

# Bodily Effort Enhances Learning and Metacognition: Investigating the Relation Between Physical Effort and Cognition Using Dual-Process Models of Embodiment

Alexander Skulmowski and Günter Daniel Rey

E-Learning and New Media, Institute for Media Research, Faculty of Humanities,  
Technische Universität Chemnitz, Germany

## ABSTRACT

Recent embodiment research revealed that cognitive processes can be influenced by bodily cues. Some of these cues were found to elicit disparate effects on cognition. For instance, weight sensations can inhibit problem-solving performance, but were shown to increase judgments regarding recall probability (judgments of learning; JOLs) in memory tasks. We investigated the effects of physical effort on learning and metacognition by conducting two studies in which we varied whether a backpack was worn or not while 20 nouns were to be learned. Participants entered a JOL for each word and completed a recall test. Experiment 1 ( $N = 18$ ) revealed that exerting physical effort by wearing a backpack led to higher JOLs for easy nouns, without a notable effect on difficult nouns. Participants who wore a backpack reached higher recall scores. Therefore, physical effort may act as a form of desirable difficulty during learning. In Experiment 2 ( $N = 30$ ), the influence of physical effort on JOLs and learning disappeared when more difficult nouns were to be learned, implying that a high cognitive load may diminish bodily effects. These findings suggest that physical effort mainly influences superficial modes of thought and raise doubts concerning the explanatory power of metaphor-centered accounts of embodiment for higher-level cognition.

## KEYWORDS

embodied cognition,  
metacognition, memory,  
judgments of learning,  
cognitive load

## INTRODUCTION

How do bodily cues affect higher-level cognitive processing, such as learning and metacognition? Previous research has shown that bodily perception is able to influence cognitive processes in a multitude of ways (Barsalou, 1999, 2008; Glenberg, 2010). As an example, the bodily perception of weight cues has been found to increase judgments regarding the importance of information on paper sheets affixed to heavy clipboards (Ackerman, Nocera, & Bargh, 2010) as well as regarding the importance of the contents of heavy books (Chandler, Reinhard, & Schwartz, 2012). As Jostmann, Lakens, and Schubert (2009) argue, the physical effort that is required for carrying heavier objects may act as a bodily trigger to put more cognitive effort into a task. In two studies, we aimed to assess whether physical effort enhances learning performance and whether metacognitive judgments are positively affected.

Furthermore, we addressed whether theoretical frameworks, such as conceptual metaphor theory (e.g., Lakoff & Johnson, 1980), provide an adequate explanation for more complex embodiment effects on cognition.

As an example for research on the bodily cue of physical effort, people tend to judge hills to be steeper when they are wearing a heavy backpack compared with not wearing a backpack (Bhalla & Proffitt,

Corresponding author: Alexander Skulmowski, E-Learning and New Media, Institute for Media Research, Faculty of Humanities, Technische Universität Chemnitz, Straße der Nationen 12, 09111 Chemnitz, Germany. Phone: 0049 371 531-34748. Fax: 0049 371 531-834748. Email: alexander.skulmowski@phil.tu-chemnitz.de

1999). In instances in which perception is closely related to a physical action, physical demands, such as carrying a heavy backpack, are said to be able to let that physical action appear to be more effortful and therefore may have the potential to bias perceptual judgments (Witt, 2011). Current research suggests that a higher physical effort may in some cases affect judgments regarding cognitive effort. As an example, Jostmann et al. (2009) assume that holding heavy objects may lead to a higher degree of cognitive elaboration, thereby supposing a link between physical and cognitive effort. A recent meta-analytic review on the role of physical workload in the perception of time (Block, Hancock, & Zakay, 2016) supports the view that physical effort has similar effects on task perception as cognitive effort, as both types of workload demand attention and therefore constrain cognitive processing. Relatedly, research on the role of arousal on cognition has established two important findings: (a) There is evidence for the claim that learning performance is highest at an optimum point of medium arousal (Teigen, 1994; Yerkes & Dodson, 1908), and (b) different levels of arousal are considered to trigger different modes of information processing (Eysenck, 1976; Schwabe & Wolf, 2013). With the present study, we aimed to assess whether physical effort is able to induce such an optimum level of arousal during a word learning task and whether the difficulty of materials (and the associated cognitive load) interacts with these bodily cues.

