org/10.1037/h0054651

Strozyk, J. V., Dudschig, C., & Kaup, B. (2019). Do I need to have my hands free to understand hand-related language? Investigating the functional relevance of experiential simulations. *Psychological Research*, *83*(3), 406–418. https://doi.org/10.1007/s00426-017-0900-8

Taylor, L. J., & Zwaan, R. A. (2008). Motor resonance and linguistic focus. *Quarterly Journal of Experimental Psychology*, *61*(6), 896–904. https://doi.org/10.1080/17470210701625519

Thornton, T., Loetscher, T., Yates, M. J., & Nicholls, M. E. R. (2013). The highs and lows of the interaction between word meaning and space. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(4), 964–973. https://doi.org/10.1037/a0030467

Tower-Richardi, S. M., Brunyé, T. T., Gagnon, S. A., Mahoney, C. R., & Taylor, H. A. (2014). Living the high life: Social status influences real estate decision making. *Journal of Applied Social Psychology*, *44*(9), 611–621. https://doi.org/10.1111/jasp.12253

Tsaregorodtseva, O., Frazier, L., Stolterfoht, B., & Kaup, B. (2023). Do linguistic stimuli activate experiential colour traces related to the entities they refer to and, if so, under what circumstances? *Quarterly Journal of Experimental Psychology*. Advance online publication. https://doi.org/10.1177/17470218231200489

van Dam, W. O., Brazil, I. A., Bekkering, H., & Rueschemeyer, S.-A. (2014). Flexibility in embodied language processing: Context effects in lexical access. *Topics in Cognitive Science, 6*(3), 407–424. https://doi.org/10.1111/tops.12100

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

Vogt, A., Kaup, B., & Dudschig, C. (2019). When words are upside down: Language-space associations in children and adults. *Journal of Experimental Child Psychology*, *186*, 142–158. https://doi.org/10.1016/j.jecp.2019.06.001

von Hecker, U., Klauer, K. C., & Sankaran, S. (2013). Embodiment of social status: Verticality effects in multilevel rank-orders. *Social Cognition*, *31*(3), 374–389. https://doi.org/10.1521/soco.2013.31.3.374

von Sobbe, L., Scheifele, E., Maienborn, C., & Ulrich, R. (2019). The space-time congruency effect : A meta-analysis. *Cognitive Science*, *43*(1), Article e12709. https://doi.org/10.1111/cogs.12709

Willems, R. M., & Casasanto, D. (2011). Flexibility in embodied language understanding. *Frontiers in Psychology*, *2*, Article 116. https://doi.org/10.3389/fpsyg.2011.00116

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, *9*(4), 625–636. https://doi.org/10.3758/BF03196322

Woods, A. T., Velasco, C., Levitan, C. A., Wan, X., & Spence, C. (2015). Conducting perception research over the Internet: A tutorial review. *PeerJ*, *3*, Article e1058. https://doi.org/10.7717/peerj.1058

Wu, X., Jia, H., Wang, E., Du, C., Wu, X., & Dang, C. (2016). Vertical position of

Chinese power words influences power judgments: Evidence from spatial compatibility task and event-related potentials. *International Journal of Psychophysiology*, *102*, 55–61. https://doi.org/10.1016/j.ijpsycho.2016.03.005

Yee, E., Chrysikou, E. G., Hoffman, E., & Thompson-Schill, S. L. (2013). Manual experience shapes object representations. *Psychological Science*, *24*(6), 909–919. https://doi.org/10.1177/0956797612464658

Zanolie, K., Van Dantzig, S., Boot, I., Wijnen, J., Schubert, T. W., Giessner, S. R., & Pecher, D. (2012). Mighty metaphors: Behavioral and ERP evidence that power shifts attention on a vertical dimension. *Brain and Cognition*, *78*(1), 50–58. https://doi.org/10.1016/j.bandc.2011.10.006

Zhai, D., Guo, Y., & Lu, Z. (2018). A dual mechanism of cognition and emotion in processing moral-vertical metaphors. *Frontiers in Psychology*, *9,* Article 1554. https://doi.org/10.3389/fpsyg.2018.01554

Zuberbühler, K. (2015). Linguistic capacity of non-human animals. *Wiley Interdisciplinary Reviews: Cognitive Science*, *6*(3), 313–321. https://doi.org/10.1002/wcs.1338

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

Zwaan, R. A., & Madden, C. J. (2005). Embodied sentence comprehension. In D. Pecher & R. A. Zwaan (Eds.), *Grounding cognition: The role of perception and action in memory, language, and thinking* (pp. 224–245). Cambridge University Press. https://doi.org/10.1017/CBO9780511499968.010

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science*, *13*(2), 168–171. https://doi.org/10.1111/1467-9280.00430

Zwaan, R. A., & Taylor, L. J. (2006). Seeing, acting, understanding: Motor resonance in language comprehension. *Journal of Experimental Psychology: General*, *135*(1), 1–11. https://doi.org/10.1037/0096-3445.135.1.1

Zwaan, R. A., & Yaxley, R. H. (2003). Spatial iconicity affects semantic relatedness judgments. *Psychonomic Bulletin & Review, 10*(4), 954–958. https://doi.org/10.3758/BF03196557

Appendix A: Article 1

Copyright notice

To cite the article

Schütt, E., Mackenzie, I. G., Kaup, B., & Dudschig, C. (2023). Replacing vertical actions by mouse movements: A web-suited paradigm for investigating vertical spatial associations. *Psychological Research, 87*(1), 194-209. https://doi.org/10.1007/s00426-022-01650-6

Psychological Research (2023) 87:194–209
https://doi.org/10.1007/s00426-022-01650-6

ORIGINAL ARTICLE

Replacing vertical actions by mouse movements: a web-suited paradigm for investigating vertical spatial associations

Emanuel Schütt[1] · Ian Grant Mackenzie[2] · Barbara Kaup[1] · Carolin Dudschig[1]

Received: 23 June 2021 / Accepted: 19 January 2022 / Published online: 7 February 2022

Abstract
The number of web-based studies in experimental psychology has been growing tremendously throughout the last few years. However, a straightforward web-based implementation does not exist for all types of experimental paradigms. In the current paper, we focus on how vertical response movements—which play a crucial role in spatial cognition and language research—can be translated into a web-based setup. Specifically, we introduce a web-suited counterpart of the vertical Stroop task (e.g., Fox & Shor, in Bull Psychon Soc 7:187–189, 1976; Lachmair et al., in Psychon Bull Rev 18:1180–1188, 2011; Thornton et al., in J Exp Psychol Hum Percept Perform 39:964–973, 2013). We employed nouns referring to entities typically located in lower or upper vertical space (e.g., "worm" and "bird", respectively) in Experiments 1 and 2, and emotional valence words associated with a crouched or an upward bodily posture (e.g., "sadness" and "excitement", respectively) in Experiment 3. Depending on the font color, our participants used their mouse to drag the words to the lower or upper screen location. Across all experiments, we consistently observed congruency effects analogous to those obtained with the lab paradigm using actual vertical arm movements. Consequently, we conclude that our web-suited paradigm establishes a reliable approach to examining vertical spatial associations.

Introduction

Within the last decade, the Internet has become increasingly relevant for behavioral research. This development is currently being boosted by the coronavirus pandemic. Experimental psychologists started to employ the Internet as a research tool in the middle of the 1990s (Krantz & Dalal, 2000; Reips, 2000, 2002a). Up to the present day, the number of studies using the Internet for delivering surveys and running experiments has grown tremendously (Gosling & Mason, 2015; Stewart et al., 2017; Woods et al., 2015). Nevertheless, the potential of the Internet with respect to conducting behavioral research—especially in terms of chronometric studies—is not yet entirely realized. In the present article, we focused on providing a setup that replaces vertical response movements by means of mouse movements on the horizontal plane inducing vertical stimulus movements on the computer screen. Importantly, this setup is easy to implement, allows web-based data collection, and can also simplify research in the lab.

Web-based data collection provides the opportunity to overcome several issues associated with classical lab research. Particularly, the Internet and online labor markets (e.g., Amazon Mechanical Turk or Prolific) enable researchers (1) to gather data of participants with a much more diverse background than the typically recruited university students (for issues with "WEIRD" samples in behavioral science see Henrich et al., 2010), (2) to sample data in short periods of time, and (3) to recruit participants with special characteristics (Birnbaum, 2004; Reips, 2000; Stewart et al., 2017; Woods et al., 2015). Naturally, we also face challenges and disadvantages when we decide to deliver surveys or conduct experiments via the Internet. Especially, researchers must deal with the fact of losing experimental control (e.g., Duffy, 2002; Gosling & Mason, 2015; Nosek et al., 2002). For example, the experimenter has reduced options for making sure that the participants follow the instructions and take their participation seriously. Due to the lack of personal interaction, there is no way of clearing up misunderstandings

Emanuel Schütt
emanuel.schuett@uni-tuebingen.de

[1] Department of Psychology, Language and Cognition Research Group, University of Tübingen, Schleichstr. 4, 72076 Tübingen, Germany

[2] Department of Psychology, Biological Psychology Research Group, University of Tübingen, Schleichstr. 4, 72076 Tübingen, Germany

or answering questions (Reips, 2002b). Moreover, several authors mention problems with respect to data security and research ethics, such as the issue that the experimenter cannot guarantee that the participants read and comprehended the informed consent statement (e.g., Buchanan & Williams, 2010; Emery, 2014; Kraut et al., 2004; Rhodes et al., 2003). In addition, lab paradigms and their web-based counterparts will rarely be fully identical (for a similar issue regarding psychological tests see Buchanan, 2002). However, there are ways and means of mitigating, managing, and (partially) solving the challenges and disadvantages that are typically related to conducting psychological research via the Internet (see, for example, Aust et al., 2013; Reips, 2000, 2002a, 2002b, 2009).

Before turning to the paradigm introduced and evaluated in the present research, we will briefly discuss the issue of timing in web-based reaction time (RT) experiments. Certainly, there exist some reservations mainly arising from arguments such as software and technology constraints and the increasing situational and technical variance in web-based data collection (Hilbig, 2016). However, it seems that the respective issues are significantly less severe than originally expected. For instance, Reimers and Stewart (2015) systematically investigated the accuracy of RT measurements in web-based experiments across different computers, operating systems, ways of implementation (Adobe Flash vs. HTML5), and browsers. Their results showed that RT effects can be detected accurately in most setups. Importantly, for a between-subjects design with two conditions, they demonstrated that the noise generated by hardware and software variability can easily be compensated by slightly raising the sample size. Moreover, for a within-subjects design with two conditions, they detected virtually no disadvantages (see Neath et al., 2011, for results with Apple computers). In a recent review, Stewart et al. (2017) stated that "it is now possible to measure reaction times sufficiently accurately in web experiments using HTML5 and Javascript" (p. 739). Additionally, several (classical) RT paradigms have been tested and validated in online settings, mostly either by replicating the results of the original lab studies or by administering the paradigm in a lab setting and in an online setting. For instance, lab and online results were comparable for the Stroop effect, the Simon effect, the flanker task, the attentional blink task, task-switching costs, visual cuing, visual search, the word frequency effect, the right-visual-field advantage for word recognition, and syntactic priming in sentence production (Corley & Scheepers, 2002; Crump et al., 2013; Hilbig, 2016; Linnman et al., 2006; McGraw et al., 2000; Semmelmann & Weigelt, 2017; Simcox & Fiez, 2014).

In general, web-based experimenting seems to be completely feasible and acceptable with respect to experimental designs that require keypresses on a standard keyboard. In addition, in recent years, mouse-based paradigms have been successfully used across various research domains to gain insights into cognitive processes (for reviews, see Freeman, 2018; Schoemann et al., 2021; Stillman et al., 2018). For instance, researchers employed mouse-tracking to examine semantic categorization (Dale et al., 2007), to study social categorization (Freeman et al., 2010), to detect response difficulty in online surveys (Horwitz et al., 2017), and to establish a potential early marker of mild cognitive decline (Seelye et al., 2015). In the present article, we aimed at making use of these developments for introducing a web-suited counterpart to vertical response movements typically recorded in lab-based settings. This sort of vertical response mode plays a crucial role in various branches of behavioral research, such as spatial cognition, numerical cognition, social cognition, and cognitive control (e.g., Dudschig & Kaup, 2020; Ito & Hatta, 2004; Koch et al., 2011; Schneider, 2020; Schubert, 2005; Schwarz & Keus, 2004).

We based our work on research using the vertical Stroop task. For example, Lachmair et al. (2011) employed such a setup (see also Dudschig & Kaup, 2017; Thornton et al., 2013) for verifying assumptions that were derived from the experiential-simulations view of language comprehension (Barsalou, 1999; Glenberg & Kaschak, 2002; Zwaan & Madden, 2005). In their vertical Stroop task, Lachmair et al. (2011, Experiment 2) presented participants with nouns referring to entities that are typically located in lower or upper vertical space (e.g., "worm" and "bird", respectively). Critically, each noun was displayed in one of four font colors. Depending on the font color, the participants performed a downward or upward arm movement employing a vertical response device mounted in front of them (see Fig. 1). Concretely, they pressed down the two middle keys of the device using their right and left hand to initiate a trial. To respond to the font color of the noun, they released the relevant middle key, pressed the corresponding lower or upper response key and returned to the released middle key. Importantly, the time period from the appearance of the noun until releasing one of the middle keys (releasing time) served as the dependent variable. In line with the original Stroop task (Stroop, 1935) correct responses solely demanded processing the font color. Thus, the task did not require participants to process the nouns. Nevertheless, Lachmair et al. observed a congruency effect of response direction and referent location, indicating that spatial features are indeed activated in a rather automatic manner when people encounter nouns that are associated with a typical vertical location (see Thornton et al., 2013, for converging results). Remarkably, the results suggest that the setup is able to detect spatial congruency effects that are likely smaller than those obtained with the standard spatial Stroop task using the words "up" and "down" (Fox & Shor, 1976; Fox et al., 1971). Hence, establishing a reliable web-based counterpart to this type of

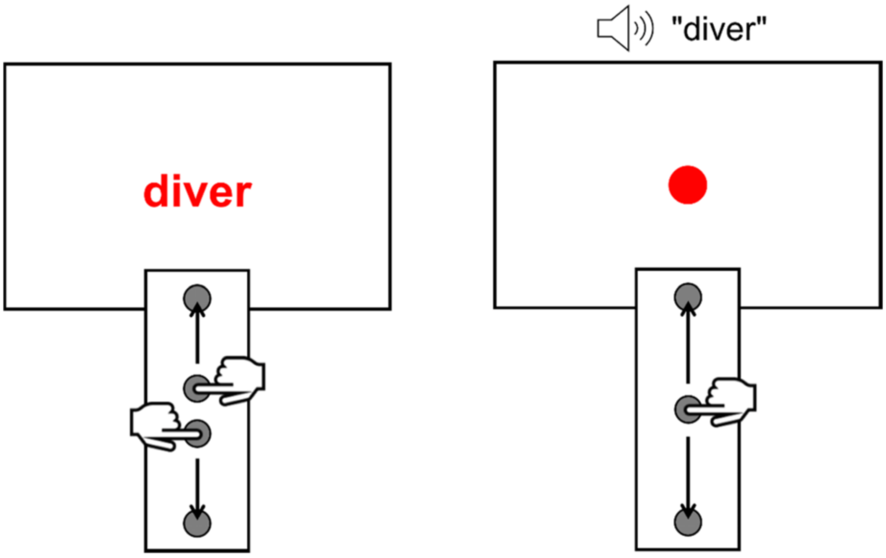


Fig. 1 Examples of setups used in research on vertical language-space associations. The illustration on the left shows the experimental setup used by Lachmair et al. (2011, Experiment 2). Participants performed a downward or an upward arm movement to respond to the font color of nouns. The vertically mounted response device had four keys: two middle keys, a lower response key, and an upper response key. Participants initiated a trial by holding both middle keys. A downward (an upward) arm movement included releasing the corresponding middle key, pressing the lower (the upper) response key, and returning to the middle key. The non-responding hand remained on the middle key. The illustration on the right shows the modified setup used by Vogt et al. (2019). In this case, participants performed downward and upward movements with a single arm. Subsequent to the auditory presentation of the language stimulus, they responded to the color of a circle displayed at the center of the screen

setup seems to be highly valuable. Moreover, in recent years, setups with a vertical response dimension have become increasingly relevant in the various fields of behavioral research (e.g., spatial cognition; language comprehension; numerical cognition; social cognition) that have an interest in spatial associations (e.g., Ahlberg et al., 2018; Dudschig et al., 2012, 2014, 2015; Gevers et al., 2006; Günther et al., 2018, 2020; Hill & Lapsley, 2009; Ito & Hatta, 2004; Meier et al., 2007; Öttl et al., 2017; Schubert, 2005; Vogt et al., 2019; Zanolie et al., 2012; Zhai et al., 2018).

In our web-suited counterpart to the manual vertical Stroop paradigm, participants used their computer mouse to respond to the font color of words: They dragged each word either to the lower or to the upper screen location. Consequently, spatial congruency was operationalized via the correspondence of spatial information associated with the language stimulus (e.g., “grave” vs. “satellite”) and the mouse movement that was directly coupled with visually sensible action effects (i.e., perceiving the word moving downwards or upwards, respectively). Noteworthy, in a series of five experiments (Schütt et al., 2021), we recently tested a similar paradigm. In this paradigm, our participants responded to the font color of nouns by performing stationary keypresses. Importantly, pressing a response key consistently produced an immediate visual feedback at the lower or upper end of the screen. Interestingly, we did not reliably observe congruency effects, indicating that visual feedback following simple keypress responses does not suffice to induce the sort of congruency effects studied in research domains that typically demand vertical responses. However, we hypothesized that coupling visual action effects with a more distinct involvement and activation of the motor system—as realized by introducing mouse movements—might be more similar to previous lab-based approaches.[1] Additionally, using mouse movements seems to be a rather convenient solution: The computer mouse is one of the standard input devices available for web-based research and allows implementing experimental procedures that involve the motor system in a more prominent way.

In three mouse movement experiments we explored whether our web-suited paradigm is appropriate for studying language-space associations in a reliable manner. We started by running web-based replications of the study by Lachmair et al. (2011, Experiment 2) that investigated the automatic activation of spatial information during processing implicit

---

[1] Of course, a mouse-based paradigm also has the potential to replace specific response devices requiring vertical actions in future lab-based research on spatial associations.

location words, once with native English speakers (Experiment 1) and once with German native speakers (Experiment 2). In Experiment 3, we applied the web-suited paradigm to another group of words, namely valence words referring to emotions that are typically associated with a slouched or an upright bodily posture (i.e., with spatial experiences that are characterized by having a down vs. an up component; see Dudschig et al., 2015; Meier & Robinson, 2004). Importantly, using the lab-based vertical Stroop paradigm of Lachmair et al., it has already been shown that these words influence subsequent vertical arm movements in accordance with the spatial experiences they are related to (see Dudschig et al., 2015, Experiment 3). Therefore, the third experiment enables us to assess the generalizability of our web-suited paradigm by means of testing it with respect to yet another well documented lab-based finding.

Experiment 1

Method

Transparency and openness

For all experiments, we report our procedure to determine the exact sample size. All data sets and the original and an updated version of the jsPsych plugin for implementing the mouse movement task are publicly available at https://doi.org/10.5281/zenodo.4557024. We will upload analysis scripts at the same place upon publication. All data exclusions, software employed for statistical analyses, and study materials are mentioned and described in the Method sections at an appropriate place. The current research was not preregistered. Ethics approval for the study was obtained from the Ethics Committee for Psychological Research at the University of Tübingen (Identifier: 2018_0831_132).

Participants

We based the sample size of the experiment on the lab-based work that served as the starting point for the current research (i.e., Lachmair et al., 2011, Experiment 2 [$N = 24$]). First, we conducted a power analysis using MorePower (Version 6.0.4; Campbell & Thompson, 2012). For this purpose, we referred to the results Lachmair et al. (2011, Experiment 2) reported for the critical congruency effect of response direction and referent location (by-participants analysis of variance; $\eta_p^2 = 0.24$). This revealed that 36 participants are needed for having a test power of 0.90 with respect to the congruency effect of response direction and referent location (for similar results obtained with a simulation-based procedure, see Günther et al., 2018). Given that primary studies in behavioral research tend to overestimate effect sizes (Fanelli & Ioannidis, 2013), we then also considered the advice of Simonsohn (2015) that replications should have 2.5 times as many observations as the original study. To incorporate both criteria when setting the sample size, we aimed at collecting the data of 60 participants. In Experiment 1, we recruited native English speakers via Amazon Mechanical Turk. In accordance with Lachmair et al., we discarded participants with a low accuracy rate (less than 90% of correct trials in at least one experimental condition). This procedure resulted in a final sample of 55 participants (i.e., five participants were excluded due to the accuracy rate criterion). The age of the participants (30 males, 25 females; 46 right-handed, eight left-handed, one ambidextrous) ranged from 21 to 65 years ($M = 38.62$ years, $SD = 10.88$ years). The experiment took about 25 min to complete. All participants gave informed consent and received $4.00 in return for participation.

Apparatus and stimuli

We created our online experiments by use of jsPsych (Version 6.1.0; de Leeuw, 2015), which is an open-source JavaScript library for generating web-based behavioral experiments. To implement the mouse movement task, we employed a custom-made jsPsych plugin (publicly available at https://doi.org/10.5281/zenodo.4557024). We instructed our participants to use only either a desktop computer or a laptop and a standard computer mouse for conducting the experiments.

A black colored plus sign served as fixation cross. The experimental stimuli were derived from the word list of Lachmair et al. (2011, Experiment 2), which comprised German nouns denoting objects that are typically associated with a lower or an upper vertical location. First, we translated the word list into American English. As a result of the translation, some of the nouns were composed of more than a single word (e.g., “roof beam”; “high wire”; “sole of foot”). However, this was not the case for any of the original German nouns. Therefore, we deleted these nouns. Finally, we randomly excluded further nouns to have a list with the same number of down and up words. This produced a final word list with 32 down words and 32 up words (see Appendix). We examined the nouns with respect to length and frequency. For determining frequency classes, we used an English news corpus that was based on

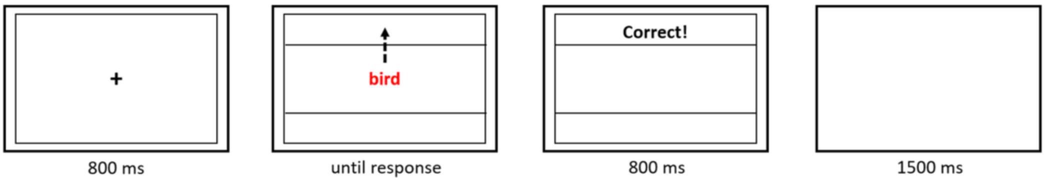


Fig. 2 Procedure of a correctly answered trial in Experiment 1. Each trial started with the presentation of a fixation cross. Concurrently, a frame appeared that marked the experimentally relevant part of the screen. Then, a lower and an upper target area was introduced by displaying a borderline at the lower and at the upper end of the framed part of the screen. Moreover, the stimulus replaced the fixation cross. The participants used their mouse to respond to the font color of the stimulus: They dragged the stimulus either to the lower or to the upper target area. Once the stimulus was completely located in one of the target areas, it was replaced by the feedback. Finally, the intertrial interval followed before the next trial started

texts from 2016 (available at https://wortschatz.uni-leipzig.de). Down words ($M = 5.47$ letters, $SD = 1.41$ letters) and up words ($M = 5.53$ letters, $SD = 1.92$ letters) did not differ significantly in length, $t(62) = -0.15$, $p = 0.883$. Likewise, there was no significant difference in terms of frequency, $t(61) = 0.67$, $p = 0.504$ (down words: $M = 12.45$, $SD = 2.10$; up words: $M = 12.06$, $SD = 2.49$).[2] Sixteen additional nouns (e.g., “book”; “letter”; “machine”) served as stimuli in the training session. These nouns were not associated with any typical vertical location (see Lachmair et al., 2011, Experiment 3). The stimuli were presented in blue (RGB: 0, 0, 255), orange (RGB: 255, 165, 0), green (RGB: 0, 128, 0), and red (RGB: 255, 0, 0) font color on a white background. In the training session as well as in the experimental session, we gave feedback by displaying either “Correct!” (if the response was correct) or “Wrong!” (if the response was wrong). The feedback was shown in black letters.

Procedure

We instructed our participants to run the experiment in an interference-free environment. The experiment consisted of a training session and an experimental session. During the training session, participants should get familiar with the task. We presented each training stimulus once in one of the four font colors. Thus, the training session comprised 16 trials. All font colors appeared equally often. The order of trials was randomized. After the training session, a self-paced break followed. This break included performance feedback (i.e., information on the percentage of correct responses and the average response time) and a reminder with respect to the experimental task.

Participants then proceeded with the experimental session, which was composed of four blocks. In each block, we displayed each of the down words and each of the up words once in one of the four font colors. Hence, an experimental block contained 64 trials. We presented each noun in a different font color in each block. Since there were four blocks and four different font colors, all nouns appeared exactly once in each font color during the experimental session. Furthermore, we balanced the font colors within the blocks (i.e., all font colors were used equally often per block). The order of the trials and the order of the blocks were randomized. After each block, there was a self-paced break. Again, this break included a performance feedback and a reminder regarding the experimental task.

Figure 2 illustrates the trial procedure. Each trial started with showing simultaneously (1) a frame that marked the experimentally relevant part of the screen (90% of the available display height and 90% of the available display width) and (2) the fixation cross (800 ms). Subsequently, a lower and an upper target area was introduced by displaying a borderline at the lower and at the upper end of the framed part of the screen. At the same time, the stimulus replaced the fixation cross. The task was to react to the font color of the stimulus as fast and as accurately as possible. For this purpose, the participants clicked on the stimulus and dragged it either to the lower or to the upper target area using their mouse. We always mapped two font colors to one response direction (e.g., blue and orange to downward dragging and green and red to upward dragging). The combination of colors to color pairs (blue-orange and green–red; blue-green and orange-red; blue-red and orange-green) and the mapping of color pairs to response directions (i.e., which of the two color pairs was assigned to which of the two response directions) was balanced between participants, resulting in six experimental versions. Once the stimulus was completely located in one of the target areas, it was replaced by the correctness feedback (800 ms). Finally, the intertrial interval (1500 ms) followed before the next trial started.

---

[2] The corpus that we used for determining the word frequency classes did not give a value for the noun “subfont”. Thus, we discarded this noun from the analysis of frequencies, which reduced the number of degrees of freedom in the t-test.

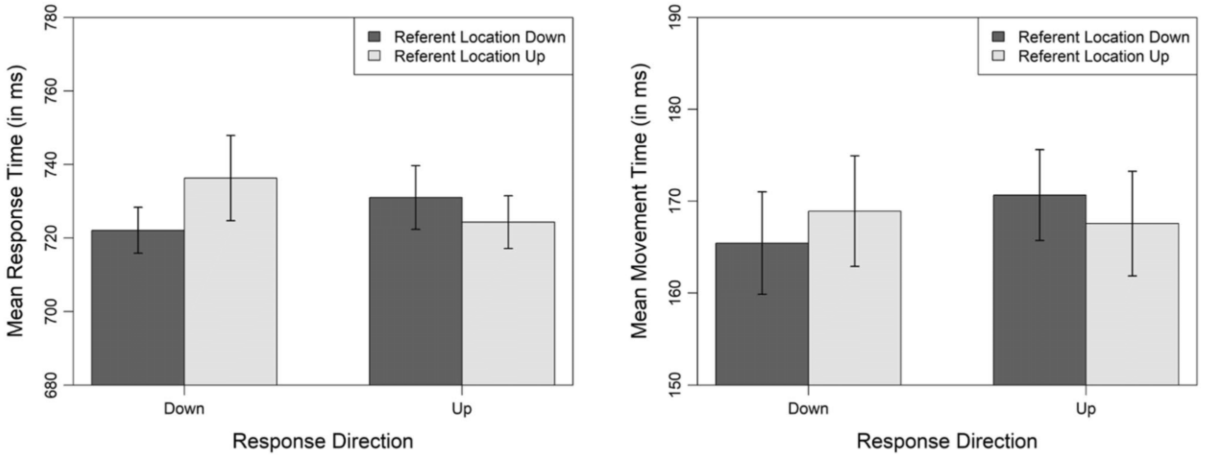


Fig. 3 Mean response and mean movement times for correctly answered trials as a function of referent location and response direction in Experiment 1. Error bars denote 95% within-subjects confidence intervals calculated as recommended by Morey (2008

Design and data analysis

The experiment had a 2 × 2 within-subjects design, with the factors referent location (down vs. up) and response direction (down vs. up). As dependent variables we used measures that likely match the dependent variables considered in the context of the lab-based vertical Stroop task we referred to (see, for example, Dudschig et al., 2015; Lachmair et al., 2011; Thornton et al., 2013). This particularly included a marker reflecting the time period until initiating a response to the presented stimulus (i.e., response selection and planning). Thus, the time period from the appearance of the noun on the screen until the initial movement of the noun (response time) served as a dependent variable. More concretely, the noun needed to be selected via mouse button press, followed by the mouse movement being started (there was no need for moving a certain distance).

We prepared and analyzed response times using the free statistical software R (Version 3.6.2). First, we removed training trials, trials with more than a single mouse click, and incorrectly answered trials. Then, we excluded trials with response times shorter than 100 or longer than 3000 ms. In line with Lachmair et al. (2011), we subsequently applied the two-step procedure suggested by Kaup et al. (2006) to eliminate further outliers. Hence, in a first step, we converted the response times of each participant to z-scores. Following this, we discarded response times with a z-score that deviated more than two standard deviations from the mean z-score of the respective noun in the respective condition. Thus, both differences among the participants and differences among the items were considered. In sum, outlier elimination reduced the data set by less than 7%. We used the R packages lme4 (Version 1.1-21; Bates et al., 2015) and lmerTest (Version 3.1-1; Kuznetsova et al., 2017) to perform a linear mixed effect analysis (see Baayen et al., 2008). Our base model contained fixed effects for referent location and response direction. In contrast to Lachmair et al., we collected data of participants with varying handedness and a rather broad range of age. Consequently, to account for any possible effects of handedness and age, we also added fixed effects for handedness and age to our base model. For determining a suitable random effect structure, we employed the data-driven model selection criterion of Matuschek et al. (2017), which aims at providing a mixed model that balances Type I error rate and power. When performing the procedure, we omitted models that had a singular fit or did not converge. This resulted in incorporating random intercepts for participants and items. For testing our hypothesis (i.e., the interaction of referent location and response direction), we compared our base model to a model that contained an additional fixed effect for the interaction of referent location and response direction by means of a likelihood ratio test.

The time period from the initial movement of the noun until the noun crossed one of the response boundaries (movement time) served as a second dependent variable. We prepared and analyzed movement times in the same way as response times, except that extreme outliers were defined as movement times longer than 1000 ms. Outlier elimination—comprising the two-step procedure recommended by Kaup et al. (2006)—reduced the data set with respect to movement times by less than 6%. The mixed models included random intercepts for participants and items.

Finally, we looked at response correctness. For this purpose, we referred to all correctly and incorrectly answered experimental trials, excluding trials with response times shorter than 100 or longer than 3000 ms (less than 2% of all trials). The analysis followed the procedure used for response and movement times, apart from the fact that we formulated generalized linear mixed models (i.e., mixed effects logistic regressions) to handle the binary outcome variable (response: correct vs. incorrect). These models comprised random intercepts for participants and items.

Results and discussion

Figure 3 illustrates the mean response and mean movement times as a function of referent location and response direction. For response times, the analysis revealed that the model with an additional fixed effect for the interaction of referent location and response direction explained the data significantly better than the base model, $\chi^2(1) = 6.18$, $p = 0.013$, effect of referent location: $\beta = 10.81$, $t = 1.99$, 95% CI [0.15, 21.47], effect of response direction: $\beta = 3.15$, $t = 0.64$, 95% CI [$-$ 6.45, 12.76], interaction of referent location and response direction: $\beta = -17.24$, $t = -2.49$, 95% CI [$-$ 30.82, $-$ 3.65]. This indicates that responses times were significantly faster when the referent location matched ($M = 723$ ms) compared to mismatched ($M = 734$ ms) the response direction. For movement times, the model with an additional fixed effect for the interaction of referent location and response direction showed a marginally better fit than the base model, $\chi^2(1) = 3.58$, $p = 0.058$, effect of referent location: $\beta = 2.25$, $t = 1.02$, 95% CI [$-$ 2.09, 6.58], effect of response direction: $\beta = 3.22$, $t = 1.56$, 95% CI [$-$ 0.84, 7.28], interaction of referent location and response direction: $\beta = -5.55$, $t = -1.89$, 95% CI [$-$ 11.30, 0.20]. Movements times were slightly faster when the referent location matched ($M = 167$ ms) compared to mismatched ($M = 170$ ms) the movement direction of the noun on the screen. There was no evidence for an interaction of referent location and response direction with respect to response correctness, $\chi^2(1) = 0.91$, $p = 0.340$, effect of referent location: $\beta = -0.16$, $z = -0.71$, 95% CI [$-$ 0.60, 0.28], effect of response direction: $\beta = -0.07$, $z = -0.31$, 95% CI [$-$ 0.51, 0.37], interaction of referent location and response direction: $\beta = 0.30$, $z = 0.95$, 95% CI [$-$ 0.32, 0.92]. Error rates were virtually identical when the referent location matched ($M = 1.20\%$) compared to mismatched ($M = 1.31\%$) the response direction.

We were successful in replicating the congruency effect of referent location and response direction, which provides first support in favor of our paradigm. Particularly, the effect emerged for response times (i.e., the period until initiating the mouse movement). Therefore, the stage of action planning seems to be crucial. Interestingly, this is in line with findings of previous lab-based work on language-space associations (e.g., Dudschig et al., 2015; Günther et al., 2018; Lachmair et al., 2011; Öttl et al., 2017). The effect also clearly tended to be mirrored in movement times. Furthermore, error rates were low, indicating that participants focused on the experimental task.

Experiment 2

In Experiment 2, we aimed at obtaining further support regarding the validity of our web-suited counterpart to the manual vertical Stroop paradigm. Therefore, we ran a replication of Experiment 1 with German materials and native German speakers.

Method

Participants

We again collected the data of 60 participants. We recruited native German speakers by sending a circular email to the students at the University of Tübingen. The exclusion of participants followed the same accuracy rate criterion as in Experiment 1. This yielded a final sample of 57 participants (i.e., three participants were excluded due to the accuracy rate criterion). The participants (40 females, 17 males; 54 right-handed, two left-handed, one ambidextrous) were between 18 and 36 years old ($M = 23.67$ years, $SD = 3.51$ years). The experiment took about 25 min to complete. All participants gave informed consent. In return for participation, the participants either received partial course credit or participated in a lottery of three vouchers with a value of €60 each.

Apparatus and stimuli

In general, apparatus and stimuli were the same as in Experiment 1. This time, however, we conducted the experiment in German. Thus, we presented the original German counterparts of the nouns that we used in Experiment 1 (see Appendix). Again, we examined the nouns with respect to length and frequency. For determining frequency classes, we made use of a German news corpus that was based on texts from 2018 (available at https://wortschatz.uni-leipzig.de). Down words ($M = 5.88$ letters, $SD = 1.70$ letters) and up words ($M = 5.69$ letters, $SD = 1.42$ letters) did not differ significantly in length, $t(62) = 0.48$, $p = 0.634$. Likewise, there was no significant difference in terms of frequency, $t(62) = 0.57$, $p = 0.571$ (down words: $M = 11.91$, $SD = 1.87$; up words: $M = 11.63$, $SD = 2.08$).

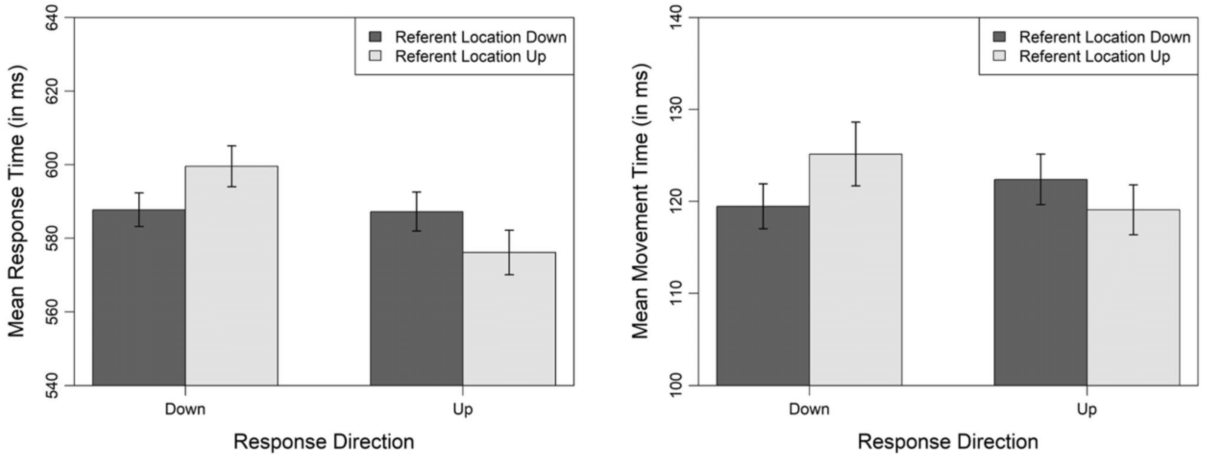


Fig. 4 Mean response and mean movement times for correctly answered trials as a function of referent location and response direction in Experiment 2. Error bars denote 95% within-subjects confidence intervals calculated as recommended by Morey (2008)

Procedure

The procedure remained as in Experiment 1.

Design and data analysis

Design, data preparation, and data analysis were identical to Experiment 1. Eliminating outliers reduced the data set by less than 5% (response times), less than 6% (movement times), and less than 1% (response correctness). Our method for determining appropriate random effect structures for the mixed models resulted in incorporating random intercepts for participants and items in all models. In addition, the mixed models for analyzing movement times also included by-item random slopes for response location.

Results and discussion

Figure 4 depicts the mean response and mean movement times as a function of referent location and response direction. For response times, the analysis showed that the model with a further fixed effect for the interaction of referent location and response direction explained the data significantly better than the base model, $\chi^2(1) = 27.42$, $p < 0.001$, effect of referent location: $\beta = 11.34$, $t = 3.63$, 95% CI [5.21, 17.47], effect of response direction: $\beta = -0.09$, $t = -0.03$, 95% CI [− 5.84, 5.65], interaction of referent location and response direction: $\beta = -21.67$, $t = -5.24$, 95% CI [− 29.78, − 13.56]. Once again, response times were significantly faster when the referent location matched ($M = 582$ ms) compared to mismatched ($M = 593$ ms) the response direction. For movement times, the model with an additional fixed effect for the interaction of referent location and response direction had a significantly better fit than the base model, $\chi^2(1) = 12.51$, $p < 0.001$, effect of referent location: $\beta = 6.30$, $t = 4.19$, 95% CI [3.36, 9.25], effect of response direction: $\beta = 3.23$, $t = 1.84$, 95% CI [− 0.20, 6.66], interaction of referent location and response direction: $\beta = -9.19$, $t = -3.72$, 95% CI [− 14.04, − 4.34]. Movements times were significantly faster when the referent location matched ($M = 119$ ms) compared to mismatched ($M = 124$ ms) the movement direction of the noun on the screen. Finally, there was an interaction of referent location and response direction regarding response correctness, $\chi^2(1) = 21.60$, $p < 0.001$, effect of referent location: $\beta = -0.65$, $z = -3.44$, 95% CI [− 1.02, − 0.28], effect of response direction: $\beta = -0.67$, $z = -3.54$, 95% CI [− 1.04, − 0.30], interaction of referent location and response direction: $\beta = 1.22$, $z = 4.64$, 95% CI [0.71, 1.74]. Errors were rare and the error rate was lower when the referent location matched ($M = 1.24\%$) compared to mismatched ($M = 2.22\%$) the response direction, ruling out a speed-accuracy tradeoff.

Just as in Experiment 1, we successfully replicated the congruency effect of referent location and response direction. Noteworthy, this time, the effect was reflected in both response times and movement times. In sum, this demonstrates that our web-suited paradigm is suitable for generating spatial congruency effects in a stable manner. Furthermore, by running the experiment in German and recruiting native German speakers, we showed that the paradigm works in different samples and languages.

Experiment 3

So far, we tested our web-suited counterpart to the manual vertical Stroop task with respect to the stimuli that were used by Lachmair et al. (2011, Experiment 2). Accordingly, we looked at the automatic activation of spatial information during the processing of implicit location words. However, there is strong evidence indicating that language-space associations also exist for various other groups of words. For instance, this applies to power-related words (e.g., "king" and "servant"; Jiang & Henley, 2012; Wu et al., 2016; Zanolie et al., 2012), words describing religious concepts (e.g., "Lord" and "Satan"; Chasteen et al., 2010; Meier et al., 2007), and valence words (e.g., "love" and "danger"; Meier & Robinson, 2004; Santiago et al., 2012). Interestingly, Dudschig et al. (2015, Experiment 3) employed the lab-based vertical Stroop paradigm of Lachmair et al. (2011, Experiment 2) to investigate whether processing a rather specific subset of valence words automatically influences subsequent arm movements. These valence words (e.g., "optimistic" and "disappointed") were characterized by referring to emotional states that are typically related to an upright or a slouched bodily posture. Crucially, responses in the lab-based experiment were faster when the vertical association of the valence words matched the response direction of the arm movements, reflecting the spatial congruency effect previously reported for implicit location words (see Lachmair et al., 2011, Experiment 2). In Experiment 3, we applied our web-suited paradigm to the posture-specific emotional valence words of Dudschig et al. (2015, Experiment 3). This enabled us to resolve the question whether our web-suited paradigm can replace vertical response movements in the context of further types of spatial associations.

Method

Participants

As in the previous experiments, we based the sample size of the current experiment on the lab-based work that served as a starting point (i.e., Dudschig et al., 2015, Experiment 3 [$N=18$]). Accordingly, we performed a power analysis using MorePower (Version 6.0.4; Campbell & Thompson, 2012) based on the results Dudschig et al. (2015, Experiment 3) observed for the interaction of response direction and vertical association (by-participants analysis of variance for posture-specific emotional valence words; $\eta_p^2 = 0.21$). This revealed that 44 participants would be necessary for obtaining a power of 0.90 with respect to replicating the interaction of response direction and vertical association. Once again, we also considered the suggestion of Simonsohn (2015) to determine the number of participants. This resulted in a target sample size of 48 participants.[3] We recruited native German speakers by sending a circular email to the students at the University of Tübingen. One participant had to be discarded by virtue of the accuracy rate criterion (less than 90% of correct trials in at least one experimental condition). Thus, the final sample consisted of 47 participants. The participants (36 females, 11 males; 42 right-handed, five left-handed) were between 18 and 56 years old ($M=22.89$ years, $SD = 5.83$ years). The experiment took about 30 min to complete. All participants gave informed consent. In return for participation, the participants either received partial course credit or participated in a lottery of three vouchers with a value of €75 each.

Apparatus and stimuli

Apparatus and stimuli remained the same as in Experiment 2, except that we replaced the experimental stimuli. In the current experiment, we showed the list of posture-specific emotional valence words that was used by Dudschig et al. (2015, Experiment 3). The list comprised 20 German valence words referring to pleasant emotions that are associated with an upright bodily posture and 20 German valence words referring to unpleasant emotions that are associated with a rather slouched bodily posture. The full word list can be found in the Appendix.

Procedure

Regarding the procedure, there were some minor changes in comparison to Experiment 2. In accordance with Dudschig et al. (2015, Experiment 3), we repeated the experimental stimuli eight times throughout the experiment. Consequently, the experimental session contained eight blocks. In each block, we displayed each of the valence words once in one of the four font colors. Hence, an experimental block comprised 40 trials. All valence words appeared exactly twice in each font color during the experiment. Everything else remained as in Experiment 2.

Design and data analysis

The experiment had a 2 × 2 within-subjects design, with the factors vertical association (down vs. up) and response direction (down vs. up). The dependent variables were defined as in the previous experiments. Data preparation and data

---

[3] Simonsohn (2015) recommended for replications to have 2.5 times as many observations as the original study. This would yield a sample size of 45 participants. However, due to balancing demands, we needed a sample size that is a multiple of six. Hence, we set the target sample size to 48 participants.

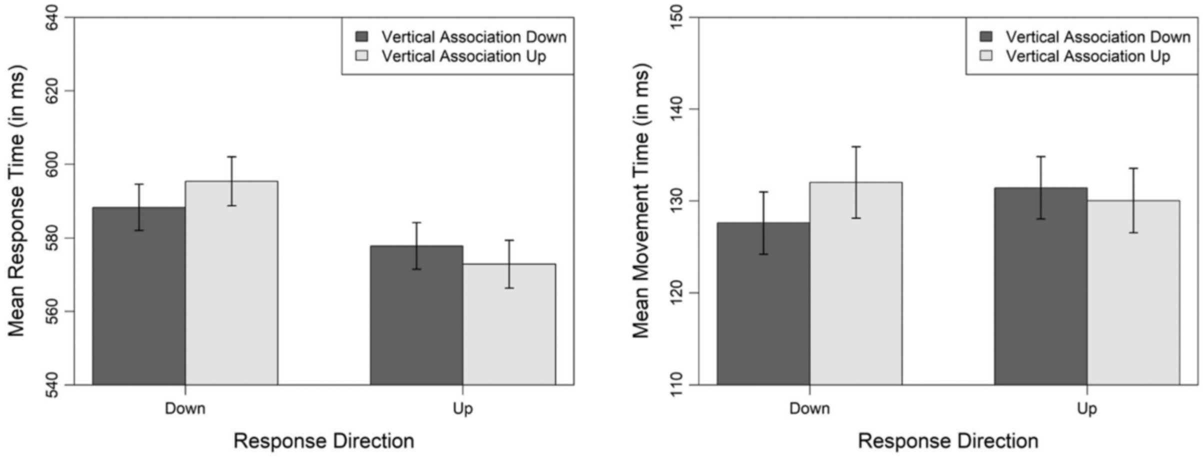


Fig. 5 Mean response and mean movement times for correctly answered trials as a function of vertical association and response direction in Experiment 3. Error bars denote 95% within-subjects confidence intervals calculated as recommended by Morey (2008).

analysis were also identical. Outlier removal reduced the data set by less than 5% (response times and movement times) and by less than 1% (response correctness). The method for determining appropriate random effect structures for the mixed models resulted in including random intercepts for participants and items in all models.

Results and discussion

Figure 5 displays the mean response and mean movement times as a function of vertical association and response direction. For response times, the analysis demonstrated that the critical model with an additional fixed effect for the interaction of the factors vertical association and response direction explained the data significantly better than the base model, $\chi^2(1)=6.69$, $p=0.009$, effect of vertical association: $\beta = 7.19$, $t=1.97$, 95% CI [0.05, 14.32], effect of response direction: $\beta = -8.97$, $t = -2.79$, 95% CI [− 15.29, − 2.66],– interaction of vertical association and response direction: $\beta = -11.79$, $t = -2.59$, 95% CI [− 20.73, − 2.86]. Hence, response times were significantly faster when the vertical association of the word matched ($M=581$ ms) compared to mismatched ($M=587$ ms) the response direction. For movement times, the model with a further fixed effect for the interaction of vertical association and response direction had a significantly better fit than the base model, $\chi^2(1)=5.99$, $p=0.014$, effect of vertical association: $\beta = 3.94$, $t=2.59$, 95% CI [0.96, 6.92], effect of response direction: $\beta = 3.67$, $t=2.61$, 95% CI [0.92, 6.42], interaction of vertical association and response direction: $\beta = -4.86$, $t = -2.45$, 95% CI [− 8.74, − 0.97]. Movement times were significantly faster when the vertical association of the word matched ($M=129$ ms) compared to mismatched ($M=132$ ms) the movement direction of the word on the screen. We obtained no evidence for an interaction of vertical association and response direction on response correctness, $\chi^2(1)=1.86$, $p=0.172$, effect of vertical association: $\beta = -0.33$, $z = -1.60$, 95% CI [− 0.73, 0.07], effect of response direction: $\beta = -0.23$, $z = -1.15$, 95% CI [− 0.62, 0.16], interaction of vertical association and response direction: $\beta = 0.39$, $z=1.40$, 95% CI [− 0.15, 0.92]. Error rates were similarly low when the vertical association matched ($M=1.28\%$) compared to mismatched ($M=1.51\%$) the response direction.

Crucially, we reproduced the spatial congruency effect with respect to posture-specific emotional valence words. Consequently, the application of our web-suited paradigm is clearly not restricted to solely investigating language-space associations that emerge from implicit location words (see Experiments 1 and 2). Instead, our paradigm seems to establish a reliable suited approach to investigating different varieties of spatial associations.

General discussion

Throughout the last years, experimental psychologists have increasingly used web-based research tools (Gosling & Mason, 2015; Stewart et al., 2017; Woods et al., 2015). Nevertheless, behavioral science is far from exploiting the full potential of web-based data collection. In the current article, we contributed a paradigm for expanding web-based behavioral research to setups that would typically require vertical response movements. For this purpose, we made use of the vertical Stroop task, which has been particularly relevant to investigating spatial associations across various research fields (e.g., Ahlberg et al., 2018; Dudschig et al., 2012, 2014, 2015; Gevers et al., 2006; Günther et al., 2018, 2020; Ito & Hatta, 2004; Lachmair et al., 2011; Müller & Schwarz, 2007; Öttl et al., 2017; Schwarz & Keus, 2004; Thornton et al., 2013; Vicovaro & Dalmaso, 2021; Vogt et al., 2019).

Interestingly, attempts to replace vertical response movements by providing visual action effects in vertical space failed (Schütt et al., 2021). The present research demonstrates that a more distinct involvement and activation of the motor system is needed. This was realized via coupling mouse movements in the horizontal plane with vertical stimulus movements on the screen. Specifically, our participants used their mouse to respond to the font color of stimuli: They dragged each stimulus either to the lower or upper screen location. In sum, three experiments showed that this paradigm is appropriate for reliably replicating spatial congruency effects.

In the first two experiments, we examined language-space associations of nouns referring to entities typically located in a lower or an upper vertical location (see Lachmair et al., 2011). We used English (Experiment 1) and German (Experiment 2) materials and recruited English (Experiment 1) and German (Experiment 2) native speakers. In both experiments, response times (i.e., the time period from the appearance of the noun on the computer screen until initiating the mouse movement) were faster when the referent's typical vertical location matched the dragging direction. This corresponds to the spatial congruency effect obtained with the lab-based version of the manual vertical Stroop paradigm (Lachmair et al., 2011, Experiment 2). Consequently, the experiments provided first empirical support in favor of our web-suited paradigm. In addition, the results revealed that our paradigm seems to work in different languages and populations.

In Experiment 3, we aimed at assessing the effectiveness of the paradigm with respect to further vertical spatial associations. To this end, we investigated valence words referring to emotions that are associated with a slouched or an upright bodily posture (i.e., with spatial experiences that are characterized by having a down vs. an up component). Interestingly, the investigation of such valence-space associations has become of ever-increasing interest since the study of Meier and Robinson (2004) was published. Crucially, we again observed a spatial congruency effect (faster responses when the vertical association of the words matched the dragging direction) that was in line with previous lab-based work using the manual vertical Stroop task (see Dudschig et al., 2015, Experiment 3). This clearly indicates that the applicability of our paradigm is not restricted to investigating a specific type of language-space association. Rather the paradigm seems to constitute a reliable web-suited approach to examining a variety of issues related to vertical spatial associations.

Moreover, the current results indicate that the emergence of spatial congruency effects stemming from language-space associations (i.e., faster responses when the vertical association of the word matches the response location; Dudschig et al., 2015, Experiment 3; Lachmair et al., 2011, Experiment 2) does not necessarily require actual vertical response movements. Whereas participants perform vertical arm movements to respond to the font color of words in standard lab-based paradigms, our participants moved their mouse on the horizontal plane to drag the words to the lower or upper screen location. Nevertheless, implementing a close relationship of manual actions (i.e., the hand movement when operating the mouse) and sensible action effects (i.e., seeing the word moving downwards or upwards, respectively) appears to be an essential factor as visual feedback per se could not reliably provoke spatial congruency effects in setups with simple stationary keypress responses (Schütt et al., 2021).

The exact origin of the spatial congruency effects obtained by applying the mouse-based paradigm remains vague. We suggest two possible mechanisms. Firstly, the congruency effects could have originated from the correspondence of spatial information conveyed by the linguistic stimuli (e.g., "worm" is related to a lower vertical location) and planning a mouse movement associated with visual action effects (i.e., a word movement) towards the lower or upper end of the screen. This approach is clearly in line with the idea of ideomotor theory, according to which representing anticipations of action effects is crucial to action planning (for overviews, see Badets et al., 2016; Hommel et al., 2001; Shin et al., 2010). Of course, it may be argued that the participants in lab-based paradigms were also exposed to visual action effects as they should have seen their arms moving upwards and downwards. For the present study, however, we can rule out the possibility that actual upward and downward arm movements caused the observed spatial congruency effects. Secondly, it could be that participants—despite conducting forward and backward mouse movements on the horizontal plane—internally re-coded their responses as "up" and "down" movements (for related suggestions see, for example, Brass et al., 2003; Eder & Rothermund, 2008). Thus, the congruency effects would result from the correspondence of the spatial information conveyed by the linguistic stimuli and an internal verbal response code. This interpretation is also relevant and even more important in the context of prior lab-based research using vertical responses movements as participants could have assigned verbal codes ("up" vs. "down") to their responses without the need of any re-coding.

In contrast to previous lab-based studies using vertical response devices with four keys (e.g., Lachmair et al., 2011), we observed spatial congruency effects not only for response times, but also for movement times. This might be due to several reasons. For instance, when operating vertical response devices with four keys (see Fig. 1), participants started trials by holding both middle keys, with each hand pressing one of the keys. They were then required to complete their response decision prior to releasing the appropriate middle key and initiating the vertical arm movement, as releasing the wrong middle key caused an error feedback. In the current studies, however, participants controlled the mouse with a single hand, which made it possible to activate the response hand before finishing the decision on the response direction. Interestingly, Vogt et al. (2019) made use of a vertical response device with a single middle key (see Fig. 1) when investigating language-space associations in children. Crucially, they also obtained a congruency effect on movement times, indicating that the described changes regarding the setup may indeed have an impact on movement times. In addition, requirements of the mouse-based paradigm with respect to response execution should be considered. For example, mouse movements could have been conducted less ballistically than response movements on vertical response devices used in lab-based setups, thus possibly being more sensitive to a correspondence of linguistic stimuli and response direction.

In the current research, we focused on collecting and evaluating data on response times, movement times, and response correctness to match our dependent variables with those measures that have typically been considered in prior lab-based research using vertical response movements to investigate language-space associations (e.g., Dudschig et al., 2015; Lachmair et al., 2011; Thornton et al., 2013). Potential future research applying our mouse-based paradigm could have a systematic look at the vast range of other metrics that have been tried and tested in the context of mouse-tracking designs, such as entropy (an indicator for movement complexity) or trajectory curvature (for a thorough overview see Wirth et al., 2020). In particular, exploiting temporally continuous real-time measures may help to gain insights into the dynamics of the investigated spatial congruency effects.

Importantly, our web-suited paradigm will be of interest to a wide range of research areas. For instance, *Conceptual Metaphor Theory* (e.g., Lakoff & Johnson, 1980, 1999) proposes that abstract concepts (e.g., "freedom, "democracy", "justice", and "love") are mentally represented by a mapping onto concrete domains that can be experienced physically. In this context, it has frequently been suggested that abstract concepts are mapped onto the spatial domain. Especially vertical space plays a crucial role for grounding concepts such as morality (e.g., Hill & Lapsley, 2009; Zhai et al., 2018), divinity (e.g., Chasteen et al., 2010; Meier et al., 2007), power (e.g., Jiang & Henley, 2012; Schubert, 2005; Wu et al., 2016; Zanolie et al., 2012), and valence (e.g., Ansorge et al., 2013; Meier & Robinson, 2004; Santiago et al., 2012). Of course, our paradigm could also easily be modified to investigate horizontal spatial associations. For example, such spatial associations are of great importance in the research on the mental representation of time (e.g., Santiago et al., 2007; Ulrich & Maienborn, 2010; Weger & Pratt, 2008). Moreover, horizontal spatial associations play a significant role in the research area of *body-specificity*, which deals with disentangling cultural, linguistic, and bodily influences with respect to the grounding of abstract concepts (e.g., Casasanto, 2009; Casasanto & Henetz, 2012; Casasanto & Jasmin, 2010; de la Fuente et al., 2015; de la Vega et al., 2012; Li & Cao, 2019). Likewise, research on the spatial numerical association of response code (SNARC) effect is inherently based on the idea of spatial grounding (see Fischer & Shaki, 2014, for a recent review). Our paradigm might be particularly useful for the web-based investigation of the vertical SNARC effect (e.g., Gevers et al., 2006; Ito & Hatta, 2004; Schwarz & Keus, 2004).

In sum, the mouse-based paradigm establishes a reliable setup for replacing vertical response movements that typically require special response devices. The method is easy to implement, allows web-based data collection, and may also facilitate future lab-based research. Of course, the paradigm can easily be adapted for investigating associations in other spatial dimensions. It thus has the potential to be a valuable tool for research across a wide range of domains interested in spatial associations.

Appendix

Experimental stimuli used in Experiments 1 to 3

Experiment 1: English location words

abyss; canyon; carpet; cellar; clover; depth; ditch; diver; earth; floor; foot; grass; grave; ground; hell; mole; mouse; rails; river; root; sidewalk; soil; sole; stone; street; subfont; submarine; subway; swamp; tunnel; underworld; worm.

alps; balloon; bird; castle; ceiling; cloud; comet; crown; eagle; gable; gallery; hawk; height; highlands; hill; kite; maximum; moon; mountains; nest; plane; planet; roof; satellite; sky; skyscraper; star; summit; sun; top; tower; universe.

Experiment 2: German location words

Abgrund; Schlucht; Teppich; Keller; Klee; Tiefe; Graben; Taucher; Erdreich; Fußboden; Fuß; Gras; Grab; Boden; Hölle; Maulwurf; Maus; Schienen; Fluss; Wurzel; Gehweg; Erde; Sohle; Stein; Straße; Untergrund; U-Boot; U-Bahn; Sumpf; Tunnel; Unterwelt; Wurm.

Alpen; Ballon; Vogel; Burg; Decke; Wolke; Komet; Krone; Adler; Giebel; Empore; Falke; Höhe; Hochland; Berg; Drachen; Höhepunkt; Mond; Gebirge; Nest; Flugzeug; Planet; Dach; Satellit; Himmel; Hochhaus; Stern; Gipfel; Sonne; Spitze; Turm; Weltall.

Experiment 3: German posture-specific emotional valence words

aufgeregt (excited); begeistert (enthusiastic); freudvoll (joyful); gutgelaunt (good-humored); stolz (proud); freudig (joyous); ekstatisch (ecstatic); erheitert (exhilarated); erfreut (pleased); erleichtert (relieved); ermutigt (encouraged); optimistisch (optimistic); beschwingt (elated); glücklich (happy); jubelnd (jubilant); strahlend (radiant); fröhlich (cheerful); vital (lively); triumphierend (triumphant); vergnügt (jolly).

traurig (sad); bedrückt (glum); demütig (humble); bekümmert (concerned); betrübt (saddened); deppresiv (depressive); erschöpft (exhausted); elend (miserable); gebrochen (broken); entmutigt (discouraged); pessimistisch (pessimistic); leidvoll (sorrowful); trübsinnig (blue); schwermütig (melancholy); enttäuscht (disappointed); verzweifelt (desperate); trauernd (grieving); ohnmächtig (helpless); deprimiert (depressed); trostlos (dismal).

English and German location words were derived from the work of Lachmair et al. (2011) and are counterparts of each other (see the Method sections). Posture-specific emotional valence words were taken from Dudschig et al. (2015); English translations are given in parentheses

Acknowledgements We thank Guo Yu for her participation in the piloting phase of the project and Roland Pfister and Robert Wirth for valuable comments on a previous version of this article.

Funding Open Access funding enabled and organized by Projekt DEAL. Our work was supported by the Deutsche Forschungsgemeinschaft (DFG)—Project IDs 75650358 and 419433647.

Availability of data and material All data sets are publicly available at https://doi.org/10.5281/zenodo.4557024. All materials are mentioned and described in the Method sections.

Code availability The original and an updated version of the custom-made jsPsych plugin for implementing the mouse movement task can be found at https://doi.org/10.5281/zenodo.4557024. Analysis scripts will be uploaded at the same place upon publication.

Declarations

Conflict of interest We have no conflicts of interest to disclose.

Ethics approval Ethics approval for our research was obtained from the Ethics Committee for Psychological Research at the University of Tübingen (Identifier: 2018_0831_132).

Consent to participate/consent for publication All individual participants included in the study gave informed consent.