Recent advances have been made in extending embodied cognition research (Barsalou, 1999, 2008) to more elaborate cognitive processes, such as problem-solving (e.g., Kaspar & Vennekötter, 2015), as well as learning and metacognition (e.g., Alban & Kelley, 2013). Kaspar and Vennekötter (2015) showed that the concurrent perception of weight can decrease performance in an anagram task, whereas studies by Alban and Kelley (2013) demonstrated that, if participants learn words printed on heavy boxes, their confidence in being able to recall these words increases in comparison with participants learning words printed on light boxes. Koriat and Nussinson (2009) found divergent effects on judgments of learning (JOLs, i.e., the predicted recall probability for learning targets; see also Koriat, 1997) when generating physical effort by letting participants contract their eyebrows. While this operationalization of physical effort ensued in lowered JOLs during a learning task without time limit, it led JOLs to increase if a time limit was imposed. The former result is explained in terms of a mechanism in which physical effort is interpreted as a higher requirement of study effort, which, in turn, is assumed to lead to a higher degree of perceived task difficulty. As a consequence, participants who contracted their eyebrows gave lower JOLs than the control group. Conversely, Koriat and Nussinson argue that under time pressure physical effort is perceived as “willful, goal-driven effort” (p. 1342) and thus leads to a higher predicted recall probability for learning targets, resulting in higher JOLs for the group exerting physical effort. However, as Koriat and Nussinson mention, facial expressions, such as contracting the eyebrows, may not only be linked to the feeling of effort but also to emotional states. Therefore, we interpreted the operationalization of physical effort in the studies by Koriat and Nussinson to be confounded and regarded further research on the role of physical effort on metacognition and learning as necessary. In sum, physical demands appear to impact the subjective

assessment and allocation of cognitive effort. Yet, a more comprehensive explanation for the variety of divergent effects of physical effort on higher-level cognition found in the literature appears to be missing.

The effects of embodiment manipulations are often explained in terms of conceptual metaphor theory (for an overview, see Lee & Schwarz, 2014), a model in which bodily cues are said to influence abstract cognitive processes by activating conceptual metaphors associated with those specific bodily experiences (Lakoff & Johnson, 1980, 1999). Consequently, judgments and decisions may be biased by the influence of the mentally activated metaphor so that, for instance, information displayed on a heavy object is judged to be more important than the same information displayed on a light object (e.g., Ackerman et al., 2010). One problematic aspect of this theory criticized in the literature is that it allows conflicting predictions regarding embodiment effects. Murphy (1996) discusses the problem of unrelated information contained within the conceptual space of the metaphorically used term being mapped to a target entity. When we apply this criticism to the problem at hand, the limitations of conceptual metaphor theory in explaining embodiment effects strikingly become apparent. In the case of weight manipulations, embodiment effects have been explained with reference to the metaphor of importance (Ackerman et al., 2010; Jostmann et al., 2009) or by referring to the metaphor of weight as a burden (Kouchaki, Gino, & Jami, 2014; Slepian, Masicampo, Toosi, & Ambady, 2012), allowing the prediction of both positive and negative effects on cognition originating from the perception of one and the same stimulus (see Kaufmann & Allen, 2014, for a related discussion of the variable meaning of weight cues).

The metaphorical link of weight to the concept of importance suggests rising JOLs and enhanced learning performance (as discussed by Alban & Kelley, 2013). However, considering weight as associated with the concept of burden may at the same time imply a negative effect on task performance, as some studies suggest (Kaspar & Vennekötter, 2015; Slepian et al., 2012), namely lower JOLs and decreased learning performance. This is a result of the range of information linked to a concept being too diverse, enabling unrelated information to be mapped to a target (Murphy, 1996). Consequently, conceptual metaphor theory appears to be an unsuitable theoretical framework for our particular experimental setting in which weight is used as a means to generate physical effort.

Even more troubling than these theoretical issues are recent findings questioning the validity of studies conducted based on the idea of weight metaphors. A variety of weight effects on judgments could not be replicated, and the issue of publication bias has been raised consequently (Rabelo, Keller, Pilati, & Wicherts, 2015). In addition, studies involving backpack manipulations of physical effort aimed at biasing judgments have become the subject of critique as well, for problems with replicability and for theoretical considerations (Durgin, Klein, Spiegel, Strawser, & Williams, 2012; Firestone, 2013). These controversies highlight that great caution must be taken when designing and interpreting studies dealing with embodiment. Some of these problems could be avoided by formulating hypotheses based on more basic theoretical entities, such as arousal, rather than metaphoric relationships.