References

Ahlberg, D. K., Bischoff, H., Kaup, B., Bryant, D., & Strozyk, J. V. (2018). Grounded cognition: Comparing language × space interactions in first language and second language. *Applied Psycholinguistics, 39*(2), 437–459. https://doi.org/10.1017/S0142716417000042X

Ansorge, U., Khalid, S., & König, P. (2013). Space-valence priming with subliminal and supraliminal words. *Frontiers in Psychology, 4*, 81. https://doi.org/10.3389/fpsyg.2013.00081

Aust, F., Diedenhofen, B., Ullrich, S., & Musch, J. (2013). Seriousness checks are useful to improve data validity in online research. *Behavior Research Methods, 45*(2), 527–535. https://doi.org/10.3758/s13428-012-0265-2

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Badets, A., Koch, I., & Philipp, A. M. (2016). A review of ideomotor approaches to perception, cognition, action, and language: Advancing a cultural recycling hypothesis. *Psychological Research Psychologische Forschung, 80*(1), 1–15. https://doi.org/10.1007/s00426-014-0643-8

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*(4), 577–660. https://doi.org/10.1017/S0140525X99002149

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

Birnbaum, M. H. (2004). Human research and data collection via the Internet. *Annual Review of Psychology, 55*, 803–832. https://doi.org/10.1146/annurev.psych.55.090902.141601

Brass, M., Ruge, H., Meiran, N., Rubin, O., Koch, I., Zysset, S., Prinz, W., & Von Cramon, D. Y. (2003). When the same response has different meanings: Recoding the response meaning in the lateral prefrontal cortex. *NeuroImage, 20*(2), 1026–1031. https://doi.org/10.1016/S1053-8119(03)00357-4

Buchanan, T. (2002). Online assessment: Desirable or dangerous? *Professional Psychology: Research and Practice, 33*(2), 148–154. https://doi.org/10.1037/0735-7028.33.2.148

Buchanan, T., & Williams, J. E. (2010). Ethical issues in psychological research on the Internet. In S. D. Gosling & J. A. Johnson (Eds.), *Advanced methods for conducting online behavioral research* (pp. 255–271). American Psychological Association. https://doi.org/10.1037/12076-016

Campbell, J. I. D., & Thompson, V. A. (2012). MorePower 6.0 for ANOVA with relational confidence intervals and Bayesian analysis. *Behavior Research Methods, 44*(4), 1255–1265. https://doi.org/10.3758/s13428-012-0186-0

Casasanto, D. (2009). Embodiment of abstract concepts: Good and bad in right- and left-handers. *Journal of Experimental Psychology: General, 138*(3), 351–367. https://doi.org/10.1037/a0015854

Casasanto, D., & Henetz, T. (2012). Handedness shapes children's abstract concepts. *Cognitive Science, 36*(2), 359–372. https://doi.org/10.1111/j.1551-6709.2011.01199.x

Casasanto, D., & Jasmin, K. (2010). Good and bad in the hands of politicians: Spontaneous gestures during positive and negative speech. *PLoS ONE, 5*(7), e11805. https://doi.org/10.1371/journal.pone.0011805

Chasteen, A. L., Burdzy, D. C., & Pratt, J. (2010). Thinking of God moves attention. *Neuropsychologia, 48*(2), 627–630. https://doi.org/10.1016/j.neuropsychologia.2009.09.029

Corley, M., & Scheepers, C. (2002). Syntactic priming in English sentence production: Categorical and latency evidence from an

Internet-based study. *Psychonomic Bulletin and Review, 9*(1), 126–131. https://doi.org/10.3758/BF03196267

Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PLoS ONE, 8*(3), e57410. https://doi.org/10.1371/journal.pone.0057410

Dale, R., Kehoe, C., & Spivey, M. J. (2007). Graded motor responses in the time course of categorizing atypical exemplars. *Memory and Cognition, 35*(1), 15–28. https://doi.org/10.3758/BF03195938

de la Fuente, J., Casasanto, D., Román, A., & Santiago, J. (2015). Can culture influence body-specific associations between space and valence? *Cognitive Science, 39*(4), 821–832. https://doi.org/10.1111/cogs.12177

de la Vega, I., de Filippis, M., Lachmair, M., Dudschig, C., & Kaup, B. (2012). Emotional valence and physical space: Limits of interaction. *Journal of Experimental Psychology: Human Perception and Performance, 38*(2), 375–385. https://doi.org/10.1037/a0024979

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods, 47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

Dudschig, C., & Kaup, B. (2017). Is it all task-specific? The role of binary responses, verbal mediation, and saliency for eliciting language-space associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(2), 259–270. https://doi.org/10.1037/xlm0000297

Dudschig, C., & Kaup, B. (2020). Negation as conflict: Conflict adaptation following negating vertical spatial words. *Brain and Language, 210*, 104842. https://doi.org/10.1016/j.bandl.2020.104842

Dudschig, C., Lachmair, M., de la Vega, I., De Filippis, M., & Kaup, B. (2012). Do task-irrelevant direction-associated motion verbs affect action planning? Evidence from a Stroop paradigm. *Memory and Cognition, 40*(7), 1081–1094. https://doi.org/10.3758/s13421-012-0201-9

Dudschig, C., de la Vega, I., & Kaup, B. (2014). Embodiment and second-language: Automatic activation of motor responses during processing spatially associated L2 words and emotion L2 words in a vertical Stroop paradigm. *Brain and Language, 132*, 14–21. https://doi.org/10.1016/j.bandl.2014.02.002

Dudschig, C., de la Vega, I., & Kaup, B. (2015). What's up? Emotion-specific activation of vertical space during language processing. *Acta Psychologica, 156*, 143–155. https://doi.org/10.1016/j.actpsy.2014.09.015

Duffy, M. E. (2002). Methodological issues in web-based research. *Journal of Nursing Scholarship, 34*(1), 83–88. https://doi.org/10.1111/j.1547-5069.2002.00083.x

Eder, A. B., & Rothermund, K. (2008). When do motor behaviors (mis) match affective stimuli? An evaluative coding view of approach and avoidance reactions. *Journal of Experimental Psychology: General, 137*(2), 262–281. https://doi.org/10.1037/0096-3445.137.2.262

Emery, K. (2014). So you want to do an online study: Ethics considerations and lessons learned. *Ethics and Behavior, 24*(4), 293–303. https://doi.org/10.1080/10508422.2013.860031

Fanelli, D., & Ioannidis, J. P. A. (2013). US studies may overestimate effect sizes in softer research. *Proceedings of the National Academy of Sciences of the United States of America, 110*(37), 15031–15036. https://doi.org/10.1073/pnas.1302997110

Fischer, M. H., & Shaki, S. (2014). Spatial associations in numerical cognition-From single digits to arithmetic. *Quarterly Journal of Experimental Psychology, 67*(8), 1461–1483. https://doi.org/10.1080/17470218.2014.927515

Fox, L. A., & Shor, R. E. (1976). Semantic gradients and interference with sorting according to color, spatial position, and numerosity. *Bulletin of the Psychonomic Society, 7*(2), 187–189. https://doi.org/10.3758/BF03337163

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*(1), 59–65. https://doi.org/10.1037/h0031850

Freeman, J. B. (2018). Doing psychological science by hand. *Current Directions in Psychological Science, 27*(5), 315–323. https://doi.org/10.1177/0963721417746793

Freeman, J. B., Pauker, K., Apfelbaum, E. P., & Ambady, N. (2010). Continuous dynamics in the real-time perception of race. *Journal of Experimental Social Psychology, 46*(1), 179–185. https://doi.org/10.1016/j.jesp.2009.10.002

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*(3), 221–233. https://doi.org/10.1016/j.actpsy.2005.11.004

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin and Review, 9*(3), 558–565. https://doi.org/10.3758/BF03196313

Gosling, S. D., & Mason, W. (2015). Internet research in psychology. *Annual Review of Psychology, 66*, 877–902. https://doi.org/10.1146/annurev-psych-010814-015321

Günther, F., Dudschig, C., & Kaup, B. (2018). Symbol grounding without direct experience: Do words inherit sensorimotor activation from purely linguistic context? *Cognitive Science, 42*(S2), 336–374. https://doi.org/10.1111/cogs.12549

Günther, F., Nguyen, T., Chen, L., Dudschig, C., Kaup, B., & Glenberg, A. M. (2020). Immediate sensorimotor grounding of novel concepts learned from language alone. *Journal of Memory and Language, 115*, 104172. https://doi.org/10.1016/j.jml.2020.104172

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*(2–3), 61–135. https://doi.org/10.1017/S0140525X0999152X

Hilbig, B. E. (2016). Reaction time effects in lab- versus web-based research: Experimental evidence. *Behavior Research Methods, 48*(4), 1718–1724. https://doi.org/10.3758/s13428-015-0678-9

Hill, P. L., & Lapsley, D. K. (2009). The ups and downs of the moral personality: Why it's not so black and white. *Journal of Research in Personality, 43*(3), 520–523. https://doi.org/10.1016/j.jrp.2009.01.014

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences, 24*(5), 849–878. https://doi.org/10.1017/S0140525X01000103

Horwitz, R., Kreuter, F., & Conrad, F. (2017). Using mouse movements to predict web survey response difficulty. *Social Science Computer Review, 35*(3), 388–405. https://doi.org/10.1177/0894439315626360

Ito, Y., & Hatta, T. (2004). Spatial structure of quantitative representation of numbers: Evidence from the SNARC effect. *Memory and Cognition, 32*(4), 662–673. https://doi.org/10.3758/BF03195857

Jiang, M., & Henley, T. B. (2012). Power and spatial relations. *Journal of Cognitive Psychology, 24*(7), 829–835. https://doi.org/10.1080/20445911.2012.702749

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

Koch, S. C., Glawe, S., & Holt, D. V. (2011). Up and down, front and back: Movement and meaning in the vertical and sagittal axes. *Social Psychology, 42*(3), 214–224. https://doi.org/10.1027/1864-9335/a000065

Krantz, J. H., & Dalal, R. (2000). Validity of web-based psychological research. In M. H. Birnbaum (Ed.), *Psychological experiments on* *the Internet* (pp. 35–60). Academic Press. https://doi.org/10.1016/b978-012099980-4/50003-4

Kraut, R., Olson, J., Banaji, M., Bruckman, A., Cohen, J., & Couper, M. (2004). Psychological research online: Report of Board of Scientific Affairs' Advisory Group on the conduct of research on the Internet. *American Psychologist, 59*(2), 105–117. https://doi.org/10.1037/0003-066X.59.2.105

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

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 and Review, 18*(6), 1180–1188. https://doi.org/10.3758/s13423-011-0158-x

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. University of Chicago Press.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to western thoughts*. Basic Books.

Li, H., & Cao, Y. (2019). The body in religion: The spatial mapping of valence in Tibetan practitioners of Bön. *Cognitive Science, 43*(4), e12728. https://doi.org/10.1111/cogs.12728

Linnman, C., Carlbring, P., Åhman, Å., Andersson, H., & Andersson, G. (2006). The Stroop effect on the Internet. *Computers in Human Behavior, 22*(3), 448–455. https://doi.org/10.1016/j.chb.2004.09.010

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language, 94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

McGraw, K. O., Tew, M. D., & Williams, J. E. (2000). The integrity of web-delivered experiments: Can you trust the data? *Psychological Science, 11*(6), 502–506. https://doi.org/10.1111/1467-9280.00296

Meier, B. P., & Robinson, M. D. (2004). Why the sunny side is up: Associations between affect and vertical position. *Psychological Science, 15*(4), 243–247. https://doi.org/10.1111/j.0956-7976.2004.00659.x

Meier, B. P., Hauser, D. J., Robinson, M. D., Friesen, C. K., & Schjeldahl, K. (2007). What's "up" with God? Vertical space as a representation of the divine. *Journal of Personality and Social Psychology, 93*(5), 699–710. https://doi.org/10.1037/0022-3514.93.5.699

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology, 4*(2), 61–64. https://doi.org/10.20982/tqmp.04.2.p061

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 and Cognition, 35*(5), 1151–1161. https://doi.org/10.3758/BF03193485

Neath, I., Earle, A., Hallett, D., & Surprenant, A. M. (2011). Response time accuracy in Apple Macintosh computers. *Behavior Research Methods, 43*(2), 353–362. https://doi.org/10.3758/s13428-011-0069-9

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). E-research: Ethics, security, design, and control in psychological research on the Internet. *Journal of Social Issues, 58*(1), 161–176. https://doi.org/10.1111/1540-4560.00254

Öttl, B., Dudschig, C., & Kaup, B. (2017). Forming associations between language and sensorimotor traces during novel word learning. *Language and Cognition, 9*(1), 156–171. https://doi.org/10.1017/langcog.2016.5

Reimers, S., & Stewart, N. (2015). Presentation and response timing accuracy in Adobe Flash and HTML5/JavaScript web experiments. *Behavior Research Methods, 47*(2), 309–327. https://doi.org/10.3758/s13428-014-0471-1

Reips, U.-D. (2000). The web experiment method: Advantages, disadvantages, and solutions. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 89–117). Academic Press. https://doi.org/10.1016/B978-012099980-4/50005-8

Reips, U.-D. (2002a). Internet-based psychological experimenting: Five dos and five don'ts. *Social Science Computer Review, 20*(3), 241–249. https://doi.org/10.1177/089443930202000302

Reips, U.-D. (2002b). Standards for Internet-based experimenting. *Experimental Psychology, 49*(4), 243–256. https://doi.org/10.1026//1618-3169.49.4.243

Reips, U.-D. (2009). Internet experiments: Methods, guidelines, metadata. *Human Vision and Electronic Imaging XIV, Proceedings of SPIE, 7240*, 724008.

Rhodes, S. D., Bowie, D. A., & Hergenrather, K. C. (2003). Collecting behavioural data using the world wide web: Considerations for researchers. *Journal of Epidemiology and Community Health, 57*(1), 68–73. https://doi.org/10.1136/jech.57.1.68

Santiago, J., Lupiáñez, J., Pérez, E., & Funes, M. J. (2007). Time (also) flies from left to right. *Psychonomic Bulletin and Review, 14*(3), 512–516. https://doi.org/10.3758/BF03194099

Santiago, J., Ouellet, M., Román, A., & Valenzuela, J. (2012). Attentional factors in conceptual congruency. *Cognitive Science, 36*(6), 1051–1077. https://doi.org/10.1111/j.1551-6709.2012.01240.x

Schneider, D. W. (2020). Alertness and cognitive control: Interactions in the spatial Stroop task. *Attention, Perception, and Psychophysics, 82*(5), 2257–2270. https://doi.org/10.3758/s13414-020-01993-5

Schoemann, M., O'Hora, D., Dale, R., & Scherbaum, S. (2021). Using mouse cursor tracking to investigate online cognition: Preserving methodological ingenuity while moving toward reproducible science. *Psychonomic Bulletin and Review, 28*(3), 766–787. https://doi.org/10.3758/s13423-020-01851-3

Schubert, T. W. (2005). Your highness: Vertical positions as perceptual symbols of power. *Journal of Personality and Social Psychology, 89*(1), 1–21. https://doi.org/10.1037/0022-3514.89.1.1

Schütt, E., Kaup, B., & Dudschig, C. (2021). *Investigating language-space associations: Can response movements be replaced by visual action effects? [Manuscript in preparation]*. University of Tübingen.

Schwarz, W., & Keus, I. M. (2004). Moving the eyes along the mental number line: Comparing SNARC effects with saccadic and manual responses. *Perception and Psychophysics, 66*(4), 651–664. https://doi.org/10.3758/BF03194909

Seelye, A., Hagler, S., Mattek, N., Howieson, D. B., Wild, K., Dodge, H. H., & Kaye, J. A. (2015). Computer mouse movement patterns: A potential marker of mild cognitive impairment. *Alzheimer's and Dementia: Diagnosis, Assessment and Disease Monitoring, 1*(4), 472–480. https://doi.org/10.1016/j.dadm.2015.09.006

Semmelmann, K., & Weigelt, S. (2017). Online psychophysics: Reaction time effects in cognitive experiments. *Behavior Research Methods, 49*(4), 1241–1260. https://doi.org/10.3758/s13428-016-0783-4

Shin, Y. K., Proctor, R. W., & Capaldi, E. J. (2010). A review of contemporary ideomotor theory. *Psychological Bulletin, 136*(6), 943–974. https://doi.org/10.1037/a0020541

Simcox, T., & Fiez, J. A. (2014). Collecting response times using Amazon Mechanical Turk and Adobe Flash. *Behavior Research Methods, 46*(1), 95–111. https://doi.org/10.3758/s13428-013-0345-y

Simonsohn, U. (2015). Small telescopes: Detectability and the evaluation of replication results. *Psychological Science, 26*(5), 559–569. https://doi.org/10.1177/0956797614567341

Stewart, N., Chandler, J., & Paolacci, G. (2017). Crowdsourcing samples in cognitive science. *Trends in Cognitive Sciences, 21*(10), 736–748. https://doi.org/10.1016/j.tics.2017.06.007

Stillman, P. E., Shen, X., & Ferguson, M. J. (2018). How mouse-tracking can advance social cognitive theory. *Trends in Cognitive*

Sciences, 22(6), 531–543. https://doi.org/10.1016/j.tics.2018.03.012

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(6), 643–662. https://doi.org/10.1037/h0054651

Thornton, T., Loetscher, T., Yates, M. J., & Nicholls, M. E. R. (2013). The highs and lows of the interaction between word meaning and space. *Journal of Experimental Psychology: Human Perception and Performance, 39*(4), 964–973. https://doi.org/10.1037/a0030467

Ulrich, R., & Maienborn, C. (2010). Left-right coding of past and future in language: The mental timeline during sentence processing. *Cognition, 117*(2), 126–138. https://doi.org/10.1016/j.cognition.2010.08.001

Vicovaro, M., & Dalmaso, M. (2021). Is "heavy" up or down? Testing the vertical spatial representation of weight. *Psychological Research Psychologische Forschung, 85*(3), 1183–1200. https://doi.org/10.1007/s00426-020-01309-0

Vogt, A., Kaup, B., & Dudschig, C. (2019). When words are upside down: Language-space associations in children and adults. *Journal of Experimental Child Psychology, 186*, 142–158. https://doi.org/10.1016/j.jecp.2019.06.001

Weger, U. W., & Pratt, J. (2008). Time flies like an arrow: Space-time compatibility effects suggest the use of a mental timeline. *Psychonomic Bulletin and Review, 15*(2), 426–430. https://doi.org/10.3758/PBR.15.2.426

Wirth, R., Foerster, A., Kunde, W., & Pfister, R. (2020). Design choices: Empirical recommendations for designing two-dimensional finger-tracking experiments. *Behavior Research Methods, 52*(6), 2394–2416. https://doi.org/10.3758/s13428-020-01409-0

Woods, A. T., Velasco, C., Levitan, C. A., Wan, X., & Spence, C. (2015). Conducting perception research over the Internet: A tutorial review. *PeerJ, 3*, e1058. https://doi.org/10.7717/peerj.1058

Wu, X., Jia, H., Wang, E., Du, C., Wu, X., & Dang, C. (2016). Vertical position of Chinese power words influences power judgments: Evidence from spatial compatibility task and event-related potentials. *International Journal of Psychophysiology, 102*, 55–61. https://doi.org/10.1016/j.ijpsycho.2016.03.005

Zanolie, K., van Dantzig, S., Boot, I., Wijnen, J., Schubert, T. W., Giessner, S. R., & Pecher, D. (2012). Mighty metaphors: Behavioral and ERP evidence that power shifts attention on a vertical dimension. *Brain and Cognition, 78*(1), 50–58. https://doi.org/10.1016/j.bandc.2011.10.006

Zhai, D., Guo, Y., & Lu, Z. (2018). A dual mechanism of cognition and emotion in processing moral-vertical metaphors. *Frontiers in Psychology, 9*, 1554. https://doi.org/10.3389/fpsyg.2018.01554

Zwaan, R. A., & Madden, C. J. (2005). Embodied sentence comprehension. In D. Pecher & R. A. Zwaan (Eds.), *Grounding cognition: The role of perception and action in memory, language, and thinking* (pp. 224–245). Cambridge University Press. https://doi.org/10.1017/CBO9780511499968.010

Appendix B: Article 2

To cite this article

Schütt, E., Kaup, B., & Dudschig, C. (2022). Investigating vertical language-space associations: Can visual action effects induce congruency effects? [Unpublished manuscript]. Department of Psychology, University of Tübingen.

Investigating Vertical Language-Space Associations:

Can Visual Action Effects Induce Congruency Effects?

Emanuel Schütt, Barbara Kaup, & Carolin Dudschig

Department of Psychology, Language and Cognition Research Group, University of Tübingen

Author Note

All data and R analysis scripts can be found at https://doi.org/10.5281/zenodo.7198737. We have no conflicts of interest to disclose. This research was supported by the Deutsche Forschungsgemeinschaft (DFG) – Project IDs 75650358 and 419433647.

Correspondence concerning this research article should be addressed to Emanuel Schütt, University of Tübingen, Department of Psychology, Language and Cognition Research Group, Schleichstr. 4, 72076 Tübingen, Germany. Email: emanuel.schuett@uni-tuebingen.de

Abstract

Language comprehenders tend to mentally simulate spatial experiences when processing words denoting entities that are physically or metaphorically associated with a lower or upper vertical location (e.g., “bird” vs. “worm”; “love” vs. “hate”). This results in spatial congruency effects. For instance, visually perceiving implicit locations words (e.g., “star” vs. “grass”) at congruent (vs. incongruent) vertical locations leads to faster response times. The current research aimed to examine whether the mere anticipation of visual action effects related to lower or upper vertical space can provoke comparable congruency effects. In a series of experiments, implicit location words were presented centered on the screen. Participants provided simple keypress responses producing immediate visual action effects at or towards the lower or upper screen location. We did not observe spatial congruency effects in a reliable manner. Importantly, this was also true when vertical space was made more salient by adapting contextual task conditions (e.g., in terms of introducing response keys located on the front-back axis of the keyboard). We conclude that visual feedback in vertical space following non-vertical stationary responses – thus only allowing the anticipation of spatially matching or mismatching action effects – appears not to be sufficient to establish the congruency effects under investigation. In contrast, however, a direct real-time coupling of visual feedback in vertical space and non-vertical responses can reliably evoke these effects (see Experiment 1).

Keywords: language-space associations, embodiment, implicit location words, ideomotor theory, visual action effects

Investigating Vertical Language-Space Associations:
Can Visual Action Effects Induce Congruency Effects?

According to the traditional approach in cognitive and neuroscientific research, higher cognitive processes (e.g., language; memory; thinking; problem-solving) are realized in heavily specialized brain areas by means of abstract representations (e.g., Fodor 1975; Anderson 1983; Pylyshyn 1984). Crucially, these brain areas are thought to be separate from those dealing with lower-level sensorimotor processes. However, there is an ever-increasing amount of behavioral and neuroscientific evidence indicating that higher cognitive processes should at least partially rely on sensorimotor brain areas and concrete representations (e.g., Glenberg and Kaschak 2002; Zwaan et al. 2002; Pecher et al. 2003; Kan et al. 2003; Borghi et al. 2004; Pulvermüller et al. 2005; Simmons et al. 2005; Bub et al. 2008). Consequently, the importance of approaches emphasizing a close link of sensorimotor systems and cognitive processes – commonly referred to as grounded or embodied approaches – has grown tremendously (Chatterjee 2010).

Embodied approaches specifically addressing language comprehension (e.g., Barsalou 1999; Glenberg and Kaschak 2002; Zwaan and Madden 2005) argue that linguistic constructions, such as words, often co-occur with the entities they refer to. As a result, it is assumed that these linguistic constructions become associated with patterns of activation in sensorimotor brain areas evoked by experiencing the respective entities. Later, when reading or hearing a word, a phrase, or a sentence, the corresponding patterns of sensorimotor brain activation are expected to be re-activated to mentally simulate and thus understand the meaning of the linguistic input.

An important branch of the research on embodied language comprehension has focused on investigating vertical language-space associations. This type of studies specifically aims at revealing whether and under what circumstances language comprehension involves mentally simulating spatial experiences. For this purpose, various groups of words have been considered.

Firstly, there are the words referring to concrete physical entities that are typically associated with a lower or an upper vertical location (*implicit location words*; e.g., “worm” and “eagle”, respectively). Crucially, several studies have provided conclusive evidence that encountering implicit location words indeed seems to trigger the activation of corresponding spatial features. For example, making semantic relatedness and word category judgments was faster when the spatial presentation of implicit location words matched the typical vertical location of their referents (e.g., Zwaan and Yaxley 2003; Šetić and Domijan 2007). Likewise, various experiments showed that implicit location words can shift visual attention in accordance with the spatial feature they convey (e.g., Estes et al. 2008; Dudschig et al. 2012a; Gozli et al. 2013). Finally, encountering implicit location words also affects subsequent motor responses. For instance, Lachmair et al. (2011, Experiment 2) presented their participants with implicit location words, such as “submarine” and “satellite”. Importantly, these words were centrally displayed in one of four different font colors. Depending on the font color of the word, participants performed either a downward or an upward arm movement. Although this task did not require any access to word meaning, response times were faster when the direction of the response movement matched the typical vertical location of the word’s referent. This strongly indicates that implicit location words automatically re-activate spatial experiences associated with their referents. Converging results were obtained in many other studies using the same or a similar vertical motor response task (e.g., Thornton et al. 2013; Dudschig et al. 2014, 2015; Öttl et al. 2017; Vogt et al. 2019). This also holds true when direction-associated motion verbs (e.g., “to dive” and “to jump”) and spatial prepositions (e.g., “above” and “below”) served as experimental stimuli (Dudschig et al. 2012b; Ahlberg et al. 2018).

Furthermore, vertical language-space associations have also become crucial to embodied approaches on language comprehension because they were employed in the context of answering the challenging question how words describing abstract concepts (e.g., “justice”, "freedom", or "love") – which do not possess a physical referent in the world and thus cannot be experienced directly – might be grounded in the sensorimotor system. The probably most influential attempt to solving this issue – the *Conceptual Metaphor Theory* (e.g., Lakoff and Johnson 1980, 1999) – proposes that abstract concepts are mentally represented by a mapping onto concrete domains that can be experienced directly. Particularly, vertical space has turned out to be a central domain for grounding abstract concepts such as valence (e.g., Meier and Robinson 2004; Santiago et al. 2012; Ansorge et al. 2013), power (e.g., Schubert 2005; Jiang and Henley 2012; Zanolie et al. 2012; Wu et al. 2016), morality (e.g., Hill and Lapsley 2009; Zhai et al. 2018), and divinity (e.g., Meier et al. 2007; Chasteen et al. 2010). As in the case of implicit locations words, numerous studies investigating the relationship of words referring to abstract concepts and vertical space have made use of behavioral paradigms stemming rather from the visual or the motor domain. For instance, Zanolie et al. (2012) centrally presented words indicating powerless and powerful persons (e.g., "servant" and "king"). Participants performed a power evaluation task, before a target stimulus appeared at the lower or upper end of the screen. Crucially, responses to targets were faster when the target location matched the power evaluation of the preceding word (i.e., powerless and lower target location; powerful and upper target location). In a very similar way, Chasteen et al. (2010) investigated whether divinity-related words orientate visual attention in vertical space. In each trial of their experiment, either a divinity-related word (e.g., "Devil" and "God") or a filler word (e.g., "Door" and "Pencil") was centrally presented on the screen. Then, participants decided whether they saw a divinity-related word. Only if so, they should proceed and respond to a target stimulus displayed in a lower or an upper vertical location. Responses to targets were faster when they appeared at a location compatible with the divinity-related words (e.g., "Devil" and lower location; "God" and upper location). Likewise, behavioral studies from the motor domain highly relied on tasks analogous to those considered for examining the link of implicit location words and vertical space. For example, Brookshire et al. (2010) looked at words with positive and negative valence (e.g., “wealthy”, “virtuous”, and “joy”; “poor”, “evil”, and “disgust”) by using a Stroop-like vertical motor response task. For responding to the font color of valence words, participants performed downward and upward arm movements. Again, a spatial congruency effect occurred (i.e., faster downward responses to negative words; faster upward responses to positive words). However, this effect was sensitive to task characteristics: It was absent when words were repeated and when attention was distracted from word meaning. In a similar vein, Dudschig et al. (2015) showed that the emergence of spatial congruency effects in Stroop-like vertical motor response tasks depends on specific features of the presented valence words. Consequently, vertical spatial simulations induced by valence words seem to be rather fragile and less automatic.

In sum, investigating word-based vertical language-space associations has played a core role in research on embodied language comprehension. Behavioral studies have particularly dealt with studying the relationship of specific groups of words and vertical space by employing tasks from the visual and the motor domain. Interestingly, congruency effects obtained with tasks from the visual domain tend to vary, as seen in the occurrence of both interference (e.g., Richardson et al. 2003; Estes et al. 2008) and facilitation (e.g., Zwaan and Yaxley 2003; Šetić and Domijan 2007). In contrast, tasks from the motor domain appear to produce consistently facilitation (e.g., Brookshire et al. 2010; Lachmair et al. 2011; Dudschig et al. 2014, 2015; Öttl et al. 2017; Vogt et al. 2019). However, these tasks usually show a confound, which – to our knowledge – has not been addressed yet: Participants typically receive some sort of visual feedback stemming from perceiving either their vertical hand movements or the vertically arranged response buttons being pressed. This aspect might also be relevant to the emergence of congruency effects in the context of tasks from the motor domain.

In particular, it is interesting to consider this aspect on the basis of ideomotor theory, which states that anticipated action effects are integrated in action planning (for reviews, see

Hommel et al. 2001; Shin et al. 2010; Badets et al. 2016). Interestingly, such mechanisms could have contributed to the emergence of spatial congruency effects obtained in experiments relying on vertical response movements. For instance, participants performing the Stroop-like motor response task (e.g., Brookshire et al. 2010; Lachmair et al. 2011; Dudschig et al. 2014, 2015; Öttl et al. 2017; Ahlberg et al. 2018; Vogt et al. 2019) responded to the font color of words by executing downward (upward) arm movements in order to press a lower (an upper) response key. Crucially, this way of responding involves several sensory events that may prompt participants to associate the lower (upper) response key with action effects at a lower (an upper) location. For example, when responding by means of pressing the lower (upper) response key, participants saw their arm moving downwards (upwards) and heard a clicking sound coming from a lower (an upper) location. Thus, it would be of interest to investigate to what extent these action effects themselves could be able to trigger congruency effects. Indeed, in a recent study, we showed that mouse movements on the horizontal plane inducing vertical movements of word stimuli (i.e., implicit location words and posture-specific valence words) reliably produce spatial congruency effects (Schütt et al. 2022). However, in this paradigm, the mouse movements and the vertical stimulus movements were coupled in a direct and real-time manner. For instance, this could have induced participants to code and process their mouse movements in terms of vertical response movements. It is therefore hardly possible to judge whether the mere anticipation of action effects associated with vertical space was the critical factor for the emergence of the observed congruency effects.

In Experiment 1, we replicated the results of the mouse movement study (Schütt et al. 2022) employing a slightly modified paradigm. We again decided to choose implicit location words as stimuli, because research has consistently shown that particularly this group of words tends to provoke reliably spatial congruency effects (e.g., Lachmair et al. 2011; Thornton et al. 2013; Dudschig et al. 2014, 2015; Öttl et al. 2017; Vogt et al. 2019). In the further experiments, we intended to examine whether mere action effects in vertical space are sufficient for eliciting spatial congruency effects. Thus, we ran a series of experiments replacing mouse movements on the horizontal plane by simple stationary keypresses generating immediate visual action effects appearing at a lower or upper screen location.

Experiment 1

In Experiment 1, we conducted a replication of the mouse movement study by Schütt et al. (2022, Experiment 2). Their participants moved their mouse on the horizontal plane to drag centrally presented implicit location words to a lower or upper target area on the screen. As soon as the word stimuli were located in one of the target areas, these were replaced by a trial-related accuracy feedback. The dragging direction was determined by the font color of the word stimuli. In the present experiment, we implemented one minor adaption to this paradigm: We refrained from providing the trial-related accuracy feedback at the final word location (i.e., at a lower or upper screen location) to rule out that this aspect critically affected the occurrence of the spatial congruency effect.

Method

Participants

For setting the number of participants, we decided to refer to the original work by Schütt et al. (2022). Noteworthy, they determined their sample size on the basis of a power analysis and the suggestion of Simonsohn (2015), which says that replications should have 2.5 times as many observations as an original study. As a result, they aimed at collecting the data of 60 participants. We thus recruited a total of 60 participants by sending an invitation email to psychology students at the University of Tübingen (partial course credit as compensation) and by making use of the online labor marketplace Prolific (£4.00 as compensation). In line with Schütt et al. (2022), we omitted participants with a low accuracy rate (less than 90% of correct responses in at least one experimental condition), which resulted in a final sample comprising 57 participants (31 males, 26 females; 46 right-handed, nine left-handed, two ambidextrous). Their ages ranged from 18 to 64 years (*M* = 28.39 years, *SD* = 9.95 years). All of them indicated to be native German speakers and gave informed consent prior to participating. The study took about 30 minutes to complete.

Apparatus and Stimuli

We programmed the experiment by means of the open-source JavaScript library jsPsych (Version 6.1.0; de Leeuw 2015) and created the mouse movement task using the custom-made jsPsych plugin provided by Schütt et al. (2022). Furthermore, participants were asked to use solely a desktop computer or a laptop together with a mouse for running the experiment.

We employed exactly the same stimuli as Schütt et al. (2022, Experiment 2). Therefore, the linguistic stimuli were determined by their list of German implicit location words. This list included 32 single words referring to entities typically encountered in the lower vertical space (down words; e.g., “Abgrund” [abyss]; “Teppich” [carpet]; “Grab” [grave]; “Maulwurf” [mole]; “Fußboden” [floor]) and 32 single words referring to entities typically encountered in the upper vertical space (up words; e.g., “Ballon” [balloon]; “Vogel” [bird]; “Wolke” [cloud]; “Hochhaus” [skyscraper]; “Flugzeug” [plane]). Crucially, it was shown that these down words and up words did not differ significantly regarding their length or frequency. Sixteen single words referring to entities not related to a specific vertical location (e.g., “Buch” [book]; “Maschine” [machine]; “Kaffee” [coffee]) were used as stimuli for the training session. Word stimuli appeared in blue (RGB: 0, 0, 255), orange (RGB: 255, 165, 0), green (RGB: 0, 128, 0), and red (RGB: 255, 0, 0) font color on a white background. The fixation cross was depicted in terms of a black plus sign.

Procedure

We closely followed the procedure of Schütt et al. (2022, Experiment 2). The participants first passed a training session. In this training session, each of the 16 training stimuli appeared once in one of the four font colors. Importantly, all font colors were displayed equally often. After completing the training session, participants received a performance feedback, including the average response time and the percentage of correct responses. Afterwards, they proceeded with the experimental session, which comprised four blocks. In each block, each down word and each up word was presented once in one of the four font colors. Moreover, each font color appeared equally often per block and each word was displayed once in each font color over the course of the experiment. We randomized the order of trials and blocks. After each block, the participants received a performance feedback and had the possibility to take a short break.

In all trials, the relevant screen area was defined by a frame (90% of the available screen width; 90% of the available screen height). Each trial started with a fixation cross located at the center of the frame (800 ms). Then – for indicating the lower and upper target area – a horizontal borderline appeared at the lower and upper end of the framed area of the screen, respectively. At the same time, the word stimulus replaced the fixation cross. Participants were asked to respond to the font color of the word by dragging it to the lower or upper target area using their computer mouse. We always mapped two font colors to one response direction. Possible color pairs (blue-orange and green-red; blue-green and orange-red; blue-red and orange-green) and the assignment of color pairs to response directions were balanced between participants by creating six different experimental versions. Trial-based correctness feedback appearing at the center of the screen was exclusively displayed in the training session (“Richtig!” [“Correct!”] vs. “Falsch!” [“Wrong!”]; black font color; 800 ms). The intertrial interval had a duration of 1500 ms.

Design and Data Analysis

The experiment implemented a 2 (referent location: up vs. down) × 2 (response direction: up vs. down) within-subjects design. In line with the study by Schütt et al. (2022), analyses were performed on response times (period from the appearance of the word until the initial movement of the word) and movement times (period from the initial movement of the word until the word crossed the borderline of the lower or upper target area). We processed and analyzed the data by using the free statistical software R (Version 3.6.2). Trials with more than a single mouse click or an incorrect response were discarded. In addition, extreme outliers were excluded (response times shorter than 100 or longer than 3000 ms and movement times longer than 1000 ms). To identify further outliers, we made use of the two-step procedure suggested by Kaup et al. (2006), which considers response differences among items and participants. Thus, we transformed the response and movement times of each participant to *z*-scores. Subsequently, we deleted trials with *z*-scores differing more than two standard deviations from the mean *z*-score of the respective word in the respective experimental condition. In sum, outlier removal reduced the data set by less than 5% (response times) and by less than 6% (movement times), respectively. We then conducted linear mixed effects analyses (see Baayen et al. 2008) employing the R packages lme4 (Version 1.1-21; Bates et al. 2015) and lmerTest (Version 3.1-1; Kuznetsova et al. 2017). Both base models (one for response times; one for movement times) included fixed effects for age, handedness, referent location, and response direction. For defining appropriate random effects structures, we referred to the proposal of Matuschek et al. (2017), which intends to balance Type I error rate and power. Some more complex models had to be skipped due to convergence issues or a singular fit. The procedure finally resulted in incorporating random intercepts for participants and items in both models. To evaluate the effect of interest, we performed likelihood ratio tests to compare each base model with a corresponding model comprising a further fixed effect for the interaction of referent location and response direction. All data and R scripts are available online (https://doi.org/10.5281/zenodo.7198737).

Results and Discussion

Table 1 gives an overview of the mean response times as a function of referent location and response direction. The model with an additional fixed effect for the interaction of referent location and response direction provided a significantly better fit for the response time data than the base model without this interaction, $\chi^2(1) = 24.80$, $p < .001$, effect of referent location: β = 12.11, *t* = 3.54, 95% CI [5.40, 18.82], effect of response direction: β = 1.89, *t* = 0.58, 95% CI [−4.45, 8.22], interaction of referent location and response direction: β = −22.73, *t* = −4.98, 95% CI [−31.67, −13.79]. Response times were significantly faster in congruent (*M* = 570 ms) than in incongruent (*M* = 582 ms) trials. For movement times, the model with an additional fixed effect for the interaction of referent location and response direction again outperformed the base model without this interaction, $\chi^2(1) = 8.62$, $p = .003$, effect of referent location: β = 2.45, *t* = 1.76, 95% CI [−0.27, 5.17], effect of response direction: β = 2.97, *t* = 2.27, 95% CI [0.40, 5.54], interaction of referent location and response direction: β = −5.44, *t* = −2.94, 95% CI [−9.07, −1.81]. Movement times were significantly faster when the referent location matched (*M* = 131 ms) compared to mismatched (*M* = 134 ms) the response direction. Participants produced less errors in congruent (*M* = 1.42%) than in incongruent (*M* = 1.93%) trials, ruling out a speed-accuracy tradeoff.[1]

Thus, we replicated the spatial congruency effect obtained when using mouse movement responses on the horizontal plane inducing vertical movements of linguistic stimuli, even though we removed the trial-based correctness feedback appearing at the lower or upper end location of the linguistic stimuli. Nevertheless, the mouse movements were still associated with a direct and real-time visual feedback in terms of vertical word movements on the screen, possibly resulting in a spatial re-interpretation of response movements. To investigate in a more conclusive manner whether mere visual action effects are sufficient to generate spatial congruency effects, we ran a series of experiments replacing mouse movements on the horizontal plane

---

[1] When computing mean error rates, we did not consider trials with more than a single mouse click or an extreme response time shorter than 100 or longer than 3000 ms.

by simple stationary keypresses associated with visual action effects located at the lower or upper screen location.

Experiment 2

In this experiment, the “A” and “L” keys of a standard US keyboard served as stationary response keys. Importantly, pressing the “A” key (“L” key) consistently produced an immediate correctness feedback appearing at the lower (upper) screen location.

Method

Participants

We again collected data from 60 participants. We recruited participants through Amazon Mechanical Turk and carried out exclusions as in Experiment 1, resulting in a final sample of 50 participants. The ages of participants (16 females, 34 males; 49 right-handed, one ambidextrous) ranged between 21 and 61 years (*M* = 34.84 years, *SD* = 9.31 years). All participants were native English speakers, gave informed consent, and received $6.50 in return for participation. It took about 50 minutes to complete the experiment.

Apparatus and Stimuli

We programmed the experiment using jsPsych (Version 6.0.5; de Leeuw 2015) and asked participants to run the experiment either on a desktop computer or on a laptop. The “A” and “L” keys of a standard US keyboard were introduced as stationary response keys.

We employed a black-colored plus sign as fixation cross. Filled blue (RGB: 0, 0, 255), orange (RGB: 255, 128, 0), lilac (RGB: 150, 0, 255), and brown (RGB: 140, 80, 20) squares, circles, triangles, and diamonds (scaled to 110 × 110 pixels) were presented as training stimuli. In the experimental session, the English version of the list of implicit location words used in the previous experiment served as stimulus set (e.g., “abyss” vs. “balloon”). Like the German word list, the English version was shown to reliably produce spatial congruency effects in the context of the paradigm involving mouse movements on the horizontal plane inducing vertical stimulus movements (Schütt et al. 2022). Implicit location words were displayed in the same colors as training stimuli. In order to encourage participants to place the index fingers on the keyboard as instructed, we inserted catch trials showing the sentence "Press the key your left (right) index finger is lying on". In each trial, we provided feedback on response correctness ("Correct!" vs. "Wrong!"). Catch trial sentences and feedback appeared in black letters (RGB: 0, 0, 0). All stimuli were presented on a white background.

Procedure

The experiment had a training session and an experimental session. Importantly, during the training session, participants learnt to associate keyboard presses with visual action effects appearing at a lower or an upper location, respectively. Each shape (square, circle, triangle, and diamond) was presented four times in each of the four colors (blue, orange, lilac, and brown). In addition, we showed eight catch trials (four trials referring to the left index finger and four trials referring to the right index finger). This resulted in a total number of 72 training trials. The trial order was randomized. After finishing the training session, participants passed the experimental session, which was divided into four blocks. We employed each down word and each up word once per block, with each word appearing in a different font color in each block. Likewise, font colors were balanced within blocks. We inserted six catch trials per block (three trials referring to the left index finger and three trials referring to the right index finger). Thus, each experimental block comprised 70 trials. The order of trials and blocks was randomized.

Figure 1 illustrates the trial procedure. Each trial started with the presentation of the fixation cross (800 ms). Afterwards, a shape (training session), a word (experimental session), or a catch trial sentence was displayed centered on the screen until the participants pressed one of the response keys. The participants were asked to respond as fast and as accurately as possible to the shape color, the font color, or the sentence, respectively. Participants placed their left index finger on the "A" key and their right index finger on the "L" key. We mapped two colors to one response key and balanced both the arrangement of color pairs and the assignment of color pairs to response keys between participants, resulting in a total of six experimental versions. After responding to the stimulus, the participants got feedback on the accuracy of their reaction (2500 ms). The feedback location depended on the key pressed: It appeared centered at the bottom of the screen after pressing the “A” key but centered at the top of the screen after pressing the “L” key. Hence, participants learnt to link the “A” key (“L” key) to visual action effects located at a lower (an upper) location. Finally, the intertrial interval (1500 ms) followed, before the next trial started.

Design and Data Analysis

The experiment implemented a 2 (referent location: up vs. down) × 2 (feedback location: up vs. down) within-subjects design. The time period from the stimulus onset until participants pressed one of the response keys served as the dependent variable. We prepared and analyzed the data in accordance with the procedure performed to analyze the response times in Experiment 1. Removing outliers reduced the data set in total by less than 7%. The base model constructed for the linear mixed effects analysis included fixed effects for age, handedness, referent location and feedback location. When following the suggestion of Matuschek et al. (2017) to define a suitable random effects structure, we again had to omit some more complex models due to convergence issues or a singular fit warning. This resulted in incorporating random intercepts for participants and items.

Results and Discussion

Table 1 shows the mean response times as a function of referent location and feedback location. The likelihood ratio test revealed that the model with an additional fixed effect for the interaction of referent location and feedback location did not fit the data significantly better than the base model without this interaction, $\chi^2(1) = 0.24$, $p = .626$, effect of referent location: $\beta = 8.89$, $t = 1.37$, 95% CI [−3.87, 21.64], effect of feedback location: $\beta = 3.53$, $t = 0.65$, 95%

CI [−7.18, 14.23], interaction of referent location and feedback location: $\beta = 3.77$, $t = 0.49$, 95% CI [−11.37, 18.91]. Thus, response times did not differ significantly when comparing congruent (M = 729 ms) and incongruent (M = 728 ms) trials.

Taken together, associating simple stationary keypresses with immediate visual feedback shown at a lower or an upper screen location did not induce spatial congruency effects usually observed with implicit location words. This pattern of results provides first evidence that mere anticipating and integrating visual action effects appears not to be sufficient for the emergence of word-based language-space associations on the vertical axis. In a next step, we aimed at making the spatial dimension more salient, as this may help visual action effects to evolve their potential impact in the context of vertical language-space associations. Therefore, we decided to introduce the "B" and "Y" keys of the standard US keyboard as stationary response keys for the following experiments. These keys are located on the front-back axis of the keyboard, which should have a stronger association with the vertical axis than the left-right axis (i.e., the "A" and "L" keys of the standard US keyboard as used in Experiment 2). Analogous to Experiment 2, visual feedback appeared centered at the lower end of the screen after pressing the "B" key but centered at the upper end of the screen after pressing the "Y" key.

Experiment 3a

Method

Participants

We again recruited a total of 60 participants via Amazon Mechanical Turk and carried out exclusions as in the previous experiments. The final sample included 50 participants (15 females, 35 males; 43 right-handed, six left-handed, one ambidextrous). The participants were between 22 and 70 years old (M = 35.74 years, SD = 12.33 years). All of them were native English speakers, gave informed consent and received $6.50 in return for participation. It took about 50 minutes to complete the experiment.

Apparatus and Stimuli

Apparatus and stimuli stayed mostly the same as in Experiment 2. This time, however, the “B” and “Y” keys of the standard US keyboard served as stationary response keys, resulting in a response axis with one response key located lower on the keyboard and one response key located upper on the keyboard.

Procedure

The procedure was identical to Experiment 2, except for the following aspects. Half of the participants placed the left index finger on the “B” key and the right index finger on the “Y” key. For the other half of the participants the index finger placement was reversed.[2] The feedback was shown centered at the bottom of the screen after pressing the “B” key (response key located lower on the keyboard) but centered at the top of the screen after pressing the “Y” key (response key located upper on the keyboard). Hence, participants learnt to link the “B” key (“L” key) to visual action effects located at a lower (an upper) location.

Design and Data Analysis

Design, data preparation, and data analysis were identical to Experiment 2. Excluding outliers reduced the data set by less than 6%. The determination of the random effects structure for the linear mixed models resulted in including random intercepts for participants and items.

Results and Discussion

Table 1 gives the mean response times as a function of referent location and feedback location. The likelihood ratio test demonstrated that the model with a further fixed effect for the interaction of referent location and feedback location fitted the data significantly better than the base model without this interaction, $\chi^2(1) = 9.65, p = .002$, effect of referent location: $\beta = 22.83$, $t = 4.13$, 95% CI [11.99, 33.68], effect of feedback location: $\beta = 22.46$, $t = 4.28$, 95% CI [12.17,

---

[2] In Experiment 2, we refrained from balancing the index finger placement in order to avoid that participants have to cross their hands.

32.74], interaction of referent location and feedback location: $\beta = -23.03$, $t = -3.11$, 95% CI [−37.56, −8.50]. Participants responded significantly faster in congruent ($M = 751$ ms) than in incongruent ($M = 764$ ms) trials. Error rates were virtually identical in congruent ($M = 2.44\%$) and incongruent trials ($M = 2.45\%$), ruling out a speed-accuracy tradeoff.[3]

Consequently, visual action effects associated with lower or upper vertical space might be able to play a role in the emergence of spatial congruency effects typically obtained with implicit location words. Importantly – if these findings reflect a stable effect – contextual task conditions in terms of the saliency of the vertical spatial dimension appear to be a crucial factor. This sort of context-dependency is in line with prior research emphasizing that embodied spatial congruency effects tend not to occur automatically across all kinds of contexts; it is argued that even central grounding features, such as space, have to be made salient (see, for instance, Lebois et al. 2015). Following, we perform a direct replication of the current experiment to test the reliability of the finding. This way of proceeding is strongly suggested by the replication crisis (for overviews, see Shrout and Rodgers 2018; Wiggins and Christopherson 2019).

Experiment 3b

Method

Participants

Once again, we recruited 60 participants through Amazon Mechanical Turk. We excluded participants using the same criteria as in the previous experiments. The final sample included 40 participants (18 females, 22 males; 38 right-handed, two left-handed), who were between 20 and 64 years old ($M = 39.50$ years, $SD = 12.23$ years). All participants indicated to be native English speakers, gave informed consent, and received \$6.50 as reimbursement. It took about 50 minutes to complete the experiment.

---

[3] When computing mean error rates, we omitted trials with an extreme response time shorter than 100 or longer than 3000 ms.

Apparatus and Stimuli

Apparatus and stimuli were identical with Experiment 3a.

Procedure

The procedure was the same as in Experiment 3a.

Design and Data Analysis

Design, data preparation, and data analysis were identical with Experiment 3a. Outlier removal reduced the data set by about 8%. The determination of the random effects structure for the linear mixed models resulted in including random intercepts for participants and items.

Results and Discussion

Table 1 gives the mean response times as a function of referent location and feedback location. The likelihood ratio test revealed that the model with an additional fixed effect for the interaction of referent location and feedback location did not fit the data significantly better than the base model without this interaction, $\chi^2(1) = 0.07$, $p = .785$, effect of referent location: $\beta = -0.03$, $t = -0.01$, 95% CI [−16.72, 16.66], effect of feedback location: $\beta = -3.70$, $t = -0.51$, 95% CI [−17.83, 10.43], interaction of referent location and feedback location: $\beta = -2.78$, $t = -0.27$, 95% CI [−22.74, 17.18]. Response times were not significantly faster in congruent ($M = 848$ ms) than in incongruent ($M = 853$ ms) trials.

Accordingly, the findings of Experiment 3a could not be replicated. Visual action effects related to lower or upper vertical space did not provoke the emergence of spatial congruency effects, even though vertical space was made more salient in comparison to Experiment 2. These results clearly show the worth of direct replications in original research, as only such attempts can tell us whether new research findings are reliable or not. Finally, we also wondered whether realizing a stable influence of action effects on the occurrence of spatial congruency effects may require contextual task conditions emphasizing the vertical spatial dimension in an even more distinct manner. In Experiment 4, we thus refrained from displaying a correctness feedback at a lower or upper vertical location. Instead, after pressing the response key located in the lower (upper) part of the keyboard, the linguistic stimulus itself moved stepwise to the lower (upper) end of the screen. Crucially, this procedure is closely related to the paradigm that relies on mouse movements causing visual action effects (i.e., stimulus movements) towards the lower or upper screen location (see Schütt et al. 2022) – but avoids the real-time coupling of the response action and visual feedback provided by vertically moving stimuli on the screen. Furthermore, no actual arm movements are involved in the response action. Hence, possible spatial congruency effects should definitely be attributed to the mere anticipation of visual action effects associated with lower and upper space.

Experiment 4

Method

Participants

In accordance with the previous experiments, we recruited 60 participants via Amazon Mechanical Turk. Again, we excluded participants with less than 90% of correct responses in at least one experimental condition, which resulted in a final sample including 39 participants. The ages of the participants (14 females, 25 males; 39 right-handed) ranged between 25 and 69 years (*M* = 37.36 years, *SD* = 11.12 years). All participants confirmed to be native English speakers, provided informed consent, and received $6.50 as reimbursement. It took about 50 minutes to complete the experiment.

Apparatus and Stimuli

Apparatus and stimuli were mostly identical with the Experiments 3a and 3b. The only difference was that we did not use the expressions “Correct!” and “Wrong!” anymore as we did not give trial-based correctness feedback in the current experiment.

Procedure

The procedure was very similar to that of the Experiments 3a and 3b. We solely replaced the feedback indicating the trial-associated response correctness. Immediately after responding to the stimulus, the stimulus moved in a stepwise manner to the bottom of the screen (when the "B" key was pressed) or to the top of the screen (when the "Y" key was pressed), respectively. This movement consisted of five single steps. Subsequent to each step, the stimulus stayed for 500 ms at its new location.

Design and Data Analysis

The experiment had a 2 (referent location: up vs. down) × 2 (stimulus movement: up vs. down) within-subjects design. The period from the appearance of the word until pressing one of the response keys served as the dependent variable. Data preparation and data analysis followed the same procedure as in the Experiments 3a and 3b. Outlier elimination reduced the data set by less than 10%. As there were only right-handed participants, the linear mixed effects models did not include a fixed effect for handedness. Moreover, we had to omit the fixed effect for age due to convergence issues. When determining an appropriate random effects structure, some more complex models had to be skipped, because we encountered singular fit warnings. Finally, we incorporated random intercepts for participants and items.

Results and Discussion

Table 1 illustrates the mean response times as a function of referent location and stimulus movement. The likelihood ratio test demonstrated that the model with a further fixed effect for the interaction of referent location and stimulus movement did not provide a significantly better data fit than the model without this interaction, $\chi^2(1) = 0.11$, $p = .744$, effect of referent location: $\beta = -3.25$, $t = -0.37$, 95% CI [−20.36, 13.86], effect of stimulus movement: $\beta = -7.21$, $t = -0.95$, 95% CI [−22.08, 7.66], interaction of referent location and stimulus movement: $\beta = 3.49$, $t = 0.33$, 95% CI [−17.49, 24.48]. Response times in congruent ($M = 926$ ms) and incongruent ($M = 925$ ms) trials did not differ significantly. Thus, emphasizing the vertical spatial dimension using a probably even more evident approach did not help to promote spatial congruency effects. Altogether, the mere anticipation of visual action effects associated with a lower or upper vertical location does not appear to be sufficient to evoke spatial congruency effects encountered in the context of implicit location words.

General Discussion

The investigation of vertical language-space associations is a prominent line of research in the field of embodied language comprehension. Especially, researchers have been working on revealing whether and under which conditions word processing involves spatial simulations. For this purpose, plenty of studies have looked at implicit location words, which are characterized by referring to entities related to a lower or upper vertical location (e.g., “worm” vs. “bird”). Other studies have aimed to examine whether words denoting abstract concepts (e.g., “justice”, “love”, or “freedom”) are comprehended in terms of a metaphorical mapping onto vertical space. The link between processing these groups of words and mentally simulating vertical space has been investigated by presenting tasks from the visual or motor domain. Inspired by ideomotor theory (for reviews, see Hommel et al. 2001; Shin et al. 2010; Badets et al. 2016), we asked whether the mere anticipation of visual action effects displayed in lower or upper vertical space might be sufficient to evoke spatial congruency effects that are typically encountered in the context of vertical language-space associations.

In the first experiment, we successfully replicated the finding that mouse movements on the horizontal plane that induce vertical movements of word-based stimuli (i.e., implicit location words) reliably produce spatial congruency effects (see also Schütt et al. 2022). The emergence of this effect could be explained by the correspondence of vertical spatial information associated with the word stimuli and the visual action effects towards a lower or an upper screen location generated by the mouse movements. However – due to the fact that the mouse movements and the vertical stimulus movements were coupled in a direct and real-time manner – it is also still possible that participants were prompted to code and process the mouse movements in terms of vertical response movements. We therefore replaced the mouse movements by simple stationary keypress responses, which produced visual action effects at a lower or an upper screen location. In a series of experiments, we did not reliably observe spatial congruency effects. Particularly, this was also true when vertical space was made more salient by contextual task conditions.

These results suggest that anticipating action effects displayed at a lower or an upper vertical location do not suffice to obtain spatial congruency effects typically encountered in the context of implicit location words. The mere anticipation of response consequences in vertical space thus appears not to be central to the occurrence of the effects under investigation. These might rather depend on a prominent activation of the sensorimotor system, endorsing the idea that word processing is closely linked to re-activating sensorimotor experiences of the word's referent. Apparently, an adequate involvement of the sensorimotor system can be realized by different means. For instance, research using the Stroop-like motor response task (e.g., Brookshire et al. 2010; Lachmair et al. 2011; Dudschig et al. 2012b, 2014, 2015; Thornton et al. 2013; Öttl et al. 2017; Ahlberg et al. 2018; Vogt et al. 2019) requires participants to perform actual vertical response movements. Typically, response times are faster when the response direction and the vertical location usually associated with the word content are congruent. Interestingly, a comparable pattern of results was obtained for participants pressing a lower or upper response key on a vertically mounted response device without conducting vertical response movements (Lachmair et al. 2011, Experiment 4). Thus, in this specific case, the spatial congruency effect might evolve from responses being related to pronounced proprioceptive perceptions connected to the lower or upper vertical space (i.e., operating response keys that were arranged vertically, but clearly separated with a distance of about 25 cm). Finally, in many studies with tasks from the visual domain, critical word stimuli such as implicit location words appear at a lower or an upper vertical screen position. Hence, performing the experimental task involves perceiving the word stimuli at a congruent or an incongruent vertical location. For instance, Zwaan and Yaxley (2003) demonstrated that decisions on the semantic relatedness of word pairs (e.g., “attic” and “basement”) are faster when the arrangement of the words corresponds to the usual location of the word referents (e.g., “attic” above “basement”). In the study by Šetić and Domijan (2007), participants categorized words presented at a lower or an upper screen location (Experiment 1: “flying animal” vs. “non-flying animal”; Experiment 2: “living entity” vs. non-living entity”). Response times were faster when the typical vertical location of the word referent matched the presentation location. Another line of research showed that processing implicit location words guides visual attention on the vertical plane and therefore affects the identification of visual targets displayed at lower or upper locations (e.g., Estes et al. 2008; Verges and Duffy 2009; Dudschig et al. 2012a; Gozli et al. 2013). Noteworthy, most studies employing tasks from the visual domain also include motor aspects, because they usually require participants to execute saccadic eye movements in the vertical space. Furthermore, it has to be mentioned that visual tasks tend to evoke diverse congruency effects, varying in terms of being related to facilitation or interference (for a more thorough discussion of this aspect, see Pecher et al. 2010).

Our set of experimental stimuli solely comprised implicit location words. Consequently, all presented words referred to entities that can be experienced in a direct way by means of the sensorimotor system. In contrast, this does not hold true for words describing abstract concepts (e.g., “power”, “morality”, or “valence”), which do not possess a respective referent in the real word and are often thought to be mentally represented by a metaphorical mapping onto vertical space (e.g., Lakoff and Johnson 1980, 1999). Therefore, understanding abstract words might rely to a much lesser extent on the sensorimotor system and the re-activation of corresponding brain activation patterns. As a result, spatial congruency effects arising from abstract words could be less closely related to the sensorimotor system and in turn be more susceptible to other factors, such as the anticipation of visual action effects in vertical space. It has already been shown that the relationship of the abstract concept of time and horizontal space is prone to be modulated by visual action effects. Participants from cultures with a left-to-right reading and writing direction respond to past-related verbal stimuli faster when pressing a left (vs. right) response key, but to future-related verbal stimuli faster when pressing a right (vs. left) response key (so-called space-time congruency effect; for a recent overview, see von Sobbe et al. 2019). This suggests that the past (the future) is mapped onto the left (right) side of a horizontal mental timeline. Importantly, however, Janczyk and Ulrich (2019) demonstrated that the pattern of the space-time congruency effect is reversed when a left (right) keypress elicits a visual action effect on the right (left) side. The authors therefore concluded that the space-time congruency effect appears not to be deeply rooted in sensorimotor brain areas, adding that this might be different for effects related to other abstract concepts. Accordingly, the influence of visual action effects on horizontal and vertical language-space associations could vary depending on the extent to which abstract concepts are linked to concrete bodily experiences.

Conclusion

Prior research has revealed that processing words describing entities that are associated with a lower or an upper vertical location (e.g., worm vs. eagle) activates corresponding spatial features. This is typically reflected in spatial congruency effects. In the current work, we aimed to investigate whether such effects can result from the mere anticipation of visual action effects related to lower or upper vertical space. In a series of several experiments, participants provided keypress responses producing immediate visual action effects at or toward the lower or the upper end of the screen. We did not observe spatial congruency effects in a reliable manner, indicating that anticipating visual action effects in vertical space is not sufficient for evoking such effects. These might rather depend on a close relationship to and therefore a distinct involvement of the sensorimotor system – for instance, in terms of vertical response movements, proprioceptive experiences in vertical space, or perceiving critical word stimuli including spatial information at vertical locations. Since we exclusively looked at implicit location words, further studies will be needed to evaluate whether our findings can be extended to other groups of words, such as those referring to abstract concepts that are thought to be metaphorically mapped onto vertical space.

References

Ahlberg DK, Bischoff H, Kaup B, Bryant D, Strozyk JV (2018) Grounded cognition: Comparing language × space interactions in first language and second language. Appl Psycholinguist 39(2):437–459. https://doi.org/10.1017/S014271641700042X

Anderson JR (1983) The architecture of cognition. Harvard University Press, Cambridge, MA

Ansorge U, Khalid S, König P (2013) Space-valence priming with subliminal and supraliminal words. Front Psychol 4:81. https://doi.org/10.3389/fpsyg.2013.00081

Baayen RH, Davidson DJ, Bates DM (2008) Mixed-effects modeling with crossed random effects for subjects and items. J Mem Lang 59(4):390–412. https://doi.org/1016/j.jml.2007.12.005

Badets A, Koch I, Philipp AM (2016) A review of ideomotor approaches to perception, cognition, action, and language: Advancing a cultural recycling hypothesis. Psychol Res 80(1):1–15. https://doi.org/10.1007/s00426-014-0643-8

Barsalou LW (1999) Perceptual symbol systems. Behav Brain Sci 22(4):577–660. https://doi.org/10.1017/S0140525X99002149

Bates D, Mächler M, Bolker BM, Walker SC (2015) Fitting linear mixed-effects models using lme4. J Stat Softw 67(1):1–48. https://doi.org/10.18637/jss.v067.i01

Borghi AM, Glenberg AM, Kaschak MP (2004) Putting words in perspective. Mem Cogn 32(6):863–873. https://doi.org/10.3758/BF03196865

Brookshire G, Ivry R, Casasanto D (2010) Modulation of motor-meaning congruity effects for valenced words. Proc Annu Meet Cogn Sci Soc 32:1940–1945. https://escholarship.org/uc/item/2ch97289

Bub DN, Masson MEJ, Cree GS (2008) Evocation of functional and volumetric gestural knowledge by objects and words. Cognition 106(1):27–58. https://doi.org/10.1016/j.cognition.2006.12.010

Chasteen AL, Burdzy DC, Pratt J (2010) Thinking of God moves attention. Neuropsychologia 48(2):627–630. https://doi.org/10.1016/j.neuropsychologia.2009.09.029

Chatterjee A (2010) Disembodying cognition. Lang Cogn 2(1):79–116. https://doi.org/10.1515/langcog.2010.004

de Leeuw JR (2015) jsPsych: A JavaScript library for creating behavioral experiments in a web browser. Behav Res Methods 47(1):1–12. https://doi.org/10.3758/s13428-014-0458-y

Dudschig C, de la Vega I, Kaup B (2014) Embodiment and second-language: Automatic activation of motor responses during processing spatially associated L2 words and emotion L2 words in a vertical Stroop paradigm. Brain Lang 132:14–21. https://doi.org/10.1016/j.bandl.2014.02.002

Dudschig C, de la Vega I, Kaup B (2015) What's up? Emotion-specific activation of vertical space during language processing. Acta Psychol (Amst) 156:143–155. https://doi.org/10.1016/j.actpsy.2014.09.015

Dudschig C, Lachmair M, de la Vega I, De Filippis M, Kaup B (2012a) From top to bottom: Spatial shifts of attention caused by linguistic stimuli. Cogn Process 13(S1):151–154. https://doi.org/10.1007/s10339-012-0480-x

Dudschig C, Lachmair M, de la Vega I, De Filippis M, Kaup B (2012b) Do task-irrelevant direction-associated motion verbs affect action planning? Evidence from a Stroop paradigm. Mem Cogn 40(7):1081–1094. https://doi.org/10.3758/s13421-012-0201-9

Estes Z, Verges M, Barsalou LW (2008) Head up, foot down: Object words orient attention to the objects' typical location. Psychol Sci 19(2):93–97. https://doi.org/10.1111/j.1467-9280.2008.02051.x

Fodor JA (1975) The language of thought. Harvard University Press, Cambridge, MA

Glenberg AM, Kaschak MP (2002) Grounding language in action. Psychon Bull Rev

9(3):558–565. https://doi.org/10.3758/BF03196313

Gozli DG, Chasteen AL, Pratt J (2013) The cost and benefit of implicit spatial cues for visual attention. J Exp Psychol Gen 142(4):1028–1046. https://doi.org/10.1037/a0030362

Hill PL, Lapsley DK (2009) The ups and downs of the moral personality: Why it's not so black and white. J Res Pers 43(3):520–523. https://doi.org/10.1016/j.jrp.2009.01.014

Hommel B, Müsseler J, Aschersleben G, Prinz W (2001) The theory of event coding (TEC): A framework for perception and action planning. Behav Brain Sci 24(5):849–878. https://doi.org/10.1017/S0140525X01000103

Janczyk M, Ulrich R (2019) Action consequences affect the space-time congruency effect on reaction time. Acta Psychol (Amst) 198:102850. https://doi.org/10.1016/j.actpsy.2019.05.002

Jiang M, Henley TB (2012) Power and spatial relations. J Cogn Psychol 24(7):829–835. https://doi.org/10.1080/20445911.2012.702749

Kan IP, Barsalou LW, Olseth Solomon K, Minor JK, Thompson-Schill SL (2003) Role of mental imagery in a property verification task: fMRI evidence for perceptual representations of conceptual knowledge. Cogn Neuropsychol 20(3–6):525–540. https://doi.org/10.1080/02643290244000257

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

Kuznetsova A, Brockhoff PB, Christensen RHB (2017) lmerTest package: Tests in linear mixed effects models. J Stat Softw 82(13):1–26. https://doi.org/10.18637/jss.v082.i13

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. Psychon Bull Rev 18(6):1180–1188. https://doi.org/10.3758/s13423-011-0158-x

Lakoff G, Johnson M (1980) Metaphors we live by. University of Chicago Press, Chicago, IL

Lakoff G, Johnson M (1999) Philosophy in the flesh: The embodied mind and its challenge to western thoughts. Basic Books, New York, NY

Lebois LAM, Wilson-Mendenhall CD, Barsalou LW (2015) Are automatic conceptual cores the gold standard of semantic processing? The context-dependence of spatial meaning in grounded congruency effects. Cogn Sci 39(8):1764–1801. https://doi.org/10.1111/cogs.12174

Matuschek H, Kliegl R, Vasishth S, Baayen H, Bates D (2017) Balancing Type I error and power in linear mixed models. J Mem Lang 94:305–315. https://doi.org/10.1016/j.jml.2017.01.001

Meier BP, Hauser DJ, Robinson MD, Friesen CK, Schjeldahl K (2007) What's "up" with God? Vertical space as a representation of the divine. J Pers Soc Psychol 93(5):699–710. https://doi.org/10.1037/0022-3514.93.5.699

Meier BP, Robinson MD (2004) Why the sunny side is up: Associations between affect and vertical position. Psychol Sci 15(4):243–247. https://doi.org/10.1111/j.0956-7976.2004.00659.x

Morey RD (2008) Confidence intervals from normalized data: A correction to Cousineau (2005). Tutor Quant Methods Psychol 4(2):61–64. https://doi.org/10.20982/tqmp.04.2.p061

Öttl B, Dudschig C, Kaup B (2017) Forming associations between language and sensorimotor traces during novel word learning. Lang Cogn 9(1):156–171. https://doi.org/10.1017/langcog.2016.5

Pecher D, Van Dantzig S, Boot I, Zanolie K, Huber DE (2010) Congruency between word position and meaning is caused by task-induced spatial attention. Front Psychol 1:30. https://doi.org/10.3389/fpsyg.2010.00030

Pecher D, Zeelenberg R, Barsalou LW (2003) Verifying different-modality properties for concepts produces switching costs. Psychol Sci 14(2):119–124. https://doi.org/10.1111/1467-9280.t01-1-01429

Pulvermüller F, Hauk O, Nikulin V V., Ilmoniemi RJ (2005) Functional links between motor and language systems. Eur J Neurosci 21(3):793–797. https://doi.org/10.1111/j.1460-9568.2005.03900.x

Pylyshyn ZW (1984) Computation and cognition: Towards a foundation for cognitive science. MIT Press, Cambridge, MA

Richardson DC, Spivey MJ, Barsalou LW, McRae K (2003) Spatial representations activated during real-time comprehension of verbs. Cogn Sci 27(5):767–780. https://doi.org/10.1016/S0364-0213(03)00064-8

Santiago J, Ouellet M, Román A, Valenzuela J (2012) Attentional factors in conceptual congruency. Cogn Sci 36(6):1051–1077. https://doi.org/10.1111/j.1551-6709.2012.01240.x

Schubert TW (2005) Your highness: Vertical positions as perceptual symbols of power. J Pers Soc Psychol 89(1):1–21. https://doi.org/10.1037/0022-3514.89.1.1

Schütt E, Mackenzie IG, Kaup B, Dudschig C (2022) Replacing vertical actions by mouse movements: A web-suited paradigm for investigating vertical spatial associations. Psychol Res 87(1):194–209. https://doi.org/10.1007/s00426-022-01650-6

Šetić M, Domijan D (2007) The influence of vertical spatial orientation on property verification. Lang Cogn Process 22(2):297–312. https://doi.org/10.1080/01690960600732430

Shin YK, Proctor RW, Capaldi EJ (2010) A review of contemporary ideomotor theory. Psychol Bull 136(6):943–974. https://doi.org/10.1037/a0020541

Shrout PE, Rodgers JL (2018) Psychology, science, and knowledge construction: Broadening perspectives from the replication crisis. Annu Rev Psychol 69:487–510. https://doi.org/10.1146/annurev-psych-122216-011845

Simmons WK, Martin A, Barsalou LW (2005) Pictures of appetizing foods activate gustatory cortices for taste and reward. Cereb Cortex 15(10):1602–1608. https://doi.org/10.1093/cercor/bhi038

Simonsohn U (2015) Small telescopes: Detectability and the evaluation of replication results. Psychol Sci 26(5):559–569. https://doi.org/10.1177/0956797614567341

Thornton T, Loetscher T, Yates MJ, Nicholls MER (2013) The highs and lows of the interaction between word meaning and space. J Exp Psychol Hum Percept Perform 39(4):964–973. https://doi.org/10.1037/a0030467

Verges M, Duffy S (2009) Spatial representations elicit dual-coding effects in mental imagery. Cogn Sci 33(6):1157–1172. https://doi.org/10.1111/j.1551-6709.2009.01038.x

Vogt A, Kaup B, Dudschig C (2019) When words are upside down: Language-space associations in children and adults. J Exp Child Psychol 186:142–158. https://doi.org/10.1016/j.jecp.2019.06.001

von Sobbe L, Scheifele E, Maienborn C, Ulrich R (2019) The space-time congruency effect : A meta-analysis. Cogn Sci 43(1):e12709. https://doi.org/10.1111/cogs.12709

Wiggins BJ, Christopherson CD (2019) The replication crisis in psychology: An overview for theoretical and philosophical psychology. J Theor Philos Psychol 39(4):202–217. https://doi.org/10.1037/teo0000137

Wu X, Jia H, Wang E, Du C, Wu X, Dang C (2016) Vertical position of Chinese power words influences power judgments: Evidence from spatial compatibility task and event-related potentials. Int J Psychophysiol 102:55–61. https://doi.org/10.1016/j.ijpsycho.2016.03.005

Zanolie K, Van Dantzig S, Boot I, Wijnen J, Schubert TW, Giessner SR, Pecher D (2012)

Mighty metaphors: Behavioral and ERP evidence that power shifts attention on a vertical dimension. Brain Cogn 78(1):50–58. https://doi.org/10.1016/j.bandc.2011.10.006

Zhai D, Guo Y, Lu Z (2018) A dual mechanism of cognition and emotion in processing moral-vertical metaphors. Front Psychol 9:1554. https://doi.org/10.3389/fpsyg.2018.01554

Zwaan RA, Madden CJ (2005) Embodied sentence comprehension. In: Pecher D, Zwaan RA (eds) Grounding cognition: The role of perception and action in memory, language, and thinking. Cambridge University Press, New York, NY, pp 224–245

Zwaan RA, Stanfield RA, Yaxley RH (2002) Language comprehenders mentally represent the shapes of objects. Psychol Sci 13(2):168–171. https://doi.org/10.1111/1467-9280.00430

Zwaan RA, Yaxley RH (2003) Spatial iconicity affects semantic relatedness judgments. Psychon Bull Rev 10(4):954–958. https://doi.org/10.3758/BF03196557

Table 1

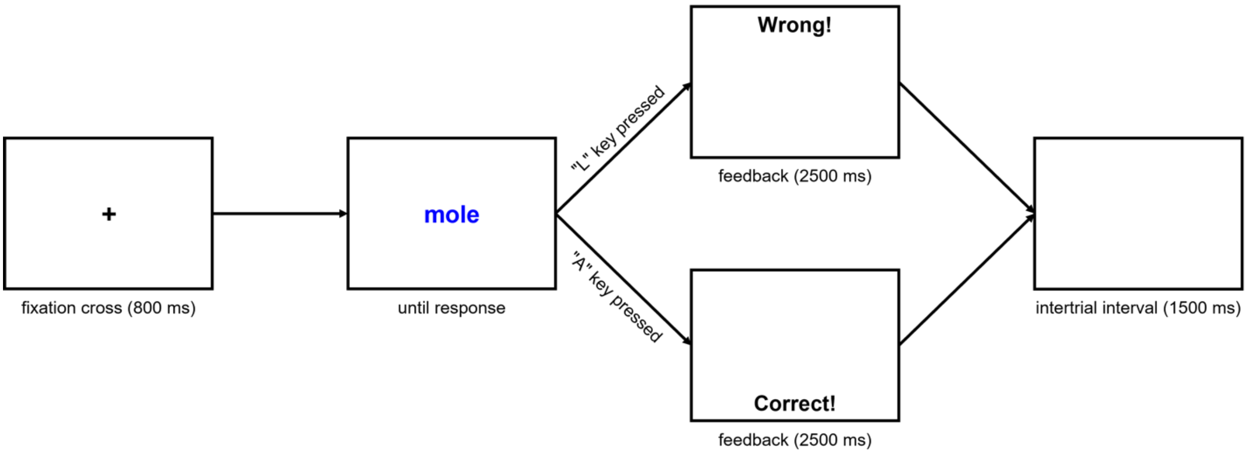
Mean Response Times in the Experiments 1 to 4

	Response/Feedback/Movement Up		Response/Feedback/Movement Down	
	Referent Up	Referent Down	Referent Up	Referent Down
Experiment 1	566 [561, 571]	577 [572, 582]	588 [583, 593]	575 [569, 581]
Experiment 2	737 [725, 749]	726 [715, 737]	731 [719, 743]	721 [710, 732]
Experiment 3a	763 [755, 771]	765 [754, 776]	764 [754, 774]	740 [730, 750]
Experiment 3b	847 [833, 861]	851 [841, 861]	854 [842, 866]	848 [836, 860]
Experiment 4	922 [904, 940]	921 [904, 938]	928 [912, 944]	930 [912, 948]

Note. Response times are given in milliseconds. The values in brackets represent 95% within-subjects confidence intervals calculated as recommended by Morey (2008).

Figure 1

Experimental Trial Procedure in Experiment 2



Note. The figure shows the procedure of an experimental trial requiring participants to press the "A" key for a correct response. Each trial began with the presentation of the fixation cross. Then, the stimulus appeared centered on the screen until participants responded to its font color. Right after giving the response, participants received an accuracy feedback. Importantly, the feedback was always displayed at the bottom of the screen after pressing the "A" key but at the top of the screen after pressing the "L" key. Consequently, the "A" key (the "L" key) was associated with visual action effects presented at a lower (an upper) vertical location.

Appendix C: Article 3

Copyright notice

To cite the article

Schütt, E., Dudschig, C., Bergen, B. K., & Kaup, B. (2023). Sentence-based mental simulations: Evidence from behavioral experiments using garden-path sentences. *Memory & Cognition*, *51*(4), 952-965. https://doi.org/10.3758/s13421-022-01367-2

Memory & Cognition (2023) 51:952–965
https://doi.org/10.3758/s13421-022-01367-2

Sentence-based mental simulations: Evidence from behavioral experiments using garden-path sentences

Emanuel Schütt[1] · **Carolin Dudschig**[1] · **Benjamin K. Bergen**[2] · **Barbara Kaup**[1]

Accepted: 11 October 2022 / Published online: 28 October 2022

Abstract

Language comprehenders activate mental representations of sensorimotor experiences related to the content of utterances they process. However, it is still unclear whether these sensorimotor simulations are driven by associations with words or by a more complex process of meaning composition into larger linguistic expressions, such as sentences. In two experiments, we investigated whether comprehenders indeed create sentence-based simulations. Materials were constructed such that simulation effects could only emerge from sentence meaning and not from word-based associations alone. We additionally asked when during sentence processing these simulations are constructed, using a garden-path paradigm. Participants read either a garden-path sentence (e.g., "As Mary ate the egg was in the fridge") or a corresponding unambiguous control with the same meaning and words (e.g., "The egg was in the fridge as Mary ate"). Participants then judged whether a depicted entity was mentioned in the sentence or not. In both experiments, picture response times were faster when the picture was compatible (vs. incompatible) with the sentence-based interpretation of the target entity (e.g., both for garden-path and control sentence: an unpeeled egg), suggesting that participants created simulations based on the sentence content and only operating over the sentence as a whole.

Keywords Sentence comprehension · Embodied cognition · Mental simulations · Garden-path sentences · Incrementality

Throughout the past two decades, the embodied cognition view has had an increasing influence on research concerned with human cognition (e.g., Chatterjee, 2010). This is especially true for the area of language comprehension. Embodied cognition views of human language comprehension (e.g., Barsalou, 1999; Bergen, 2012; Glenberg & Kaschak, 2002; Zwaan & Madden, 2005) propose that comprehenders grasp the meaning of a word by mentally simulating the word's referent. Concretely, upon hearing or reading a word, comprehension is effected by reactivating sensorimotor experiences that are associated with its referent. For example, when hearing the word "sky", comprehenders might reactivate the experience of perceiving the typical color of the sky (i.e., blue) or the experience of looking up. Importantly, words are hypothesized to activate such sensorimotor experiences as—particularly during childhood—they tend to regularly co-occur with their referents in everyday life (Vogt et al., 2019). For instance, a mother may point at a cat and say to her child: "Look! A cat".

Naturally, words are typically encountered together with other words forming phrases and sentences that convey meaning beyond the word level. Embodied cognition views of language comprehension suggest that comprehenders combine reactivated word-based sensorimotor experiences to create mental simulations corresponding to the meaning of the phrase or sentence in question. Thus, the literature proposes two types of simulation mechanism: Word-based simulations that are sensorimotor experiences triggered by individual words and sentence-based simulations that result from merging word-based simulations to obtain a combined meaning on phrasal or sentential level (see, for instance, Kaup et al., 2016).

Evidence for the word-based mechanism is extensive. A number of behavioral studies have demonstrated that comprehenders reactivate spatial experiences when they encounter words whose referents are typically associated with an

---

✉ Emanuel Schütt
emanuel.schuett@uni-tuebingen.de

[1] Department of Psychology, Language and Cognition Research Group, University of Tübingen, Schleichstr. 4, 72076 Tübingen, Germany

[2] Department of Cognitive Science, University of California at San Diego, La Jolla, CA, USA

upper or lower vertical location (*implicit location words*; e.g., "satellite" vs. "grave"). For example, in a study by Lachmair et al. (2011), implicit location words were shown centered on the screen and in different font colors. Participants were asked to respond to the font color of the words by performing upward and downward arm movements. Crucially, even though the task did not require lexical access, response times were faster when the movement direction matched the typical vertical location of the implicit location word. The identical pattern of results was obtained in further studies using highly similar materials and experimental procedures (e.g., Ahlberg et al., 2018; Dudschig et al., 2012, 2014, 2015; Öttl et al., 2017; Schütt et al., 2022; Thornton et al., 2013; Vogt et al., 2019). In another study by Dunn et al. (2014), participants made lexical decisions on auditorily presented implicit location words, non-spatial words, and non-words. Concretely, they provided their decision by fixating a target located above or below the screen center. As it turned out, initiating saccadic eye movements was faster when the vertical location typically associated with the implicit location word matched the saccade direction (e.g., when participants performed an upward saccade to state that "moon" is a word; see also Dudschig et al., 2013). Interestingly, Ansorge et al. (2010) used simple German words referring to an upper or lower spatial position on the vertical axis (translated into English: "on top"; "above"; "upward"; "high"; "downward"; "deep"; "down"; "below") as primes and targets in a masked priming paradigm. The task was to press a higher response key when the target referred to a higher spatial position and to press a lower response key when the target referred to a lower spatial position. The results revealed that response times were faster when prime and target words had the same spatial feature (e.g., this was true for the prime–target pair "above–high"). This clearly illustrates that even rather unconscious word processing can influence subsequent sensorimotor processing. Moreover, there is also evidence from neuroscience suggesting that word processing involves reactivating sensorimotor experiences. For instance, reading odor-related words (e.g., "cinnamon"; "garlic"; "jasmine") compared with reading odor-neutral words (e.g., "coat"; "poker"; "glasses") induced an elicited activation in the primary olfactory cortex (González et al., 2006). Similarly, reading action verbs associated with movements of the face, the arm, or the leg (e.g., to "lick"; "pick"; "kick") evoked a somatotopic activation in motor and premotor brain areas that are related to actual movements of the tongue, the fingers, or the feet (Hauk et al., 2004; but see Miller et al., 2018).

In contrast, the situation is much less clear for sentence-based simulations conveying meaning beyond the word level. Even though there has been research using sentence-based materials and producing results usually considered as simulation effects, it is uncertain whether these results reflect specifically sentence-based and not merely word-based simulation processes. A prototypical example for this issue is the seminal work of Zwaan et al. (2002), which asked whether language comprehenders mentally simulate the shapes of mentioned entities. Participants read sentences referring to entities in specific locations that modulated the implied shape of the entity in question. For instance, reading the sentence "The ranger saw the eagle in the sky" should trigger the simulation of an eagle with outstretched wings, but reading the sentence "The ranger saw the eagle in its nest" should be more likely to trigger the simulation of an eagle with folded wings. After reading sentences like these, participants saw an image of the entity and decided whether it had been mentioned in the sentence. The results revealed faster responses when the shape implied by the sentence matched the shape depicted in the image. Even though these effects were in response to sentences, it is also plausible that they were driven by individual words within those sentences, irrespective of sentential meaning composition. For example, associations with the words "eagle" and "sky" might have produced the mental simulation of an eagle with outstretched wings, whereas encountering the words "eagle" and "nest" might have elicited the mental simulation of an eagle with folded wings. In line with this interpretation, Kaup et al. (2007) found comparable simulation effects when sentences included a negation marker (e.g., "The eagle was *not* in the sky/nest"). Response times were faster when the picture matched the situation that was negated (e.g., an eagle with outstretched/folded wings) than when the picture matched the situation that was actually conveyed by the sentential meaning (e.g., an eagle with folded/outstretched wings). In addition, participants have been found to react faster to picture probes showing specific entity shapes indicated by sets of content words presented in word lists (Kaup et al., 2012). In sum, this line of work shows that simulation effects in response to sentences may or may not be attributable to simulation processes beyond the word level.

This same confound applies to other influential work in the area. The action-sentence compatibility effect (ACE; Glenberg & Kaschak, 2002) is the observation that participants are faster to judge the sensibility of sentences when the direction of the response movement matches (vs. mismatches) the direction of the action described in the sentences (e.g., when reacting with a movement towards the body to sentences such as "Courtney handed you the notebook" or "Andy delivered the pizza to you" compared with sentences such as "You handed Courtney the notebook" or "You delivered the pizza to Andy"). This ACE has recently been found to be hard to replicate (Morey et al., 2022; Winter et al., 2022). But even if the effect is real, it does not necessarily reflect simulation processes regarding the sentential meaning as a whole. Rather, participants might be reactivating sensorimotor experiences related to those words mentioned at the end of the sentence (e.g., “to you” and “handed you”: movement towards the body; “to Andy” and “handed Courtney”: movement away from the body), immediately before engaging their own motor response. The same is true of a similar paradigm, in which participants read sentences describing clockwise or counterclockwise manual rotations (e.g., “Jenny screwed in the light bulb” vs. “Liza opened the pickle jar”) while turning a knob device clockwise or counterclockwise (e.g., Capuano et al., 2022; Claus, 2015; Zwaan & Taylor, 2006). Once again, ACEs obtained in the context of this paradigm could equally be explained in terms of word-based effects.

Finally, this issue also appears to apply to experiments conducted in the context of a set of studies addressing the activation of specific hand-action representations during language comprehension (e.g., Bub & Masson, 2010; Masson et al., 2013; Masson, Bub, & Newton-Taylor, 2008; Masson, Bub, & Warren, 2008). In general, these studies distinguish functional hand actions related to interacting with an object according to its common function (e.g., pulling the trigger of a water pistol) and volumetric hand actions related to picking up or holding an object. For instance, in one experiment by Bub and Masson (2010) that might be affected by the confound discussed here, participants were presented with context sentences implying a functional or a volumetric hand action (e.g., “David wrote with the pencil” vs. “Bert picked up the pencil”). After a short or long delay (300 vs. 750 ms), which was accompanied by an image of the target referent (here: a pencil), a cue appeared prompting participants to perform an unrelated hand action or the functional or volumetric hand action typically associated with the object referenced in the sentence. Irrespective of the sentence context, there was a priming effect on response latencies for functional and volumetric actions after the short delay. In contrast, after the long delay, a priming effect was only present when the sentence context and the hand action matched. This might suggest that participants created a hand-action representation reflecting the sentential meaning over the course of time. However, it is again conceivable that these effects resulted from single words included in the sentences. For example, associations with the words “wrote” and “pencil” could have caused a functional hand-action representation, whereas reading the words “picked up” and “pencil” might have evoked a volumetric hand-action representation. Consequently—just as in the case of the examples outlined previously—the obtained results can be attributed to either sentence-based or word-based simulation effects.

The most compelling evidence for sentence-based simulation effects comes from studies using grammatical modifications to sentences for driving changes in simulation effects. For instance, Taylor and Zwaan (2008) observed that the rotational ACE persisted when a postverbal adverb referred to the matching action (e.g., “He found a new light bulb which he screwed in rapidly”) but ended when the postverbal adverb addressed the acting individual (e.g., “On the shelf, he found a closed jar which he opened hungrily”). Similarly, Bergen and Wheeler (2010) found that progressive sentences (e.g., “Beverley is closing/opening the drawer”) induce an ACE, whereas perfect sentences (e.g., “Beverley closed/opened the drawer”) do not. Another line of work (Bergen et al., 2007) showed that verbs of upwards or downwards motion provoke simulation effects when combined with concrete nouns (e.g., “The cork rocketed”), but not when combined with abstract nouns (e.g., “The numbers rocketed”). Moreover, a study by Bidet-Ildei, Gimenes, Toussaint, Almecija, et al. (2017) revealed that sentence plausibility can affect the judgment about biological motions. The visual detection capacity for human actions displayed under point-light conditions was better when an auditorily presented sentence including a congruent action verb was plausible compared with implausible (e.g., “The neighbor is running in the garden” vs. “The garden is running in the neighbor”), suggesting that simulations were influenced by contextual aspects beyond the word level (for related work, see Bidet-Ildei et al., 2020; Bidet-Ildei, Gimenes, Toussaint, Beauprez, et al., 2017). In general, however, findings suggesting sentence-based simulations are few and in some cases rely on effects that are hard to replicate.

Taken together, there is little doubt that comprehenders indeed generate word-based simulations, whereas clear evidence in favor of sentence-based simulations is still sparse. Therefore, the first aim of the present research was to provide a new method for investigating whether comprehenders engage in creating mental simulations beyond the word level when processing sentential materials. To this end, we adapted the sentence–picture verification framework (see Zwaan et al., 2002), building sentential materials containing words that independently should equally well activate both entity shapes that match the final sentence meaning and entity shapes that do not. For instance, a sentence like “The egg was in the fridge as Mary ate” includes the word “egg”, denoting an object that can take on different shapes, such as intact in its shell (i.e., unpeeled) versus cracked open and peeled. The sentence by design comprises a word associated with each of these shapes—“fridge” with the intact egg and “ate” with the cracked and peeled egg. A sentence–picture compatibility effect to a sentence like this would therefore be unlikely to derive from lexical associations alone.

As a second-order question, we also interrogated the time course of simulation processes during sentence comprehension. If simulations are constructed on the basis of language structures larger than the word alone, then does this occur incrementally over the course of processing an utterance, or does it wait until the end of a sentence, manifesting as a sort of sentential wrap-up effect? The incrementality of sentential mental simulations is an issue that has barely been tackled. Available evidence stems from research investigating the modulation of the rotational ACE during sentence comprehension. For instance, in the study by Zwaan and Taylor (2006), participants turned a knob device clockwise or counterclockwise to read sentences describing manual rotations frame by frame in a self-paced manner (e.g., "He / realized / that / the music / was / too loud / so he / turned down / the / volume"). Interestingly, the authors found that the rotational ACE occurred when encountering the critical verb region referring to the manual rotation movement (e.g., "turned down"). This suggests that participants immediately created motor simulations and did not wait until the end of the sentence. In another line of work, Sato et al. (2013) made use of the verb-final word order of the Japanese language. In one of their experiments, they investigated whether comprehenders build specific object shape simulations even before reaching the verb at the end of a sentence. Participants were presented with sentences generating the expectation of a certain object shape prior to encountering the verb. For instance, an item paraphrased as "Mother put the shirt neatly in the drawer" was arranged in the typical Japanese word order: "Mother-NOM shirt-ACC drawer-LOC neatly put". Crucially, reading the preverbal arguments could provide sufficiently constraining information for the comprehender to infer that the shape of the shirt was folded. To test for this early activation of scene-compatible object shape, participants responded to a picture probe *before* the verb appeared. The results showed faster responses when the shape implied by the preverbal phrase matched the depicted shape, indicating that detailed object shape simulations were created even though critical information was still missing. However—as the authors themselves noted—it is well possible that exposure to the picture probe itself prompted the participants to form detailed object shape simulations to perform the task, even if they would not have done so spontaneously during more naturalistic language processing. Thus, based on the few currently existing findings, it remains unknown whether language comprehenders routinely create incremental simulations during sentence processing. A second aim of our research therefore was to evaluate whether comprehenders construct mental simulations incrementally when reading sentences.

In order to investigate whether language comprehenders engage in forming sentence-based simulations and whether these are built in an incremental manner over the course of sentence processing, we presented participants with two kinds of sentences. The first were unambiguous sentences (e.g., "The egg was in the fridge as Mary ate") comprising words that were associated with multiple possible shapes as described above. The second were manipulated versions of those same sentences, so-called garden-path sentences, which used the same words but were transitionally ambiguous (e.g., "As Mary ate the egg was in the fridge"). Typically, comprehenders interpret the first verb of such garden-path sentences as transitive (in the example: "Mary ate the egg"). If comprehenders formulate incremental simulations, they should thus activate an initial shape interpretation (e.g., a ready-to-eat egg). However, when arriving at the second verb, where they have to reanalyze the sentence, they should then activate the final sentence-based shape interpretation (e.g., an unpeeled egg in its shell). Previous research on incrementality of semantic and syntactic processing has found that the initial syntactic or semantic interpretation created during garden-path processing tends to linger after the sentence has been reanalyzed (Christianson et al., 2001; Patson et al., 2009; Slattery et al., 2013). So, if participants construct incremental simulations, there should be evidence of both object shape interpretations being activated at the end of the garden-path sentences. By contrast, the unambiguous control sentences should only evoke a single entity interpretation reflecting the sentence-based object shape interpretation (e.g., an unpeeled egg in its shell).

In each trial, participants read either a garden-path or a control sentence, followed by a picture probe displaying the target entity (e.g., a ready-to-eat egg vs. an unpeeled egg in its shell). If language comprehenders create mental simulations on the basis of the sentence as a whole, we should see faster picture-verification times when the picture probe matched the sentence-based interpretation of the target entity. However, if simulation effects are driven by independent word associations, then there should be no such difference—sentences like "The egg was in the fridge as Mary ate" include the same number of words consistent with each of the two possible depicted shapes of an egg. Moreover, if language comprehenders create sentence-level simulations incrementally, then the sentence–picture compatibility effect should be larger for unambiguous sentences than for garden-path sentences; since they will have representations corresponding to both shapes active at the end of the sentence in the garden-path condition, there should be a smaller difference between response times to the pictures or none at all.

Experiment 1

Method

Participants

We aimed to collect data from $N = 96$ participants through Amazon Mechanical Turk. All participants reported being right-handed native English speakers. They also declared normal or corrected-to-normal vision. Their ages ranged from 23 to 60 years (M = 38.21 years, SD = 9.60 years). There were 41 female and 55 male participants. In total, 19 additional participants completed the experiment, but were excluded and replaced due to an error rate higher than 25% in at least one experimental condition or on the filler trials. All participants gave informed consent and received $4.00 in return for participation. It took about 20 to 30 minutes to conduct the experiment. The study was approved by the Ethics Committee for Psychological Research at the University of Tübingen (Identifier: 2018_0831_132).

Apparatus and stimuli

We employed the open-source JavaScript library jsPsych (Version 6.1.0; de Leeuw, 2015) to implement a browser-based experiment. Participants were explicitly asked to use a laptop or a desktop computer for participating. They pressed the space bar to start trials and indicate that they had read and understood a sentence. The "d" key and the "k" key served as response keys in the sentence–picture verification task.

For experimental trials, we created 36 pairs of critical sentences. Each pair included one garden-path sentence (e.g., "While Amber hunted the turkey was on the table"; "As Zoe bathed the baby slept in the bed"; "While Ryan won the car was in poor condition") and one matching unambiguous control sentence (e.g., "The turkey was on the table while Amber hunted"; "The baby slept in the bed as Zoe bathed"; "The car was in poor condition while Ryan won"). As described above, comprehenders tend to interpret the first verb of such garden-path sentences as transitive, constructing an initial interpretation of the target entity mentioned in the sentence (e.g., a living turkey; a baby in a bathtub; a car in brand-new condition; see, for instance, Christianson et al., 2001). Importantly, this initial interpretation corresponds to an intermediate processing step as comprehenders have to reanalyze the sentence when encountering the second verb, which should lead to creating a final sentence-based interpretation of the target entity (e.g., a ready-to-eat turkey; a dressed baby lying in the bed; a squalid car). Unambiguous control sentences, however, required comprehenders to form only a single interpretation of the target entity reflecting the sentence-based meaning (e.g., a ready-to-eat turkey; a dressed baby lying in the bed; a squalid car). As correctly answering experimental sentences always meant giving a "yes" response during the sentence–picture verification task, we also generated 36 filler sentences demanding a "no" response. Three fourths of the filler sentences followed the structure of unambiguous control sentences (e.g., "The scarf was in the washing machine as Bill knitted"); the remaining fourth of the filler sentences had the same structure as the garden-path sentences (e.g., "While Samuel ordered the fish swam upstream"). This reduced the proportion of sentences with garden-path structure participants encountered throughout the experiment. This in turn gave them fewer chances to learn the sentence structures and draw conclusions, reducing the likelihood that they would develop specific strategies with respect to garden-path processing (e.g., avoiding the initial entity interpretation). Four additional experimental sentences and four filler sentences were created for the practice session. Our sentential materials were partially adapted from or inspired by prior research (Christianson et al., 2001; Slattery et al., 2013; van Gompel et al., 2006).

For each pair of critical sentences, there were two pictures showing the respective target entity mentioned in the sentences. One of the pictures depicted the entity in the shape implied by the initial interpretation that could be inferred during garden-path processing. The other picture displayed the entity in the shape corresponding to the sentence-based interpretation, which was always the same for both garden-path and unambiguous control sentences. For instance, for the garden-path sentence "While Amber hunted the turkey was on the table" and the corresponding unambiguous control sentence "The turkey was on the table while Amber hunted", one picture showed a living turkey (initial entity interpretation during garden-path processing), whereas the other picture depicted a ready-to-eat turkey as served at Thanksgiving (sentence-based entity interpretation). A pretest ensured that the pictures referring to the target entity were comparable with respect to how clearly they depicted the entity irrespective of the shape (all $ps > .05$).[1] In filler trials, we presented participants with pictures showing an entity not mentioned in the respective filler sentence. For example, the filler sentence "The scarf was in the washing machine as Bill knitted" was followed by a picture of green olives. Pictures were in color and scaled to a size of 768 (width) × 576 (height) pixels. Some example materials are given in Table 1 as well as in Fig. 1. For copyright reasons, we are not able to make the pictures publicly available. However, all sentential and pictorial materials will be made accessible upon scientific request (please contact the corresponding author).

Procedure

We instructed our participants to participate in the experiment in an interference-free environment. Each trial started with the prompt "Please press the space bar to initiate the trial". After pressing the space bar, a fixation cross ("+"; 800 ms) appeared centered on the screen. Then the fixation cross was replaced by a critical or a filler sentence. Participants

---

[1] For the pretest, we recruited another 40 participants via Amazon Mechanical Turk. In each trial, we initially displayed the word denoting one of the target entities (e.g., "turkey"). Then, we presented the picture related to one of the two entity shapes (e.g., a living turkey or a ready-to-eat turkey). The task was to judge whether the picture showed the entity the word referred to.

Table 1 Examples of the sentential materials used in Experiments 1 and 2

Sentence type	Entity interpretation	
	Initial	Final
Garden-path		
1: As Mary ate the egg was in the fridge.	ready-to-eat egg	egg in its shell
2: While Edward painted the house was afire.	intact house	burning house
3: As Eve walked the dog lay on the ground.	walking dog	lying dog
4: While Miranda stirred the coffee was roasted.	cup of coffee	coffee beans
Control (Experiment 1)		
1: The egg was in the fridge as Mary ate.	not available	egg in its shell
2: The house was afire while Edward painted.	not available	burning house
3: The dog lay on the ground as Eve walked.	not available	lying dog
4: The coffee was roasted as Miranda stirred.	not available	coffee beans
Control (Experiment 2)		
1: As Mary ate the egg the butter was in the fridge.	not available	ready-to-eat egg
2: While Edward painted the house the forest was afire.	not available	intact house
3: As Eve walked the dog the cat lay on the ground.	not available	walking dog
4: While Miranda stirred the coffee the potato was roasted.	not available	cup of coffee

Comprehenders tend to interpret the first verb of garden-path sentences as transitive, resulting in an initial entity interpretation. However, when arriving at the second verb, the sentence must be reanalyzed, inducing the final sentence-based entity interpretation. By contrast, comprehenders should only create a single (sentence-based) entity interpretation in unambiguous control sentences

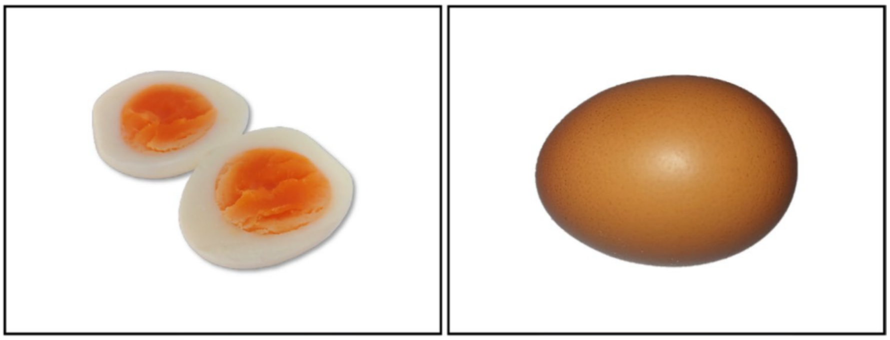


Fig. 1 Example of a picture pair used in Experiments 1 and 2. The picture pair refers to the sentence pair comprising the garden-path sentence "As Mary ate the egg was in the fridge" and the unambiguous control sentence "The egg was in the fridge as Mary ate" (Experiment 1) or "As Mary ate the egg the butter was in the fridge" (Experiment 2), respectively. The ready-to-eat egg (picture on the left) showed the initial entity interpretation during garden-path processing, whereas the unpeeled egg (picture on the right) should illustrate the sentence-based entity interpretation. For the unambiguous control sentence, we expected the participants to generate a single entity interpretation corresponding to the sentence-based entity interpretation (Experiment 1: an unpeeled egg; Experiment 2: a ready-to-eat egg)

were asked to read the sentence at a normal pace and to press the space bar. After this, a blank screen followed (500 ms), before participants were presented with the picture. Their task was to judge whether the depicted entity was mentioned in the previous sentence or not. Half of the participants pressed the "d" key for a "yes" response and the "k" key for a "no" response. For the other half of the participants, the response mapping was reversed. They were instructed to provide their response as fast as possible. During practice trials, participants received feedback regarding their response accuracy ("Correct!" vs. "Wrong!"; 1000 ms). The intertrial interval was 1500 ms. Initially, participants participated in a practice session. Subsequently, they performed three experimental blocks, each consisting of 12 critical and 12 filler trials. The conditions for each item (sentence type: garden-path vs. control sentence; picture type: compatible vs. incompatible with the sentence-based entity interpretation) were counterbalanced using four lists. Likewise, conditions were counterbalanced within the blocks. The order of trials was randomized. After each block, there was a self-paced break.

Design and data analysis

The experiment had a 2×2 within-subjects design, including the factors sentence type (garden-path vs. unambiguous control sentence) and picture type (compatible vs. incompatible with the sentence-based entity interpretation). Importantly, regarding the factor "picture type", the level "incompatible with the sentence-based entity interpretation" reflected the initial entity interpretation during garden-path processing (i.e., this particular entity interpretation should not be formed during control sentences). The time period from the occurrence of the picture on the screen until pressing the response key (picture response time) served as dependent variable. All data and R analysis scripts are publicly available online (https://doi.org/10.5281/zenodo.6504181).

We preprocessed and analyzed picture response times using the free statistical software R (Version 4.1.1). First, we removed filler trials and incorrectly answered critical trials. After this, extreme outliers were eliminated (picture response times shorter than 150 or longer than 3000 ms, respectively). Finally, to detect further outliers, we applied the two-step procedure proposed by Kaup et al. (2006). We transformed the picture response times of each participant to *z*-scores and discarded picture response times with a *z*-score that deviated more than two and a half standard deviations from the mean *z*-score of the respective item in the respective condition. In all, outlier exclusion reduced the data set by less than 4%. We made use of the R package lme4 (Version 1.1-27.1; Bates et al., 2015) to build a linear mixed model (see Baayen et al., 2008). Our model contained fixed effects for sentence type, picture type, and the interaction of both factors. In order to arrive at a suitable random effects structure, we referred to the data-driven model selection criterion introduced by Matuschek et al. (2017), which aims at balancing Type I error rate and power. When performing the procedure, however, we obtained warning messages indicating singular fits and convergence issues with respect to more complex models. Consequently, our model was finally restricted to include random intercepts for participants and items. For assessing the significance of the fixed effects, we employed the function *mixed* from the R package afex (Version 1.0-1; Singmann et al., 2021), which estimates mixed models based on lme4. We calculated *p* values through likelihood ratio tests (i.e., we chose the option "LRT" for the argument *method* within the function *mixed*). Generally, this means that the goodness of fit of a model with a specific fixed effect and the goodness of fit of a model without this specific effect were compared by referring to the ratio of their likelihoods. In case of the function *mixed*, the complete model with all fixed effects under consideration must be entered; the function then automatically builds suitable reduced models and performs likelihood ratio tests to compute *p* values for all fixed effects included in the complete model.

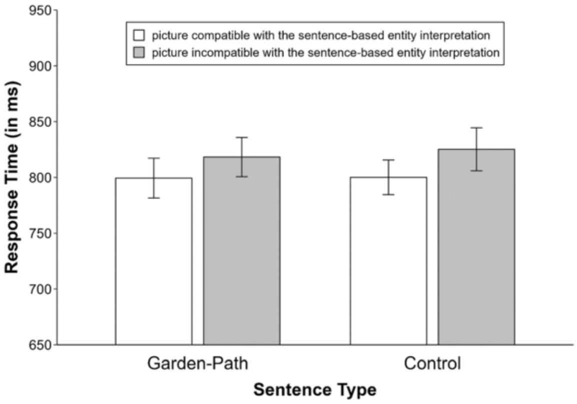


Fig. 2 Mean response times for the sentence–picture verification task in Experiment 1. Error bars denote 95% within-subjects confidence intervals calculated as recommended by Morey (2008)

We also evaluated reading times (i.e., the time period from the appearance of the sentence on the screen until pressing the space bar). Particularly, we were interested in whether reading times were modulated by the factor sentence type. For this purpose, we processed and analyzed reading times in the same way as picture response times, except for the following adaptations. First, we defined extreme outliers as reading times shorter than 500 or longer than 7000 ms, respectively. In total, removing outliers—including the two-step procedure suggested by Kaup et al. (2006)—reduced the data set by less than 7%. Second, the mixed model solely comprised a fixed effect for the factor sentence type. Again, we skipped some more complex models due to a singular fit when determining the random effects structure. The final model contained random intercepts for participants and items and by-item random slopes for sentence type.

Results and discussion

The data analysis revealed that there was a significant effect of sentence type on reading times, $\chi^2(1) = 17.78$, $p < .001$. As expected, participants needed more time to read garden-path sentences ($M = 1812$ ms) than unambiguous control sentences ($M = 1677$ ms). Figure 2 depicts the mean response times for the sentence–picture verification task as a function of sentence type and picture type. The effect of sentence type turned out not to be significant, $\chi^2(1) = 0.80$, $p = .372$. However, the results showed that there was a significant effect of picture type, $\chi^2(1) = 9.42$, $p = .002$, with participants responding faster when the picture probe matched (M = 800 ms) compared with mismatched (M = 822 ms) the sentence-based entity interpretation.[2] This effect was not significantly modulated by sentence type, $\chi^2(1) = 0.58$, $p = .446$.

First, these results can be interpreted as evidence with respect to the validity of the experimental procedure. Specifically, reading times were slower for garden-path sentences than for control sentences. This most likely reflects the additional processing difficulties associated with garden-path sentences (e.g., Ferreira & Henderson, 1991; Frazier & Rayner, 1982; Pickering & Traxler, 1998). This provides indirect evidence that participants read the sentences for comprehension and indeed created an intermediate, incremental interpretation of some kind when facing garden-path sentences. Participants also responded faster to picture probes compatible with the sentence-based entity interpretation. Since our materials were explicitly constructed to be less prone to word-based effects, this suggests that participants indeed generated sentence-based simulations. As argued, this comes against a backdrop of quite sparse evidence in favor of sentence-based simulations.

However, it should be noted that limitations in the construction of experimental materials did not allow us to manipulate which entity shape interpretation and thus which picture of a target entity was associated with the sentence-based interpretation in an individual item. For instance, for the garden-path sentence "As Mary ate the egg was in the fridge" and the control sentence "The egg was in the fridge as Mary ate", the sentence-based entity interpretation could not be varied and always corresponded to the unpeeled egg in its shell. Therefore—even though the pretest indicated that both pictures related to a target entity similarly clearly depicted this target entity—we cannot rule out the possibility that picture probes referring to the sentence-based shape interpretation were somehow preferred for the entities in question, and thus led to faster picture response times. Moreover, it remains possible that even if we included words in sentences aligned with each of the two entity shape interpretations, nevertheless these might have had unbalanced effects such that one shape was more consistent with the aggregate lexical associations of the sentence.

To address these limitations, we conducted a second experiment, adapting the materials in such a way that the sentence-based entity interpretation in control sentences was linked to the opposite shape and picture from the current experiment. Importantly, if the observed effect is due to comprehenders simulating the sentence-based meaning—and not an artefact of picture preference—the advantage for pictures compatible with the sentence-based entity interpretation should still occur in both garden-path sentences and control sentences. Since we did not observe evidence for the creation of incremental simulations during sentence comprehension in the present experiment (there was no significant interaction of picture type and sentence type), our material adaptions additionally aimed to create more fertile conditions for garden-path effects by making it more difficult to quickly identify this sentence type upon sentence presentation. Prior to starting data collection, we preregistered the experiment (https://aspredicted.org/tq6p9.pdf).

Experiment 2

Method

Participants

Based on the results of a pilot study, we conducted a simulation-based power analysis for the linear mixed model by using the R package mixedpower (Kumle et al., 2021). This revealed that we would need about 200 participants to reach a power of at least .80 for each of the fixed effects included in the model (sentence type; picture type; interaction of sentence type and picture type). We recruited participants via the crowdsourcing online labor marketplace Prolific. All participants (173 females, 27 males) reported themselves to be right-handed native English speakers. Their ages ranged between 18 and 55 years ($M = 23.51$ years, $SD = 5.85$ years). As in Experiment 1, we excluded and replaced participants with an error rate higher than 25% in at least one experimental condition or in filler trials. For this reason, there were seven additional participants who performed the experiment. All participants gave informed consent and received £4.00 in compensation for participation. It took about 20 to 30 minutes to finish the experiment. The study was approved by the Ethics Committee for Psychological Research at the University of Tübingen (Identifier: 2018_0831_132).

Apparatus and stimuli

Sentential materials were modified from Experiment 1. The garden-path sentence in each pair of critical sentences stayed exactly the same, but we replaced the unambiguous control sentences. Like the original control sentences, the newly introduced control sentences were transformed versions of the corresponding garden-path sentences. However, these were created by inserting an additional noun phrase

---

[2] As there was virtually no difference in error rates (picture compatible with sentence-based entity interpretation: 2.20%; picture incompatible with sentence-based entity interpretation: 2.26%), we can rule out an explanation in terms of a speed-accuracy tradeoff.

referring to a task-irrelevant entity prior to the second verb of the garden-path sentence. For instance, the garden-path sentence "As Mary ate the egg was in the fridge" was transformed into the unambiguous control sentence "As Mary ate the egg *the butter* was in the fridge". Thus, as in Experiment 1, understanding the unambiguous control sentences should involve forming a single entity interpretation fitting the sentence-based meaning (e.g., a ready-to-eat egg in the control sentence mentioned above). This time, however, the sentence-based entity interpretation always referred to the opposite shape and picture from Experiment 1. Furthermore, in Experiment 2, garden-path sentences and the newly introduced control sentences had very similar structure, which we hoped would make identifying the sentence type at hand and applying specific strategies (e.g., avoiding the initial entity shape interpretation during garden-path processing) less likely to occur. Filler and training sentences were adapted in an analogous manner. Table 1 provides further sample materials.

Procedure

The procedure was identical to that of Experiment 1.

Design and data analysis

The design was the same as in Experiment 1, except for the crucial fact that the sentence-based entity interpretation in control sentences reflected the opposite shape and picture from Experiment 1 (this was always identical to the initial entity interpretation during garden-path processing).[3] All data and R scripts can be found online (https://doi.org/10.5281/zenodo.6504181). Data preprocessing and data analysis fully followed the procedure employed in Experiment 1. Outlier elimination reduced the data set with respect to picture response times and reading times by less than 4%. When performing the method for determining an appropriate random effects structure for the mixed models, some more complex models were rejected due to convergence issues or a singular fit. Ultimately, the model for analyzing picture response times included random intercepts for participants and items and by-item random slopes for sentence type, picture type, and the interaction of sentence type and picture type. The model for reading times comprised random intercepts for participants and items as well as by-item random slopes for sentence type.

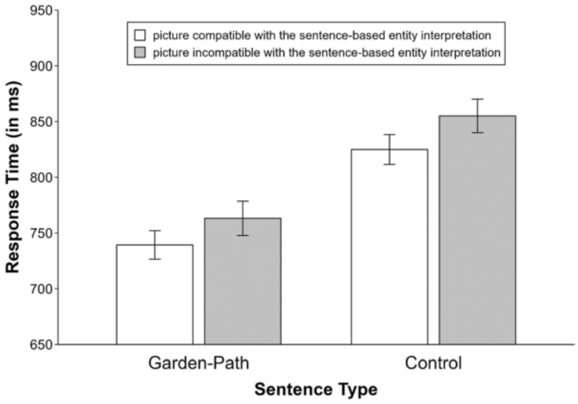


Fig. 3 Mean response times for the sentence–picture verification task in Experiment 2. Error bars denote 95% within-subjects confidence intervals calculated as recommended by Morey (2008)

Results and discussion

There again was a significant effect of sentence type on reading times, $\chi^2(1) = 67.55$, $p < .001$. This time, participants needed more time to read control sentences ($M = 2322$ ms) than garden-path sentences ($M = 1946$ ms), which is not surprising given the fact that control sentences were now longer due to the additional noun phrase. Figure 3 provides an overview of the mean response times in the sentence–picture verification task. The effect of sentence type was significant, $\chi^2(1) = 35.64$, $p < .001$. Participants responded faster to pictures after garden-path sentences ($M = 751$ ms) than after control sentences ($M = 840$ ms). More importantly, the effect of picture type was also significant, $\chi^2(1) = 11.11$, $p < .001$. Just as in Experiment 1, response times were faster when the picture probe matched ($M = 781$ ms) compared with mismatched ($M = 808$ ms) the sentence-based entity interpretation.[4] Once again, this effect was not significantly modulated by sentence type, $\chi^2(1) = 0.00$, $p = .946$.

Thus, all critical effects remained as observed in the first experiment. Most importantly, participants again responded faster to picture probes compatible with the sentence-based entity interpretation than to picture probes incompatible with the sentence-based entity interpretation. Since the sentence-based entity interpretation in control sentences was related to the opposite shape from the first experiment, we can rule out effects of picture preference or aggregate

---

[3] Please note that we used the same levels for the factor "picture type" as in Experiment 1, even though other levels were mentioned in the preregistration for Experiment 2. We did so to be consistent across all experiments.

[4] As participants barely produced erroneous responses and error rates did not differ significantly (picture compatible with sentence-based entity interpretation: 2.44%; picture incompatible with sentence-based entity interpretation: 3.03%), we can rule out a speed–accuracy tradeoff.

word association bias. This reinforces the conclusion that participants indeed generated sentence-based simulations. In line with the results of the first experiment, there was no evidence suggesting that participants created incremental simulations during garden-path processing. However, this time, response times to picture probes as well as reading times were faster in trials with garden-path sentences than in trials with control sentences. This could be due to several reasons. First, the control sentences now included an additional noun phrase and were therefore longer than the garden-path sentences. Second, research stemming from the context of the discourse model approach indicates that establishing and accessing more referents produces longer processing times (e.g., Murphy, 1984). It is conceivable that this is the result of mentally simulating the additional referent. Third, encountering two referents one after another—such as in the control sentence "As Mary ate the egg the butter was in the fridge"—may have provoked a shift of the focus to the entity mentioned second (i.e., the butter; foreground), thus leaving the actual target entity shown on the picture (i.e., the egg) in the background (for a review on the use of situation models in research on language comprehension including foregrounding, see Zwaan & Radvansky, 1998).

General discussion

Over the last few years, embodied accounts of cognition have become increasingly important in theories and empirical research on human language comprehension. Crucially, these approaches (e.g., Barsalou, 1999; Glenberg & Kaschak, 2002; Zwaan & Madden, 2005) propose that we conceive the meaning of language via mental simulations, which are created by means of activating and combining sensorimotor experiences related to the referents of the linguistic input (e.g., words, phrases, or idioms). As of yet, however, there has been little focus on the processes underlying meaning composition during embodied sentence comprehension. In fact, prior research largely leaves it unknown whether reported effects resulted from sentence-based simulations or simulations based on single words or bags of words. In two experiments, we aimed to overcome this issue. We also investigated whether language comprehenders create incremental simulations during sentence processing, an aspect that has also been treated in a very limited way in research on sentence comprehension.

We presented participants with a sentence–picture verification task. In each trial, their task was to decide whether the entity shown in the picture was mentioned in the sentence they had previously read. Importantly, the sentential materials consisted of garden-path sentences (e.g., "While Mary ate the egg was in the fridge") and unambiguous control sentences (e.g., "The egg was in the fridge while Mary ate"). Both for garden-path sentences and control sentences, we expected participants to respond faster to a picture probe showing the final sentence-based entity shape interpretation (e.g., an unpeeled egg in its shell). Moreover—as the other picture probe always displayed the initial entity shape interpretation incrementally created during garden-path processing (e.g., a ready-to-eat-egg)—we also hypothesized that this effect should be less pronounced for garden-path sentences than for unambiguous control sentences.

In the first experiment, participants indeed responded significantly faster when the picture probe was compatible with the sentence-based entity shape interpretation. However, as this effect was not modulated by sentence type, we did not find any evidence indicating that comprehenders create incremental simulations during sentence comprehension. In the follow-up experiment, we intended to exclude the possibility that the observed effect simply occurred because there was a general preference for the pictures displaying the sentence-based entity interpretation. Thus, we adapted the materials in such a way that the sentence-based entity shape interpretation in control sentences was related to the opposite shape and picture than in the first experiment. Nevertheless, participants again responded significantly faster when the picture probe corresponded to the sentence-based entity shape interpretation. Just as in the first experiment, this effect did not vary as a function of sentence type.

Most importantly, these results clearly indicate that participants tended to create sentence-based mental simulations when facing both garden-path and unambiguous sentences. Compared with previous research that might suggest similar conclusions (e.g., Glenberg & Kaschak, 2002; Zwaan et al., 2002), the sentential materials of the current experiments were constructed to be less prone to word-based interpretations. By ruling out certain alternative explanations, the findings presented here enrich the empirical evidence that mental simulation in sentence comprehension is compositional.

As do many studies using similar methods, the current work leaves unresolved the issue of whether mental simulations are functionally implicated in understanding linguistic meaning. It remains conceivable that sentence-based mental simulations are only formed after an amodal symbolic meaning composition has taken place. In this case, mental simulations of sentential meaning would constitute a by-product of the language comprehension processes. Although functional relevance is clearly one of the most interesting and pressing open issues in the area of embodied language comprehension, there has been little behavioral research on this topic. In addition, the few existing studies on functional role have largely focused on word-based mental simulations (e.g., Shebani & Pulvermüller, 2013; Strozyk et al., 2019; Yee et al., 2013). Future research is needed that comes up with research methods suited to produce informative results regarding the functional relevance of sentence-based mental simulations.

However, even though our experiments were not explicitly designed to examine the functional role of sentence-based mental simulations, they still provide at least some preliminary insights into this issue. In both experiments, we found no evidence indicating that participants formed incremental simulations during the processing of garden-path sentences. This suggests that mental simulations might not functionally contribute to understanding the meaning of language—indeed, prior research (e.g., Christianson et al., 2001; Patson et al., 2009; Slattery et al., 2013) has convincingly shown that comprehenders build a representation of the initial (i.e., incremental) entity interpretation when processing the sort of garden-path sentences used in the here reported experiments. Thus, the absence of evidence for incremental simulations could propose that generating mental simulations is an optional by-product of language comprehension processes.

However, the observed pattern of results (i.e., no evidence for mental simulations of the initial entity interpretation during garden-path processing) is inconsistent with prior findings by Sato et al. (2013), who also examined incremental simulations of entity shapes. Critically—and in contrast to the current experiments—they used a sentence–picture verification task including a picture probe located in the middle of the sentence. On the one hand, it is clearly possible that this procedure prompted participants to form the object shape simulation in question and thus artificially induced their observations pointing towards the existence of incremental simulations. On the other hand, a direct approach like the one they adopted could be necessary to prevent detection problems due to the de-activation or overwriting of early, incremental simulations in the further course of reading. Interestingly, Hoeben Mannaert et al. (2019) recently provided some initial insights into this issue when investigating the dynamics of mental simulations *across several* sentences. They presented participants with short narratives of two or four sentences. Most importantly, these narratives differed in whether they implied a change in the shape of a target entity (e.g., change of shape: "The eagle was moving through the air. That evening the eagle was resting in its nest."; constant shape: "The eagle was moving through the air. That evening the eagle was still moving through the air."). Then, a picture of the target entity appeared, either matching or mismatching the final entity shape (e.g., an eagle with folded vs. outstretched wings), and participants judged whether the entity was mentioned in the narrative or not. The results showed that response times were significantly faster when the picture matched with the shape implied by the final sentence of the narrative. This effect was not modulated by shape condition (change of shape vs. constant shape), even though in trials with a change of shape an initial and a final simulation of the entity shape should have been created. The authors thus proposed that the simulation of the final entity shape may have replaced the simulation of the initial entity shape. Although this finding refers to the dynamics of mental simulations *across* sentences, similar processes could occur *within* sentences. By this reasoning, our participants may have created incremental simulations but did not retain them sufficiently at the critical picture probe following the sentence presentation, making it hard or even impossible to detect them. On the contrary, there is also research indicating that comprehenders are able to preserve mental simulations of the orientation and shape of an entity over longer periods of time (e.g., Pecher et al., 2009) and that incremental entity interpretations during garden-path processing linger after the sentence has been reanalyzed (e.g., Christianson et al., 2001; Patson et al., 2009; Slattery et al., 2013). Altogether, it currently seems to be premature and inappropriate to draw any definite conclusions regarding the existence or relevance of incremental simulations during sentence comprehension.

Finally, the lack of evidence in favor of incremental simulations could also result from participants employing strategies to avoid forming the initial entity interpretation during garden-path processing. In the first experiment, garden-path and control sentences differed structurally, which subjects could have learned to attend to in order to process garden-path sentences more efficiently. We tried to overcome this issue by limiting the number of sentences that followed the garden-path structure—three fourths of the fillers had similar structure to the unambiguous control sentences. The intent was to make it harder for participants to become accustomed to garden-path sentences and develop specific processing strategies. Moreover, in the second experiment, participants should have had a harder time discriminating garden-path and control sentences at first glance, since they were identical for roughly the first half. Nonetheless, the pattern of results was similar in both experiments. This seems to argue for the validity of the procedure and against artefacts due to specific processing strategies.

In general, the sentence–picture verification framework used in the current experiments constitutes an important and well-established standard paradigm in the research on embodied language comprehension. Numerous behavioral studies rely on such tasks to investigate which aspects of meaning comprehenders tend to simulate when they encounter written linguistic stimuli, including the shape, size, color, orientation, and visibility of objects described in a sentence (e.g., Connell, 2007; de Koning et al., 2017a, 2017b; Stanfield & Zwaan, 2001; Yaxley & Zwaan, 2007; Zwaan & Pecher, 2012; Zwaan et al., 2002). More recently, the sentence–picture verification task has also served as a tool for revealing dynamic changes of mental simulations across sentences (see Hoeben Mannaert et al., 2019). Nevertheless, it is reasonable to question whether the approach is sufficiently sensitive to detect incremental processing steps during sentence comprehension because it does not allow for a direct look at on-line processes (for more general criticism on the validity of the sentence–picture verification task, see Ostarek et al., 2019).

In sum, we found evidence indicating that comprehenders tend to build sentence-based mental simulations. Importantly, the sentential materials were constructed to be less sensitive to simulation effects resulting from single words or bags of words than materials used in previous studies. Certainly, it remains an important and unanswered question whether these sentence-based simulations are functionally relevant for language comprehension. Beyond that, the pattern of results found no evidence that participants created incremental simulations related to the initial entity shape interpretation during garden-path processing. However, it cannot be ruled out that this finding emerged from methodological characteristics of the experiments. Future research needs to test the stability of these results by means of modified paradigms.

In conclusion, our experiments provide clear evidence for the idea that simulation effects are not limited to the word level but pertain to sentential meaning as well. Hence, our findings confirm a core assumption of embodied views of human language comprehension, which—as of yet—has not received much conclusive support in the literature. Furthermore, the results suggest that mental simulations might be created globally after meaning composition has taken place.

Funding Open Access funding enabled and organized by Projekt DEAL. This research was supported by the Deutsche Forschungsgemeinschaft (DFG)—Project IDs 75650358 and 419433647.

Data availability All data generated during this research can be found online (https://doi.org/10.5281/zenodo.6504181). Experiment 2 was preregistered (https://aspredicted.org/tq6p9.pdf).

Code availability All R analysis scripts are publicly available and can be found at the same place as the data.

Declarations

Ethics approval Ethics approval for this research was obtained from the Ethics Committee for Psychological Research at the University of Tübingen (Identifier: 2018_0831_132).

Conflicts of interest We have no conflicts of interest to disclose.

References

Ahlberg, D. K., Bischoff, H., Kaup, B., Bryant, D., & Strozyk, J. V. (2018). Grounded cognition: Comparing language × space interactions in first language and second language. *Applied Psycholinguistics, 39*(2), 437–459. https://doi.org/10.1017/S014271641700042X

Ansorge, U., Kiefer, M., Khalid, S., Grassl, S., & König, P. (2010). Testing the theory of embodied cognition with subliminal words. *Cognition, 116*(3), 303–320. https://doi.org/10.1016/j.cognition.2010.05.010

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*(4), 577–660. https://doi.org/10.1017/S0140525X99002149

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

Bergen, B. K. (2012). *Louder than words: The new science of how the mind makes meaning*. Basic Books.

Bergen, B. K., & Wheeler, K. (2010). Grammatical aspect and mental simulation. *Brain and Language, 112*(3), 150–158. https://doi.org/10.1016/j.bandl.2009.07.002

Bergen, B. K., Lindsay, S., Matlock, T., & Narayanan, S. (2007). Spatial and linguistic aspects of visual imagery in sentence comprehension. *Cognitive Science, 31*(5), 733–764. https://doi.org/10.1080/03640210701530748

Bidet-Ildei, C., Gimenes, M., Toussaint, L., Almecija, Y., & Badets, A. (2017). Sentence plausibility influences the link between action words and the perception of biological human movements. *Psychological Research, 81*(4), 806–813. https://doi.org/10.1007/s00426-016-0776-z

Bidet-Ildei, C., Gimenes, M., Toussaint, L., Beauprez, S.-A., & Badets, A. (2017). Painful semantic context modulates the relationship between action words and biological movement perception. *Journal of Cognitive Psychology, 29*(7), 821–831. https://doi.org/10.1080/20445911.2017.1322093

Bidet-Ildei, C., Beauprez, S.-A., & Badets, A. (2020). A review of literature on the link between action observation and action language: Advancing a shared semantic theory. *New Ideas in Psychology, 58*, 100777. https://doi.org/10.1016/j.newideapsych.2019.100777

Bub, D. N., & Masson, M. E. J. (2010). On the nature of hand-action representations evoked during written sentence comprehension. *Cognition, 116*(3), 394–408. https://doi.org/10.1016/j.cognition.2010.06.001

Capuano, F., Claus, B., & Kaup, B. (2022). The experiential basis of compatibility effects in reading-by-rotating paradigms. *Psychological Research.* Advance online publication. https://doi.org/10.1007/s00426-022-01663-1

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

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology, 42*(4), 368–407. https://doi.org/10.1006/cogp.2001.0752

Claus, B. (2015). Verb gapping: An action-gap compatibility study. *Acta Psychologica, 156*, 104–113. https://doi.org/10.1016/j.actpsy.2014.07.009

Connell, L. (2007). Representing object colour in language comprehension. *Cognition, 102*(3), 476–485. https://doi.org/10.1016/j.cognition.2006.02.009

de Koning, B. B., Wassenburg, S. I., Bos, L. T., & van der Schoot, M. (2017a). Mental simulation of four visual object properties: Similarities and differences as assessed by the sentence–picture verification task. *Journal of Cognitive Psychology, 29*(4), 420–432. https://doi.org/10.1080/20445911.2017.1281283

de Koning, B. B., Wassenburg, S. I., Bos, L. T., & van der Schoot, M. (2017b). Size does matter: Implied object size is mentally simulated during language comprehension. *Discourse Processes, 54*(7), 493–503. https://doi.org/10.1080/0163853X.2015.1119604

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods, 47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

Dudschig, C., Lachmair, M., de la Vega, I., De Filippis, M., & Kaup, B. (2012). Do task-irrelevant direction-associated motion verbs affect action planning? Evidence from a Stroop paradigm. *Memory & Cognition, 40*(7), 1081–1094. https://doi.org/10.3758/s13421-012-0201-9

Dudschig, C., Souman, J., Lachmair, M., de la Vega, I., & Kaup, B. (2013). Reading "sun" and looking up: The influence of language on saccadic eye movements in the vertical dimension. *PLOS ONE, 8*(2), e56872. https://doi.org/10.1371/journal.pone.0056872

Dudschig, C., de la Vega, I., & Kaup, B. (2014). Embodiment and second-language: Automatic activation of motor responses during processing spatially associated L2 words and emotion L2 words in a vertical Stroop paradigm. *Brain and Language, 132*, 14–21. https://doi.org/10.1016/j.bandl.2014.02.002

Dudschig, C., de la Vega, I., & Kaup, B. (2015). What's up? Emotion-specific activation of vertical space during language processing. *Acta Psychologica, 156*, 143–155. https://doi.org/10.1016/j.actpsy.2014.09.015

Dunn, B., Kamide, Y., & Scheepers, C. (2014). Hearing "moon" and looking up: Word-related spatial associations facilitate saccades to congruent locations. *Proceedings of the Annual Meeting of the Cognitive Science Society, 36*, 433–438 https://escholarship.org/uc/item/78z895zd

Ferreira, F., & Henderson, J. M. (1991). Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language, 30*(6), 725–745. https://doi.org/10.1016/0749-596X(91)90034-H

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14*(2), 178–210. https://doi.org/10.1016/0010-0285(82)90008-1

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review, 9*(3), 558–565. https://doi.org/10.3758/BF03196313

González, J., Barros-Loscertales, A., Pulvermüller, F., Meseguer, V., Sanjuán, A., Belloch, V., & Ávila, C. (2006). Reading cinnamon activates olfactory brain regions. *NeuroImage, 32*(2), 906–912. https://doi.org/10.1016/j.neuroimage.2006.03.037

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

Hoeben Mannaert, L. N., Dijkstra, K., & Zwaan, R. A. (2019). How are mental simulations updated across sentences? *Memory & Cognition, 47*(6), 1201–1214. https://doi.org/10.3758/s13421-019-00928-2

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

Kaup, B., Yaxley, R. H., Madden, C. J., Zwaan, R. A., & Lüdtke, J. (2007). Experiential simulations of negated text information. *Quarterly Journal of Experimental Psychology, 60*(7), 976–990. https://doi.org/10.1080/17470210600823512

Kaup, B., Lüdtke, J., & Steiner, I. (2012). Word- vs. sentence-based simulation effects in language comprehension. In B. Stolterfoht & S. Featherston (Eds.), *Empirical approaches to linguistic theory: Studies in meaning and structure* (pp. 267–290). De Gruyter Mouton. https://doi.org/10.1515/9781614510888

Kaup, B., de la Vega, I., Strozyk, J., & Dudschig, C. (2016). The role of sensorimotor processes in meaning composition. In M. H. Fischer & Y. Coello (Eds.), *Foundations of embodied cognition: Conceptual and interactive embodiment* (pp. 46–70). Routledge. https://doi.org/10.4324/9781315751962

Kumle, L., Võ, M. L. H., & Draschkow, D. (2021). Estimating power in (generalized) linear mixed models: An open introduction and tutorial in R. *Behavior Research Methods, 53*(6), 2528–2543. https://doi.org/10.3758/s13428-021-01546-0

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*(6), 1180–1188. https://doi.org/10.3758/s13423-011-0158-x

Masson, M. E. J., Bub, D. N., & Newton-Taylor, M. (2008). Language-based access to gestural components of conceptual knowledge. *Quarterly Journal of Experimental Psychology, 61*(6), 869–882. https://doi.org/10.1080/17470210701623829

Masson, M. E. J., Bub, D. N., & Warren, C. M. (2008). Kicking calculators: Contribution of embodied representations to sentence comprehension. *Journal of Memory and Language, 59*(3), 256–265. https://doi.org/10.1016/j.jml.2008.05.003

Masson, M. E. J., Bub, D. N., & Lavelle, H. (2013). Dynamic evocation of hand action representations during sentence comprehension. *Journal of Experimental Psychology: General, 142*(3), 742–762. https://doi.org/10.1037/a0030161

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language, 94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

Miller, J., Brookie, K., Wales, S., Wallace, S., & Kaup, B. (2018). Embodied cognition: Is activation of the motor cortex essential for understanding action verbs? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(3), 335–370. https://doi.org/10.1037/xlm0000451

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology, 4*(2), 61–64. https://doi.org/10.20982/tqmp.04.2.p061

Morey, R. D., Kaschak, M. P., Díez-Álamo, A. M., Glenberg, A. M., Zwaan, R. A., Lakens, D., Ibáñez, A., García, A., Gianelli, C., Jones, J. L., Madden, J., Alifano, F., Bergen, B., Bloxsom, N. G., Bub, D. N., Cai, Z. G., Chartier, C. R., Chatterjee, A., Conwell, E., ... Ziv-Crispel, N. (2022). A pre-registered, multi-lab non-replication of the action-sentence compatibility effect (ACE). *Psychonomic Bulletin & Review, 29*(2), 613–626. https://doi.org/10.3758/s13423-021-01927-8

Murphy, G. L. (1984). Establishing and accessing referents in discourse. *Memory & Cognition, 12*(5), 489–497. https://doi.org/10.3758/BF03198311

Ostarek, M., Joosen, D., Ishag, A., de Nijs, M., & Huettig, F. (2019). Are visual processes causally involved in "perceptual simulation" effects in the sentence–picture verification task? *Cognition, 182*, 84–94. https://doi.org/10.1016/j.cognition.2018.08.017

Öttl, B., Dudschig, C., & Kaup, B. (2017). Forming associations between language and sensorimotor traces during novel word learning. *Language and Cognition, 9*(1), 156–171. https://doi.org/10.1017/langcog.2016.5

Patson, N. D., Darowski, E. S., Moon, N., & Ferreira, F. (2009). Lingering misinterpretations in garden-path sentences: Evidence from a paraphrasing task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(1), 280–285. https://doi.org/10.1037/a0014276

Pecher, D., Van Dantzig, S., Zwaan, R. A., & Zeelenberg, R. (2009). Language comprehenders retain implied shape and orientation of objects. *Quarterly Journal of Experimental Psychology, 62*(6), 1108–1114. https://doi.org/10.1080/17470210802633255

Pickering, M. J., & Traxler, M. J. (1998). Plausibility and recovery from garden paths: An eye-tracking study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(4), 940–961. https://doi.org/10.1037/0278-7393.24.4.940

Sato, M., Schafer, A. J., & Bergen, B. K. (2013). One word at a time: Mental representations of object shape change incrementally during sentence processing. *Language and Cognition, 5*(4), 345–373. https://doi.org/10.1515/langcog-2013-0022

Schütt, E., Mackenzie, I. G., Kaup, B., & Dudschig, C. (2022). Replacing vertical actions by mouse movements: A web-suited paradigm for investigating vertical spatial associations. *Psychological Research*. Advance online publication. https://doi.org/10.1007/s00426-022-01650-6

Shebani, Z., & Pulvermüller, F. (2013). Moving the hands and feet specifically impairs working memory for arm- and leg-related action words. *Cortex, 49*(1), 222–231. https://doi.org/10.1016/j.cortex.2011.10.005

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2021). *Afex: Analysis of factorial experiments* (R package) [Computer software]. https://cran.r-project.org/package=afex

Slattery, T. J., Sturt, P., Christianson, K., Yoshida, M., & Ferreira, F. (2013). Lingering misinterpretations of garden path sentences arise from competing syntactic representations. *Journal of Memory and Language, 69*(2), 104–120. https://doi.org/10.1016/j.jml.2013.04.001

Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science, 12*(2), 153–156. https://doi.org/10.1111/1467-9280.00326

Strozyk, J. V., Dudschig, C., & Kaup, B. (2019). Do I need to have my hands free to understand hand-related language? Investigating the functional relevance of experiential simulations. *Psychological Research, 83*(3), 406–418. https://doi.org/10.1007/s00426-017-0900-8

Taylor, L. J., & Zwaan, R. A. (2008). Motor resonance and linguistic focus. *Quarterly Journal of Experimental Psychology, 61*(6), 896–904. https://doi.org/10.1080/17470210701625519

Thornton, T., Loetscher, T., Yates, M. J., & Nicholls, M. E. R. (2013). The highs and lows of the interaction between word meaning and space. *Journal of Experimental Psychology: Human Perception and Performance, 39*(4), 964–973. https://doi.org/10.1037/a0030467

van Gompel, R. P. G., Pickering, M. J., Pearson, J., & Jacob, G. (2006). The activation of inappropriate analyses in garden-path sentences: Evidence from structural priming. *Journal of Memory and Language, 55*(3), 335–362. https://doi.org/10.1016/j.jml.2006.06.004

Vogt, A., Kaup, B., & Dudschig, C. (2019). When words are upside down: Language-space associations in children and adults. *Journal of Experimental Child Psychology, 186*, 142–158. https://doi.org/10.1016/j.jecp.2019.06.001

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

Yaxley, R. H., & Zwaan, R. A. (2007). Simulating visibility during language comprehension. *Cognition, 105*(1), 229–236. https://doi.org/10.1016/j.cognition.2006.09.003

Yee, E., Chrysikou, E. G., Hoffman, E., & Thompson-Schill, S. L. (2013). Manual experience shapes object representations. *Psychological Science, 24*(6), 909–919. https://doi.org/10.1177/0956797612464658

Zwaan, R. A., & Madden, C. J. (2005). Embodied sentence comprehension. In D. Pecher & R. A. Zwaan (Eds.), *Grounding cognition: The role of perception and action in memory, language, and thinking* (pp. 224–245). Cambridge University Press. https://doi.org/10.1017/CBO9780511499968.010

Zwaan, R. A., & Pecher, D. (2012). Revisiting mental simulation in language comprehension: Six replication attempts. *PLOS ONE, 7*(12), e51382. https://doi.org/10.1371/journal.pone.0051382

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123*(2), 162–185. https://doi.org/10.1037/0033-2909.123.2.162

Zwaan, R. A., & Taylor, L. J. (2006). Seeing, acting, understanding: Motor resonance in language comprehension. *Journal of Experimental Psychology: General, 135*(1), 1–11. https://doi.org/10.1037/0096-3445.135.1.1

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science, 13*(2), 168–171. https://doi.org/10.1111/1467-9280.00430

Appendix D: Article 4

Copyright notice

To cite the article

Schütt, E., Weicker, M., & Dudschig, C. (2024). Multimodal aspects of sentence comprehension: Do facial and color cues interact with processing negated and affirmative sentences? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *50*(6), 957-966. https://doi.org/10.1037/xlm0001302

Journal of Experimental Psychology:
Learning, Memory, and Cognition

ISSN: 0278-7393

2024, Vol. 50, No. 6, 957–966
https://doi.org/10.1037/xlm0001302

Multimodal Aspects of Sentence Comprehension: Do Facial and Color Cues Interact With Processing Negated and Affirmative Sentences?

Emanuel Schütt[1], Merle Weicker[2], and Carolin Dudschig[1]
[1] Department of Psychology, Language and Cognition Research Group, University of Tübingen
[2] Department of Psycholinguistics and Didactics of German, Goethe University Frankfurt/Main

Negation is usually considered as a linguistic operator reversing the truth value of a proposition. However, there are various ways to express negation in a multimodal manner. It still remains an unresolved issue whether nonverbal expressions of negation can influence linguistic negation comprehension. Based on extensive evidence demonstrating that language comprehenders are able to instantly integrate extralinguistic information such as a speaker's identity, we expected that nonverbal cues of negation and affirmation might similarly affect sentence comprehension. In three preregistered experiments, we examined how far nonverbal markers of negation and affirmation—specifically, the so-called "not face" (see Benitez-Quiroz et al., 2016) and red or green color (see Dudschig et al., 2023)—interact with comprehending negation and affirmation at the sentential level. Participants were presented with photos ("not face" vs. positive control; Experiments 1 and 2) or color patches (red vs. green; Experiment 3). They then read negated and affirmative sentences in a self-paced manner or judged the sensibility of negated and affirmative sentences (e.g., "No, I do not want to sing" vs. "Yes, I would like to buy a sofa"). Both frequentist statistics and Bayes factors resulting from linear mixed-effects analyses showed that processing times for negated and affirmative sentences were not significantly modulated by the nonverbal features under investigation. This indicates that their influence might not extend to sentential negation or affirmation comprehension.

Keywords: language comprehension, multimodal negation and affirmation, facial expressions, color

Even though individuals from other primate species also tend to communicate with each other, the human language system and its complexity are unique (e.g., Zuberbühler, 2015). One fundamental aspect of human language is the ability to negate, which is universally present in all natural languages (e.g., Zeijlstra, 2007). By contrast, the existence of negation in nonhuman species cannot be taken for granted (Cameron, 1991; Dautriche et al., 2022; Schneider et al., 2010). Negation is a complex phenomenon with diverse manifestations. The most basic function of negation expressed earliest in the human lifespan appears to be rejection and refusal (e.g., in terms of saying "no" and pushing an object away); another rather early available but linguistically and cognitively more demanding type of negation is the expression of nonexistence, disappearance, or unfulfilled expectations (e.g., in terms of saying "no apple" to indicate that there is no apple in place; Dimroth, 2010; Litowitz, 1998; Pea, 1980). Probably, logical denial (i.e., the reversal of the truth value of a proposition as in "It is not raining") can be considered as the most mature and sophisticated function of negation. In addition to purely verbal markers (e.g., "no" or "not"), means of nonverbal communication can also be associated with negative meaning. These nonverbal expressions of negation include gestures like the head shake or the vertically oriented open palm (e.g., Bressem & Müller, 2014; Harrison, 2018; Kendon, 2002, 2004). Most importantly for the current work, recent research findings indicated that producing negated utterances seems to be accompanied by the so-called "not face"—a unique negation marker combining facial signs of contempt, anger, and disgust, which was shown to be articulated by speakers of different languages and with different cultural backgrounds (Benitez-Quiroz et al., 2016). Another recent study (Dudschig et al., 2023) revealed that performing "yes" ("no") responses in a lexical decision task is facilitated when the color of the response button is green (red). In the present work, we aimed to explore whether the effects of the "not face" and of red and green color cues might extend to the comprehension of sentential negation and affirmation.

Traditionally, negation is understood as a verbal linguistic operator undertaking specific functions within sentences. In psycholinguistics, it has regularly been demonstrated that negation is rather difficult to integrate during meaning composition, resulting in longer processing times, erroneous responses, and nonintegration as reflected in electrophysiological measures (e.g., Dudschig et al., 2018, 2019; Fischler et al., 1983; for an overview, see Kaup & Dudschig, 2020). Given the complexity of negation, this might be

---

Emanuel Schütt https://orcid.org/0000-0002-8257-167X

This research was supported by the Deutsche Forschungsgemeinschaft (DFG; Project ID 419433647). The authors thank two anonymous reviewers for valuable comments on a previous version of this article and Lea Köster and Jonathan Göth for assistance with stimulus creation and data collection. The authors have no conflicts of interest to disclose. All experiments were preregistered prior to starting data collection (Experiment 1: https://aspredicted.org/dj85z.pdf; Experiment 2: https://aspredicted.org/m4de2.pdf; Experiment 3: https://aspredicted.org/5d62j.pdf). Sentence materials, experimental data, and analysis scripts are publicly available online (https://doi.org/10.5281/zenodo.7389652).

Correspondence concerning this article should be addressed to Emanuel Schütt, Department of Psychology, Language and Cognition Research Group, University of Tübingen, Schleichstraße 4, 72076 Tübingen, Germany. Email: emanuel.schuett@uni-tuebingen.de

considered a plausible finding at first glance. On the other hand, such observations are surprising, because we frequently use negated expressions in everyday language. Researchers taking up this discrepancy suggested that integration difficulties observed in studies investigating negation processing could emerge from a pragmatically infelicitous use of negation. For instance, Nieuwland and Kuperberg (2008) explored in their seminal study the influence of truth value and pragmatic licensing on processing affirmative and negated statements. In licensed negated sentences, words that made a proposition false produced a larger N400 than words that made a proposition true (e.g., "With proper equipment, scuba-diving isn't very *safe* vs. *dangerous*"), which corresponded to the pattern of results observed for affirmative sentences. In nonlicensed negated sentences, however, both types of words induced a similar N400 (e.g., "Bulletproof vests aren't very *dangerous* vs. *safe*"). Consequently, the authors suggested that pragmatically licensed negation is incrementally integrated and does not produce additional processing costs.

Even though psycholinguistic research has focused on investigating negation in terms of a verbal linguistic operator, it is well known that negation can also be expressed nonverbally. In a recent study, Benitez-Quiroz et al. (2016) instructed participants to show a facial expression of negation. Before the participants were photographed, they were given some examples of specific situations that could evoke negative reactions. The authors found that participants with different ethnic and cultural backgrounds consistently expressed negation using a unique subset of facial articulations that are related to negative moral judgment, involving expressions of anger, disgust, and contempt. They concluded that this "not face" constitutes a universal facial expression of negation. In a further step, the authors could also demonstrate that the "not face" occurs when reproducing negated sentences, but virtually not when reproducing affirmative sentences. They suggested that specific facial expressions of emotions, including the "not face", should be considered the evolutionary origin of verbal grammatical markers, thus probably carrying semantic meaning. In addition, in another recent study, Dudschig et al. (2023) showed that negative and positive facial expressions condensed in an emoticon (frowning vs. smiling) are able to modulate the speed of performing "no" and "yes" decisions. A second possible visual way of expressing negation explored by the same authors was red color. Specifically, participants were presented with words and nonwords (e.g., "guitar" vs. "mipe") in the context of a lexical decision task. Two response button symbols appearing next to each other on the screen indicated whether the left or the right response key had to be pressed to provide a "yes" or a "no" response. Importantly, in each trial, one of these response button symbols was red-colored, whereas the other one was green-colored. The assignment of colors and response labels ("yes" vs. "no") to response button symbols was randomized on a trial-by-trial basis. The participants performed "no" responses faster when the response button symbol denoting the "no" response was red-colored (compared to green-colored) and "yes" responses faster when the response button symbol denoting the "yes" response was green-colored (compared to red-colored). This suggests that "no" responses are associated with red color and "yes" responses with green color. In the present research, we asked whether the outlined findings on the effects of facial expressions and color information also apply to the comprehension of sentential negation and affirmation.

In particular, we aimed to examine whether visual negation in terms of the "not face" or red color cues and visual affirmation in terms of green color cues can influence the processing of negated or affirmative sentences. The primary objective was to explore whether presenting negation or affirmation in a multimodal manner facilitates comprehension, thus indicating that visual cues are rapidly integrated in a one-step fashion during the processing of negated and affirmative sentences. We started by showing that the "not face" and the color cues under investigation are associated with negation or affirmation in an explicit rating task. Subsequently, these visual cues served as primes in standard reading tasks. Previous studies using similar setups demonstrated that visual information can indeed be integrated with verbal information early during sentence comprehension. For instance, in a study by Rück et al. (2017), participants saw pictures of male and female faces while they were reading sentences phrase by phrase. These sentences matched or mismatched the prototypical gender identity conveyed by the pictures (e.g., "Last week, I bought an elegant *dress* [mismatch for male; match for female] vs. *tuxedo* [match for male; mismatch for female] for my sister's wedding"). When the picture and sentence content mismatched, reading times for the critical phrase (e.g., the word "dress" or "tuxedo") were prolonged. This indicates that participants rapidly integrated the visual information about the speaker's gender identity during sentence comprehension, thus supporting the view that verbal and visual information is processed in a one-step fashion. Moreover, in a second experiment with negated sentences (e.g., "Last week, I did not buy an elegant *dress* [match for female; plausible mismatch for male] vs. *tuxedo* [plausible mismatch for female; match for male] for my sister's wedding"), reading times suggested that visual information about the speaker is taken up early during comprehension, but that integration in the sentence meaning might need some time. Apart from considering extralinguistic information arising from depictions of faces, comprehenders are also able to quickly integrate the voice of a speaker (e.g., Van Berkum et al., 2008), event-based expectations (e.g., Matsuki et al., 2011), and world knowledge (e.g., Dudschig et al., 2016a, 2016b; Hagoort et al., 2004) when processing language.

Based on the described evidence convincingly proving that nonverbal information is routinely incorporated during language comprehension, we asked whether nonverbal negation in terms of the "not face" or red and green color information might be integrated during sentence processing. In a series of three experiments, participants read affirmative and negated sentences in a self-paced manner or judged the sensibility of affirmative and negated sentences. Prior to the sentence onset, a "not face" versus a positive control face (Experiments 1 and 2) or a red versus a green color patch (Experiment 3) was displayed. If the "not face" serves as a nonverbal marker for negation that is integrated during language processing, we expected faster reading and sensibility judgment times for negated sentences preceded by a "not face" (compared to negated sentences preceded by a positive control face). Similarly, if red (green) color can be integrated as a marker for negation (affirmation), we hypothesized that sensibility judgment times should be faster for negated (affirmative) sentences preceded by red (green) color patches. All experiments were preregistered on AsPredicted (Experiment 1: https://aspredicted.org/dj85z.pdf; Experiment 2: https://aspredicted.org/m4de2.pdf; Experiment 3: https://aspredicted.org/5d62j.pdf).

Experiment 1

Method

Participants

We determined the sample size by referring to prior experimental work investigating the influence of visual information about the speaker on sentence processing (Rück et al., 2017). For this purpose, we used MorePower (Version 6.0.4; Campbell & Thompson, 2012) to run a power analysis based on the results Rück et al. (2017, Experiment 1) reported for the congruency effect of visual and linguistic information on reading times in a critical phrase of the sentence (by-participants paired t test; $d = 0.25$). The analysis revealed that a total of 130 participants would be required to obtain a test power of .80. Due to balancing demands, the final target sample size was set to 132 participants.

We recruited participants through email messages to the students at the University of Tübingen. We planned to exclude participants with an overall error rate higher than 30% in the comprehension task, which was introduced to make sure that participants read the sentences properly. However, none of the participants scored below this threshold. All participants (92 females, 37 males, and three nonbinaries; 113 right-handers, 17 left-handers, and two ambidexters) reported being native German speakers. Their ages ranged from 18 to 59 years ($M = 26.76$ years, $SD = 9.24$ years). All of them gave informed consent prior to participating. As compensation, participants could either receive partial course credit or take part in a raffle of two vouchers, each with a value of €20. In sum, the experiment took 15–20 min to complete. The study was approved by the Ethics Committee for Psychological Research at the University of Tübingen (Identifier: 2021_0505_226).

Apparatus and Stimuli

We employed the open-source JavaScript library jsPsych (Version 6.1.0; de Leeuw, 2015) to create a browser-based experiment and asked participants to use a desktop computer or laptop to run the experiment (using a smartphone or tablet was generally not possible as crucial jsPsych plugins did not work on mobile devices). Participants pressed the spacebar to indicate that they had read the test sentence. The “c” key and the “m” key of a standard keyboard were the response keys in the comprehension task.

We constructed 80 pairs of simple German sentences, each with an affirmative and a negated counterpart. The sentences were presented in three syntactic frames. Each sentence contained a modal verb (e.g., “wollen” [“to want”]) and either a transitive main verb (e.g., “kaufen” [“to buy”]), an intransitive main verb (e.g., “singen” [“to sing”]), or no main verb at all (which is possible in German when the main verb can be derived from context). Negation was realized by starting the sentence with “Nein” (“no”) and using the negation particle “nicht” if the main verb was intransitive (e.g., “Nein, ich will nicht singen” [“No, I do not want to sing”]) or the negative indefinite determiner “kein/e” combined with the object noun phrase if the main verb was transitive or there was no main verb at all (e.g., “Nein, ich möchte kein Sofa kaufen” [“No, I would not like to buy a sofa”]; “Nein, ich möchte kein Kissen” [“No, I would not like to have a pillow”]). The negation operator always followed the modal verb. By contrast, affirmative sentences started with “Ja” (“yes”). The adverb “jetzt” (“now”) and the indefinite determiner “ein/e” (“a/an”) replaced the negation operators (e.g., “Ja, ich will jetzt singen” [“Yes, I want to sing now”]; “Ja, ich möchte ein Sofa kaufen” [“Yes, I would like to buy a sofa”]; Ja, ich möchte ein Kissen” [“Yes, I would like to have a pillow”]). A list including all experimental sentences can be found online (https://doi.org/10.5281/zenodo.7389652). Four additional negated and affirmative sentences served as stimuli in the training session.

We selected 10 photos showing facial expressions of emotions corresponding to the “not face” as described by Benitez-Quiroz et al. (2016). Another 10 photos displaying positive facial expressions were used as controls. All photos were scaled to a size of 400 (width) × 300 (height) pixels. Twelve raters, who did not participate in the experiment, judged on a 5-point Likert scale (1 = *definitely no*; 2 = *rather no*; 3 = *neutral*; 4 = *rather yes*; 5 = *definitely yes*) to what extent they associated the facial expressions shown on the photos with “no” (negation) or “yes” (affirmation). The rating scores for photos with positive facial expressions differed significantly from the rating scores for photos displaying a “not face”, $t(11) = -17.46$, $p < .001$. Whereas the “not face” was associated with “no” ($M = 1.65$, $SD = 0.50$), positive facial expressions were associated with “yes” ($M = 4.55$, $SD = 0.63$). For copyright reasons, we are not able to make the photos publicly available. However, all photos will be provided upon scientific request.

To ensure that participants processed the sentences properly, they were presented with a comprehension question after 20 of the 80 experimental sentences they read over the course of the experiment. The questions referred to aspects of the sentential content and required a “yes” or “no” response (e.g., “Enthält die Aussage der Person das Wort Strand?” [“Does the statement of the person contain the word beach?”]; “Möchte die Person fotografieren?” [“Does the person want to take photos?“]; “Geht es in der Aussage der Person ums Wandern?” [“Is the statement of the person about hiking?”]).

Procedure

Each trial started with a fixation cross (“+”; 800 ms). Then, a photo showing a “not face” or positive facial expression appeared at the center of the screen; 500 ms after the photo onset, a negated or an affirmative sentence was displayed right below the photo. Participants were asked to read the sentence attentively. Both picture and sentence stayed on the screen until participants clicked on the spacebar to indicate that they had read and understood the sentence. In 25% of the trials, a comprehension question followed. Participants used the “c” key to give a “yes” response and the “m” key to give a “no” response. The intertrial interval was 1,000 ms.

Participants passed a training session before proceeding with two experimental blocks, each including 40 trials. The possible conditions associated with presenting an experimental sentence (polarity: negated vs. affirmative; photo type: “not face” vs. positive control) were counterbalanced creating four lists. The specific photo appearing on the screen was defined by randomly choosing one of the photos from the pool of photos showing a “not face” or positive facial expressions, respectively. We also balanced the conditions within blocks and randomized the order of trials.

Design and Data Analysis

The experiment implemented a 2 × 2 within-subjects design, with the factors sentence polarity (negated vs. affirmative) and photo type (“not face” vs. positive). The time period from the onset of the sentence on the screen until pressing the spacebar (reading time) served as the dependent variable. All experimental data and analysis scripts are publicly available online (https://doi.org/10.5281/zenodo.7389652).

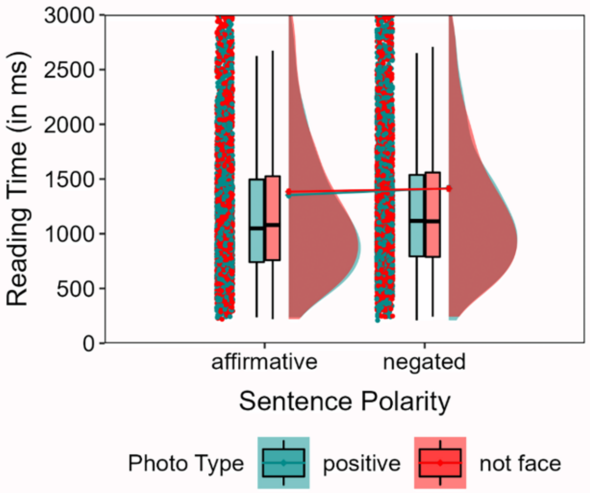
We processed and analyzed the data using the free statistical software R (Version 4.1.1). First, we removed reading times reflecting anticipatory responses (<200 ms) or disruption of performance (>8,000 ms). This reduced the data set by about 1%. After that, we created a linear mixed-effects model (see Baayen et al., 2008) employing the R package lme4 (Version 1.1-27.1; Bates et al., 2015). The model contained fixed effects for sentence polarity, photo type, and the interaction of these factors. To determine the random effects structure, we referred to the model selection criterion suggested by Matuschek et al. (2017), which aims to balance Type I error rate and power. When performing the procedure, several more complex models had to be skipped as these produced convergence or singular fit issues. The final model included random intercepts for participants and items. We explored the significance of fixed effects by performing likelihood ratio tests, which were controlled by the R package afex (Version 1.0-1; Singmann et al., 2021). Furthermore, we also computed Bayes factors (BFs) using the R package BayesFactor (Version 0.9.12-4.3; Morey & Rouder, 2021). The values represent the evidence for H_1 over H_0. Verbal interpretations follow the common classification scheme (e.g., Lee & Wagenmakers, 2014).

Results

Figure 1 shows a raincloud plot (see Allen et al., 2021) for reading times as a function of sentence polarity and photo type. The likelihood ratio tests indicated that there was a significant effect of sentence polarity, $\chi^2(1) = 10.76$, $p < .001$. Similarly, the BF (4.71) suggested that there was moderate evidence for the existence of the effect. Reading times were shorter for affirmative sentences ($M = 1,369$ ms, $SD = 937$ ms) than for negated sentences ($M = 1,415$ ms, $SD = 947$ ms). The effect of photo type turned out not to be significant, $\chi^2(1) = 1.45$, $p = .228$. This was also mirrored by the BF (0.05; strong evidence for the absence of the effect). Most importantly, however, we did not observe a significant interaction of sentence polarity and photo type, $\chi^2(1) = 2.25$, $p = .134$. The BF (0.09) strongly favored the absence of the interaction effect.

Figure 1

Reading Times as a Function of Sentence Polarity and Photo Type in Experiment 1



Note. The raincloud plot was created using R code provided by Allen et al. (2021). See the online article for the color version of this figure.

Experiment 2

In the previous experiment, participants read affirmative and negated sentences in a self-paced manner. As they were not prompted to immediately provide a response associated with the sentence meaning, we cannot be sure that sentence processing was completed when they pressed the spacebar to indicate that they had read the sentence. We thus replaced the self-paced reading task by a sensibility judgment task in Experiment 2. This ensured that participants established a mature representation of the sentence meaning to be able to provide their response.

Method

Participants

For defining the sample size, we referred to the power analysis conducted in the context of Experiment 1 ($N = 130$ for a test power of .80 with respect to the congruency effect of visual information and sentential content on sentence comprehension). We again considered balancing demands and therefore set the target sample size to 136 participants. Participants were recruited by sending an email to the students at the University of Tübingen. We excluded participants with an overall error rate higher than 30% in the sensibility judgment task, which reduced the sample to 134 participants. All participants (104 females, 28 males, and two nonbinaries; 119 right-handers, 13 left-handers, and two ambidexters) indicated to be native German speakers. They were between 18 and 56 years old ($M = 23.01$ years, $SD = 6.04$ years) and provided informed consent. In return for participation, they could receive partial course credit or take part in a raffle of five vouchers, each with a value of €50.[1] The participants needed about 20 min to complete the experiment. The study was approved by the Ethics Committee for Psychological Research at the University of Tübingen (Identifier: 2021_0505_226).

Apparatus and Stimuli

Apparatus and stimuli were largely similar to Experiment 1. However, due to replacing the reading task by a sensibility judgment task, the following aspects were different. First, the "f" key and the "j" key of a standard keyboard were employed as response keys in the sensibility judgment task. Second, we created 40 affirmative and 40 negated nonsensible filler sentences, which had the same structure as the sensible sentences used in Experiment 1 but exhibited semantic violations (e.g., "Ja, ich will eine Glühbirne trösten" ["Yes, I want to solace a lightbulb"]; "Ja, ich will jetzt geschält

---

[1] The guidelines for reimbursing participants via raffles changed at our department. Hence, the number and the value of vouchers differed from Experiment 1.

werden" ["Yes, I want to be peeled now"]; "Nein, ich möchte keine Steckdose unterrichten" ["No, I would not like to teach a wall socket"]; "Nein, ich will keine Arterien jonglieren" ["No, I do not want to juggle with arteries"]). We also constructed two affirmative and two negated nonsensible filler sentences for the training session. Third, comprehension questions were not shown anymore, as the sensibility judgment task required participants to read the sentences properly.

Procedure

The trial procedure was identical to Experiment 1, except that participants indicated as fast as and as accurately as possible whether the displayed sentence was sensible or nonsensible. Both picture and sentence remained on the screen until participants pressed one of the response keys. Half of the participants pressed the "f" key for a "yes" response and the "j" key for a "no" response. For the other half of the participants, the response assignment was reversed. In the training session, we also presented trial-based accuracy feedback ("Richtig!" ["Correct!"] vs. "Falsch!" ["Wrong!"]; 1,000 ms).

The training session comprised two affirmative and two negated sensible sentences and two affirmative and two negated nonsensible sentences. Participants passed four experimental blocks, each including 20 sensible and 20 nonsensible sentences. The conditions related to the presentation of a sensible sentence (polarity: negated vs. affirmative; photo type: "not face" vs. positive control) were balanced across participants using four lists. In contrast, each nonsensible filler sentence was displayed under fixed conditions (i.e., a specific nonsensible filler sentence appeared in the same polarity and was preceded by the same photo type for all participants). The photo on the screen was defined by randomly choosing one of the photos included in the pool of photos showing a "not face" or positive facial expressions, respectively. We counterbalanced the conditions of sensible and nonsensible sentences within blocks and randomized the trial order.

Design and Data Analysis

The factors were the same as in Experiment 1. This time, however, the time period from the onset of the sentence on the screen until pressing one of the response keys (response time) served as dependent variable. All experimental data and analysis scripts are publicly available online (https://doi.org/10.5281/zenodo.7389652).

The procedure for analyzing response times was mostly identical to Experiment 1. We performed the analysis on all trials with sensible sentences, excluding incorrect responses (less than 4% of the trials). Removing anticipatory responses ($<$200 ms) and trials associated with disruption of performance ($>$8,000 ms) reduced the remaining data set by less than1%. Again, we had to reject some more complex models when defining the random effects structure for the linear mixed model due to convergence issues or singular fit warning. The final model included random intercepts for participants and items.

We also analyzed error rates for trials with sensible sentences—a measure not available for the self-paced reading task in Experiment 1. Again, trials related to anticipatory responses or disruption of performance were removed. In contrast to the response time analysis, we created a generalized mixed model (mixed-effects logistic regression) as response accuracy (correct vs. incorrect) was a binary outcome variable. The method for setting the random effects structure resulted in incorporating random intercepts for participants and items. We estimated BFs from the Bayesian information criterion (Wagenmakers, 2007) because the R package BayesFactor (Version 0.9.12-4.3; Morey & Rouder, 2021) does not support computing BFs for generalized mixed models.

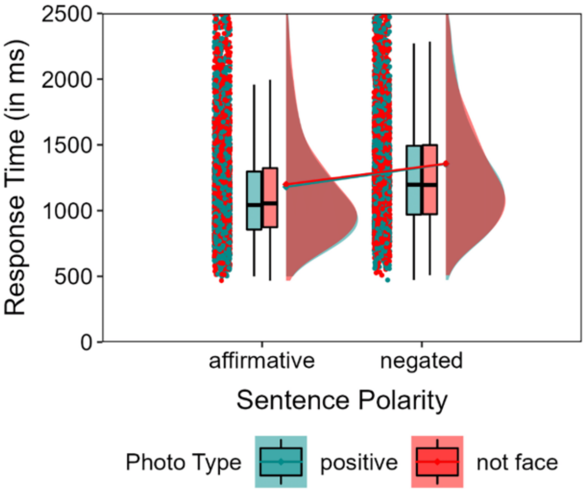
Results

Figure 2 shows a raincloud plot (see Allen et al., 2021) for response times as a function of sentence polarity and photo type. In accordance with Experiment 1, there was a significant effect of sentence polarity, $\chi^2(1) = 339.48$, $p < .001$. The BF ($>$100) provided extreme evidence for the existence of this effect. Thus, participants responded faster when the sentence was affirmative ($M =$ 1,190 ms, $SD = 522$ ms) than when the sentence was negated ($M =$ 1,356 ms, $SD = 608$ ms). The effect of photo type was not significant, $\chi^2(1) = 2.76$, $p = .097$. The BF (0.10) proposed that there was moderate evidence for the absence of the effect. Again, we did not observe a significant interaction of sentence polarity and photo type, $\chi^2(1) = 2.71$, $p = .100$. This was also reflected by the BF (0.13), suggesting that there was moderate evidence for the absence of the interaction.

The analysis of response accuracy revealed that there was a significant interaction, $\chi^2(1) = 13.78$, $p < .001$. Similarly, the BF (9.48) indicated that there was moderate evidence for the existence of the interaction. To further explore this effect, we performed follow-up comparisons using the R package emmeans (Version 1.7.4-1; Lenth, 2022). Interestingly, in trials including affirmative sentences, the error rate was significantly lower when the presented photo showed positive facial expressions ($M = 2.35\%$, $SD =$ 3.91%) than when the presented photo showed a "not face" ($M = 4.44\%$, $SD = 10.60\%$), $z = -4.22$, $p < .001$. By contrast, in trials including negated sentences, the error rate did not differ significantly as a function of photo type ("not face": $M =$

Figure 2

Response Times as a Function of Sentence Polarity and Photo Type in Experiment 2



Note. The raincloud plot was created using R code provided by Allen et al. (2021). See the online article for the color version of this figure.

4.35%, $SD = 6.14\%$; positive control: $M = 4.76\%$, $SD = 8.04\%$), $z = 0.68$, $p = .494$. Due to the specific pattern of the interaction, the explanatory value of the main effects appears to be limited. However, there was a significant main effect of sentence polarity, $\chi^2(1) = 13.27$, $p < .001$, with a BF (1.90) pointing toward anecdotal evidence for the existence of the effect. Participants committed less errors in trials with affirmative sentences ($M = 3.40\%$, $SD = 5.62\%$) than in trials with negated sentences ($M = 4.56\%$, $SD = 5.91\%$). There was also a significant main effect of photo type, $\chi^2(1) = 7.92$, $p = .005$, whereas the BF (0.14) suggested that there was moderate evidence for the absence of this effect. Participants produced less errors when the photo showed positive expressions ($M = 3.56\%$, $SD = 4.92\%$) than when it showed a “not face” ($M = 4.40\%$, $SD = 7.18\%$).

Experiment 3

In this experiment, we investigated the effect of red and green color on the processing of affirmative and negated sentences. Thus, in comparison to the previous experiment, the facial cues were replaced by red and green color patches.

Method

Participants

As in Experiment 2, we collected data from 136 participants. We recruited participants by sending an email to the students at the University of Tübingen and via the online labor marketplace Prolific. We again planned to exclude participants with an overall error rate higher than 30% in the sensibility judgment task. However, there was no participant scoring below this threshold. Hence, the final sample included 136 participants (83 females, 53 males; 122 right-handers, 13 left-handers, and one ambidexter). All of them were confirmed to be native German speakers. Their ages ranged between 18 and 60 years ($M = 29.60$ years, $SD = 10.23$ years). They provided informed consent prior to participating and received partial course credit (student participants) or £3.50 (participants on Prolific) as reimbursement. The participants needed about 20 min to complete the experiment. The study was approved by the Ethics Committee for Psycholog-ical Research at the University of Tübingen (Identifier: 2021_0505_226).

Apparatus and Stimuli

Apparatus and stimuli were identical with Experiment 2, except that we introduced new nonverbal cues. Instead of photos displaying positive facial expressions or a “not face”, we presented color patches. For this purpose, a picture of a green rectangle (RGB color value: 0, 176, 80) and a picture of a red rectangle (RGB color value: 255, 0, 0) was created. We scaled both pictures to the same size as the photos used in the previous experiments. Furthermore, the same raters who had judged the photos evaluated on a 5-point Likert scale (1 = *definitely no*; 2 = *rather no*; 3 = *neutral*; 4 = *rather yes*; 5 = *definitely yes*) to what extent they associated the green and the red rectangle with “no” (negation) or “yes” (affirmation). The rating scores for the rectangles differed significantly, $t(11) = 4.43$, $p = .001$. The raters associated the red rectangle rather with “no” ($M = 2.25$, $SD = 0.97$) and the green rectangle rather with “yes” ($M = 3.92$, $SD = 0.67$).

Procedure

The procedure was the same as in Experiment 2, apart from the fact that we replaced the photos (“not face” vs. positive control) by the pictures of the colored rectangles (red vs. green).

Design and Data Analysis

The experiment implemented a 2 × 2 within-subjects design, with the factors sentence polarity (negated vs. affirmative) and color patch (red vs. green). The dependent variable was the same as in Experiment 2. All experimental data and analysis scripts are publicly available online (https://doi.org/10.5281/zenodo.7389652).

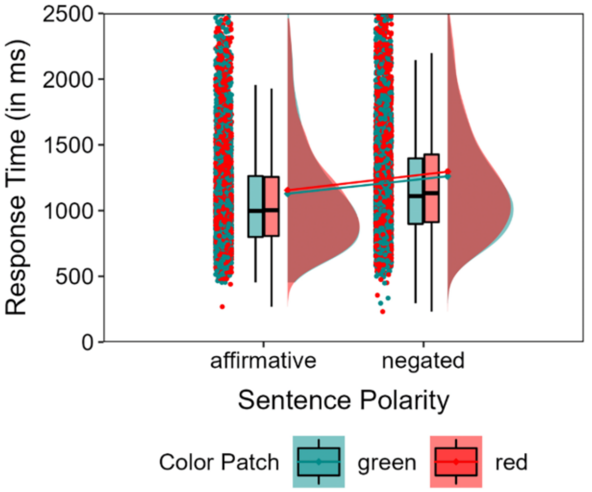
Data analysis followed the procedure used in Experiment 2. The exclusion of incorrect reactions from the response time analysis reduced the data set by less than 4%. Removing trials associated with anticipatory responses or disruption of performance reduced the data set by less than 1%. As in the prior experiments, several more complex models had to be disregarded when determining the random effects structure for the mixed models because we encountered singular fit or convergence issues. The models for both response time and response accuracy comprised random intercepts for participants and items.

Results

Figure 3 shows a raincloud plot (see Allen et al., 2021) for response times as a function of sentence polarity and color patch. Once again, there was a significant effect of sentence polarity, $\chi^2(1) = 241.56$, $p < .001$, which was confirmed in terms of the BF (>100; extreme evidence for the existence of the effect). Participants responded faster to affirmative sentences ($M = 1{,}141$ ms, $SD = 550$ ms) than to negated sentences ($M = 1{,}278$ ms, $SD = 587$ ms). The effect of color patch was also significant, $\chi^2(1) = 12.59$, $p < .001$.

Figure 3

Response Times as a Function of Sentence Polarity and Color Patch in Experiment 3



Note. The raincloud plot was created using R code provided by Allen et al. (2021). See the online article for the color version of this figure.

The associated BF (9.64) proposed that there was moderate evidence in favor of the existence of the effect. Response times were faster in trials with a green patch (*M* = 1,194 ms, *SD* = 546 ms) than in trials with a red patch (*M* = 1,225 ms, *SD* = 598 ms). Most importantly, we did not observe a significant interaction of sentence polarity and color patch, $\chi^2(1) = 0.15$, $p = .701$. The BF (0.03) strongly supported the absence of an interaction effect.

The analysis of response accuracy showed that there was a significant effect of sentence polarity, $\chi^2(1) = 21.58$, $p < .001$. Similarly, the BF (>100) provided extreme evidence for the existence of the effect. Participants committed less errors to affirmative sentences (*M* = 2.98%, *SD* = 4.02%) than to negated sentences (*M* = 4.58%, *SD* = 5.10%). Moreover, the effect of color patch was also significant, $\chi^2(1) = 5.50$, $p = .019$. However, the BF (0.15) suggested that there was moderate evidence for the absence of this effect. The mean error rate was lower in trials including a green color patch (*M* = 3.37%, *SD* = 4.19%) than in trials including a red color patch (*M* = 4.20%, *SD* = 4.76%). The interaction of sentence polarity and color patch was not significant, $\chi^2(1) = 0.06$, $p = .800$. The BF (<0.01) extremely favored the absence of the interaction.

General Discussion

Psycholinguistic research has investigated negation predominantly in terms of a purely verbal operator. However, there are various ways to express negation in a nonverbal manner, such as the head shake or the vertically oriented open palm (e.g., Kendon, 2002, 2004). Most interestingly for the current research, recent findings propose that producing negated contents appears to be associated with the so-called "not face", which includes facial expressions of disgust, anger, and contempt (Benitez-Quiroz et al., 2016). Moreover, it has been shown that pictorially presented red (green) color information is able to facilitate providing "no" ("yes") responses (Dudschig et al., 2023). In a series of three experiments, we examined the effect of perceiving the "not face" or red and green color patches—that is, visual information that has been linked to negation or affirmation—on the processing of verbal negation and affirmation. Participants were presented with photos showing a "not face" versus a positive facial expression (Experiments 1 and 2) or a red versus green color patch (Experiment 3). Subsequently, they read negated and affirmative sentences in a self-paced manner or judged the sensibility of negated and affirmative sentences. We expected to observe respective compatibility effects if the "not face" and red color information (green color information) should serve as a visual marker for negation (affirmation), as redundant information can accelerate information processing (see Miller, 1982 for a study from the nonlinguistic domain). This would indicate that visual cues are rapidly integrated in a one-step fashion during processing sentential negation and affirmation.

However, the results provided no evidence pointing toward compatibility effects induced by multimodal presentation of negation and affirmation; this was true for both the "not face" and red and green color information. Language processing (operationalized in terms of reading and sensibility judgment times for negated and affirmative sentences) was not significantly facilitated when nonverbal information ("not face"; red or green color) and sentence polarity matched. This evaluation is not solely based on traditional frequentist statistics but also on BF analyses. In addition, prior research convincingly demonstrated that language comprehenders generally possess the capability to integrate different types of nonverbal information during meaning composition, such as the voice of a speaker, event-based expectations, and aspects of world knowledge (e.g., Dudschig et al., 2016b; Hagoort et al., 2004; Matsuki et al., 2011; Van Berkum et al., 2008). This also applies to experiments involving visual presentations of nonverbal information separate from verbal stimuli like words or sentences (e.g., Dudschig et al., 2021; Lüdtke et al., 2008; Rück et al., 2017), thus implementing similar presentation modes as the present research. In sum, our results therefore appear to indicate that the "not face" and red and green color information might not affect the online processing of verbal negation or affirmation on the sentential level.

The present research was primarily interested in investigating the comprehension of multimodal negation and affirmation and related integration phenomena. Despite our findings, it is thus still conceivable that the "not face" is routinely produced during negated utterances (Benitez-Quiroz et al., 2016); however, it could rather constitute a mere by-product of language production than being a relevant nonverbal marker used by recipients to facilitate negation processing. Prior meta-analytic findings indeed suggested that gestural information might be less advantageous for comprehension when it solely duplicates information that is also conveyed in a verbal manner (Hostetter, 2011)—just as it occurs in the present experiments. Results from the same meta-analysis also revealed that adults tend to benefit less from multimodal information than children, which can be aligned with our findings as well. In addition, previously reported effects of red and green color might be limited to saying "no" and "yes" (Dudschig et al., 2023) and not extend to processing negation and affirmation in the context of larger linguistic units such as sentences. Apart from the obvious interpretation that the specific types of visual information investigated in the current research are not central to processing negation or affirmation on the sentential level, this could even indicate that relevant cognitive processes generally operate on abstract representations that are independent from any nonverbal (sensorimotor) information. Such an idea would clearly match with the traditional view of cognition (e.g., Chomsky, 1980; Fodor, 1983; Pylyshyn, 1984), which claims that higher-level cognitive processes such as language comprehension are based on abstract amodal symbols (for opposing points of view, see Barsalou, 1999; Bidet-Ildei et al., 2020; Glenberg & Kaschak, 2002; Zwaan & Madden, 2005). However, it must be taken into account that the set of primes was limited and did not include the full range of potential nonverbal markers of negation and affirmation. Hence, to draw solid conclusions with respect to the representational format, further studies are needed that examine whether our findings on negation and affirmation comprehension can be replicated for other types of nonverbal information, such as the head shake or the vertical open palm (e.g., Kendon, 2002, 2004).

Methodological aspects might also explain the apparent discrepancies between the findings of the present and prior research. As outlined, our experiments were dedicated to the question whether comprehending negated and affirmative sentences can be influenced by multimodal information. Standard sentence comprehension tasks including visual primes were used to investigate the effect of nonverbal information on comprehension speed. By contrast, Benitez-Quiroz et al. (2016) explicitly asked participants to show facial expressions of negation and analyzed facial expression of participants during the reproduction of negated and affirmative sentences from memory. The authors concluded that the "not face" could be an evolutionary antecedent of grammaticalized negation, thus possibly carrying semantic meaning. In line with this idea, participants of our rating tasks semantically associated the "not face" with negation—but this did not extend to online language comprehension processes. Furthermore, in the experiments conducted by Dudschig et al. (2023), participants were presented with a task exploring the influence of emoticons (frowning vs. smiling) and color cues (red vs. green) on processing a fairly specific type of affirmation and negation. They provided simple "yes" and "no" answers to categorize stimuli as words or nonwords. This task does not reflect a natural language comprehension scenario and probably uncovered cognitive processes that are related to the decision but not to the comprehension level. At least, it remains unresolved whether response time effects emerged from processing the labels "yes" and "no" shown on the response buttons (red-colored vs. green-colored; including a frowning vs. smiling emoticon) or the response decision stage involved in the task (e.g., Sternberg, 1969).

In all experiments presented here, participants responded significantly slower to negated sentences than to affirmative sentences, replicating typical processing difficulties prior research regularly reported with respect to negation (for an overview, see Kaup & Dudschig, 2020). This suggests that the sentence materials properly established the experimental conditions of affirmation and negation and that the task was suitable to detect them. However, one limitation associated with the setup could be the fact that nonverbal cues (i.e., the "not face" vs. positive controls; red vs. green color patches) always appeared separated from sentences (these were displayed below the nonverbal cues) and not in an integrated manner. This aspect clearly differs from natural settings, such as a dialogue of two persons, where potential nonverbal markers of negation and affirmation and verbal contents tend to be conveyed in a simultaneous manner. Nevertheless, several studies employing similar setups involving the visual presentation of nonverbal information separate from written words or sentences reported effects demonstrating the integration of nonverbal and verbal information. For instance, this applies to rather obvious nonverbal information on the gender identity of speakers (Rück et al., 2017), to more subtle nonverbal information on the surface of materials (Dudschig et al., 2021), but also to pictures of situations described in affirmative and negated sentences (Lüdtke et al., 2008). However, future research exploring the integration of multimodal information on negation and affirmation during sentence processing could clearly benefit from incorporating video-based materials. This would allow directly combining different sources of information, thereby simulating natural language processing scenarios. Furthermore, adopting time-locked presentation methods, where nonverbal negation markers are displayed precisely when verbal negation markers occur in spoken language (see Willems et al., 2008), might be a promising approach. To gain deeper insights into potential integration mechanisms that are specifically related to critical phrases of negated and affirmative sentences, one could implement techniques enabling a more direct examination of online comprehension processes (see Dudschig et al., 2016, 2018; Rück et al., 2017). This would be particularly relevant if effects of multimodal integration might be overwritten by other processing issues or decay until the end of a sentence.

Finally, it is conceivable that the "not face" and the color cues under investigation tend to interact solely with processing very specific types of negated and affirmative sentences, which were not sufficiently covered by the sentential materials used in the experiments (e.g., "No, I would not like to buy a sofa"; "Yes, I want to sing now"; "No, I would not like to have a pillow"). In particular, the "not face" combines facial expressions of anger, disgust, and contempt (Benitez-Quiroz et al., 2016). It is therefore possible that its influence is limited to understanding negated sentences related to these emotions (e.g., "No, do not lie at me!"; "No, I would not like to eat locusts!"; "No, I do not want to be compared to such persons!"). One might also speculate that the potential association of red color with negation and green color with affirmation stems from the fact that these colors typically serve as stop or go signal and are used to ask for action inhibition or action initiation in our everyday life (e.g., in terms of red and green traffic lights). Consequently, red and green color cues could affect especially the comprehension of sentences denying or approving concrete motor actions (e.g., "Do not touch the stove!"; "Open the door, please!"; "Push the button to start the engine!").

To conclude, our experiments did not reveal any significant influence of perceiving facial expressions ("not face" vs. positive controls) or color information (red vs. green) on processing negated and affirmative sentences. These findings appear to stand in contrast to prior research that focused on the expression of the "not face" during uttering negated content (Benitez-Quiroz et al., 2016) or the effect of green and red color cues on the speed of providing simple "yes" and "no" responses in a lexical decision task (Dudschig et al., 2023). However, the present research specifically aimed to explore whether particular visual primes can affect the comprehension of negated and affirmative sentences. Future studies simulating natural language comprehension scenarios, implementing methods allowing a more direct examination of online processes, and using specific sentential contents might be a promising avenue to clarify whether certain facial expressions or color information—but also other types of nonverbal information (e.g., the head shake)—can modulate the comprehension of sentential negation and affirmation at all. If not, this might suggest that cognitive processes associated with understanding sentential negation and affirmation rely on more abstract representational formats.

References

Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., & Kievit, R. A. (2021). Raincloud plots: A multi-platform tool for robust data visualization. *Wellcome Open Research*, *4*, Article 63. https://doi.org/10.12688/wellcomeopenres.15191.2

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*(4), 577–660. https://doi.org/10.1017/S0140525X99002149

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

Benitez-Quiroz, C. F., Wilbur, R. B., & Martinez, A. M. (2016). The not face: A grammaticalization of facial expressions of emotion. *Cognition*, *150*, 77–84. https://doi.org/10.1016/j.cognition.2016.02.004

Bidet-Ildei, C., Beauprez, S.-A., & Badets, A. (2020). A review of literature on the link between action observation and action language: Advancing a shared semantic theory. *New Ideas in Psychology*, *58*, Article 100777. https://doi.org/10.1016/j.newideapsych.2019.100777

Bressem, J., & Müller, C. (2014). A repertoire of German recurrent gestures with pragmatic functions. In C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, & J. Bressem (Eds.), *Body-language-communication: An international handbook on multimodality in human interaction* (Vol. 2, pp. 1575–1591). De Gruyter Mouton. https://doi.org/10.1515/9783110302028.1575

Cameron, D. L. (1991). *The use of signs for negation and affirmation by five cross-fostered chimpanzees* (Publication No. 1346764) [Master's thesis, University of Nevada]. ProQuest Dissertations and Theses Global.

Campbell, J. I. D., & Thompson, V. A. (2012). MorePower 6.0 for ANOVA with relational confidence intervals and Bayesian analysis. *Behavior Research Methods*, *44*(4), 1255–1265. https://doi.org/10.3758/s13428-012-0186-0

Chomsky, N. (1980). *Rules and representations*. Columbia University Press.

Dautriche, I., Buccola, B., Berthet, M., Fagot, J., & Chemla, E. (2022). Evidence for compositionality in baboons (*Papio papio*) through the test case of negation. *Scientific Reports*, *12*(1), Article 19181. https://doi.org/10.1038/s41598-022-21143-1

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, *47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

Dimroth, C. (2010). The acquisition of negation. In L. R. Horn (Ed.), *The expression of negation* (pp. 39–71). De Gruyter Mouton. https://doi.org/10.1515/9783110219302

Dudschig, C., Kaup, B., Leuthold, H., & Mackenzie, I. G. (2021). Conceptual representation of real-world surface material: Early integration with linguistic labels indicated in the N400 component. *Psychophysiology*, *58*(12), Article e13916. https://doi.org/10.1111/psyp.13916

Dudschig, C., Kaup, B., & Mackenzie, I. G. (2023). The grounding of logical operations: The role of color, shape, and emotional faces for "yes" or "no" decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *49*(3), 477–492. https://doi.org/10.1037/xlm0001181

Dudschig, C., Mackenzie, I. G., Leuthold, H., & Kaup, B. (2018). Environmental sound priming: Does negation modify N400 cross-modal priming effects? *Psychonomic Bulletin & Review*, *25*(4), 1441–1448. https://doi.org/10.3758/s13423-017-1388-3

Dudschig, C., Mackenzie, I. G., Maienborn, C., Kaup, B., & Leuthold, H. (2019). Negation and the N400: Investigating temporal aspects of negation integration using semantic and world-knowledge violations. *Language, Cognition and Neuroscience*, *34*(3), 309–319. https://doi.org/10.1080/23273798.2018.1535127

Dudschig, C., Mackenzie, I. G., Strozyk, J., Kaup, B., & Leuthold, H. (2016). The sounds of sentences: Differentiating the influence of physical sound, sound imagery, and linguistically implied sounds on physical sound processing. *Cognitive, Affective, & Behavioral Neuroscience*, *16*(5), 940–961. https://doi.org/10.3758/s13415-016-0444-1

Dudschig, C., Maienborn, C., & Kaup, B. (2016a). Is there a difference between stripy journeys and stripy ladybirds? The N400 response to semantic and world-knowledge violations during sentence processing. *Brain and Cognition*, *103*, 38–49. https://doi.org/10.1016/j.bandc.2016.01.001

Dudschig, C., Maienborn, C., & Kaup, B. (2016b). These lemons are sour: Investigating the influence of demonstrative determiners on the N400 complex. *Neuroscience Letters*, *630*, 141–146. https://doi.org/10.1016/j.neulet.2016.07.031

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

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. MIT Press. https://doi.org/10.7551/mitpress/4737.001.0001

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, *9*(3), 558–565. https://doi.org/10.3758/BF03196313

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, *304*(5669), 438–441. https://doi.org/10.1126/science.1095455

Harrison, S. (2018). *The impulse to gesture: Where language, minds, and bodies intersect*. Cambridge University Press. https://doi.org/10.1017/9781108265065

Hostetter, A. B. (2011). When do gestures communicate? A meta-analysis. *Psychological Bulletin*, *137*(2), 297–315. https://doi.org/10.1037/a0022128

Kaup, B., & Dudschig, C. (2020). Understanding negation: Issues in the processing of negation. In V. Déprez & M. T. Espinal (Eds.), *The Oxford handbook of negation* (pp. 635–655). Oxford University Press.

Kendon, A. (2002). Some uses of the head shake. *Gesture*, *2*(2), 147–182. https://doi.org/10.1075/gest.2.2.03ken

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge University Press.

Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge University Press. https://doi.org/10.1017/CBO9781139087759

Lenth, R. V. (2022). *emmeans: Estimated marginal means, aka least-squares means* (R package version 1.7.4-1) [Computer software]. https://doi.org/https://cran.r-project.org/package=emmeans

Litowitz, B. E. (1998). An expanded developmental line for negation: Rejection, refusal, denial. *Journal of the American Psychoanalytic Association*, *46*(1), 121–148. https://doi.org/10.1177/00030651980460010 8

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

Matsuki, K., Chow, T., Hare, M., Elman, J. L., Scheepers, C., & McRae, K. (2011). Event-based plausibility immediately influences on-line language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(4), 913–934. https://doi.org/10.1037/a0022964

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology*, *14*(2), 247–279. https://doi.org/10.1016/0010-0285(82)90010-X

Morey, R. D., & Rouder, J. N. (2021). *BayesFactor: Computation of Bayes factors for common designs* (R package version 0.9.12-4.3) [Computer software]. https://doi.org/https://cran.r-project.org/package=BayesFactor

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

Pea, R. D. (1980). The development of negation in early child language. In D. R. Olson (Ed.), *The social foundations of language and thought: Essays in honor of Jerome Seymour Bruner* (pp. 156–186). W.W. Norton.

Pylyshyn, Z. W. (1984). *Computation and cognition: Towards a foundation for cognitive science*. MIT Press.

Rück, F., de la Vega, I., Leuthold, H., & Kaup, B. (2017). Integration of visual information about the speaker during sentence processing. In S. Featherston, R. Hörnig, R. Steinberg, B. Umbreit, & J. Wallis (Eds.), *Proceedings of Linguistic Evidence 2016: Empirical, theoretical, and computational perspectives*. University of Tübingen. https://doi.org/10.15496/publikation-19060

Schneider, C., Call, J., & Liebal, K. (2010). Do bonobos say NO by shaking their head? *Primates*, *51*(3), 199–202. https://doi.org/10.1007/s10329-010-0198-2

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2021). *afex: Analysis of factorial experiments* (R package version

1.0-1) [Computer software]. https://doi.org/https://cran.r-project.org/package=afex

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, *30*, 276–315. https://doi.org/10.1016/0001-6918(69)90055-9

Van Berkum, J. J. A., van den Brink, D., Tesink, C. M. J. Y., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, *20*(4), 580–591. https://doi.org/10.1162/jocn.2008.20054

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review*, *14*(5), 779–804. https://doi.org/10.3758/BF03194105

Willems, R. M., Özyürek, A., & Hagoort, P. (2008). Seeing and hearing meaning: ERP and fMRI evidence of word versus picture integration into a sentence context. *Journal of Cognitive Neuroscience*, *20*(7), 1235–1249. https://doi.org/10.1162/jocn.2008.20085

Zeijlstra, H. (2007). Negation in natural language: On the form and meaning of negative elements. *Language and Linguistics Compass*, *1*(5), 498–518. https://doi.org/10.1111/j.1749-818x.2007.00027.x

Zuberbühler, K. (2015). Linguistic capacity of non-human animals. *Wiley Interdisciplinary Reviews: Cognitive Science*, *6*(3), 313–321. https://doi.org/10.1002/wcs.1338

Zwaan, R. A., & Madden, C. J. (2005). Embodied sentence comprehension. In D. Pecher & R. A. Zwaan (Eds.), *Grounding cognition: The role of perception and action in memory, language, and thinking* (pp. 224–245). Cambridge University Press. https://doi.org/10.1017/CBO9780511499968.010

Received January 23, 2023
Revision received August 8, 2023
Accepted September 17, 2023 ■