Lee and Schwarz (2014) offer an overview of processes involved in embodiment manipulations that may help to better understand the effects of physical effort on cognition. They distinguish between embodiment effects that involve metaphors and those that do not. On the one hand, they assume the existence of embodiment effects not mediated by metaphors including bodily influences on cognition related to the general arousal level or facial expressions. On the other hand, they argue that some embodiment effects ensuing from metaphorically laden bodily cues are indeed able to influence people's behavior and decision-making as a result of being reminded of certain mental states in a manner that bears a relation to that metaphor. We designed our experiment in a way that avoids a link between the bodily sensation and metaphors. Instead of using a weight manipulation physically attached to the learning target, such as clipboards or printed objects, we used a heavy backpack to induce physical effort (following Bhalla & Proffitt, 1999). Applying weight directly to the body instead of varying the weight of objects on which the learning targets are imprinted spatially disconnects the weight sensation from the learning materials, making a metaphor-based explanation implausible (see Kaufmann & Allen, 2014, for a discussion of a similar use of a backpack manipulation). Rather, we assume that bodily cues, such as physical effort, may be utilized as an indicator for cognitive effort required by a task, causing more attention to be allocated towards a physically more demanding task (as proposed by Jostmann et al., 2009). A related argument was put forth by Alban and Kelley (2013), who claim that strong differences in weight may be able to direct attention towards stimuli, thereby potentially enhancing learning performance.

Choi, Van Merriënboer, and Paas (2014) stress the importance of the physical surroundings in learning tasks. Among other effects that the environment can have on learners, they cite R. A. Bjork's (1994) work on "desirable difficulties." Desirable difficulties are defined as properties of learning tasks that promote learning through generating difficulty, such as the use of learning strategies making learning less predictable (E. L. Bjork & Bjork, 2011). An example for learning task factors introducing difficulties that might initially appear harmful to learning but has been found to actually increase learning performance is the use of learning materials presented in low visual quality (Eitel, Köhl, Scheiter, & Gerjets, 2014). Similarly, exerting physical effort during learning may prove to be a desirable difficulty, promoting a stronger focus on the learning contents.

To more closely assess the interplay between the relevant cognitive pathways for embodied information processing, we also aim to investigate the limits of bodily influences in higher-level cognitive processes, such as learning and metacognition. Maglio and Trope (2012) have provided evidence for the claim that embodiment manipulations primarily affect judgments if a low-level thought mode has been triggered, suggesting that more deliberate cognitive processes are not as easily influenced by bodily cues. In one of their studies, participants were asked to estimate the length of a hallway while either wearing a heavy backpack or not wearing a backpack. As a second experimental factor, Maglio and Trope induced high- and low-level modes of thinking by giving their participants different writing tasks before the estimation

task. Only those participants whose superficial thinking mode had been activated were susceptible to the embodiment manipulation. Furthermore, a higher overall cognitive load diminished effects on judgment processes in experiments by Jiang and Hong (2014). In their study, processing fluency influenced judgments, but only when the overall cognitive load required for the task was low. Consequently, we assumed that cognitive load might have a similar inhibitory influence on embodiment effects. In the research area of metacognition, Alter, Oppenheimer, Epley, and Eyre (2007) proposed a dual-process model of metacognition stating that difficulty acts as a trigger for more thorough modes of thought. Considering the results of Maglio and Trope, this model provides further evidence for the claim that embodiment effects on metacognition may be limited to superficial modes of information processing. In accordance with these findings, the influence of weight cues may be restricted to superficial cognitive processing routes, hence only affecting the JOLs of easy words in a learning task. In addition, the presented findings suggested that cognitive overload may negate bodily effects on JOLs.

We conducted two experiments in which participants learned words of varying difficulty levels and provided JOLs while either exerting physical effort or not. Based on the reviewed literature, we hypothesized that people wearing a backpack would report higher JOLs but only during superficial processing triggered by easy words. Similarly, we predicted higher learning outcomes for participants wearing a backpack, based on the assumption that the physical effort required by carrying the backpack would lead to an increase in perceived task demands, guiding attention towards the learning targets. An increased overall cognitive load induced by highly difficult words should nullify these effects.

## EXPERIMENT 1

### Method

#### PARTICIPANTS

Since previous studies using comparable backpack manipulations yielded large effect sizes (e.g., Proffitt, Stefanucci, Banton, & Epstein, 2003), we calculated our sample size based on an estimated effect of  $\eta_p^2 = .12$ , with a power of .80, resulting in a required sample size of 18 participants. We tested 18 participants (14 female, 4 male;  $M_{\text{age}} = 25.6$ ,  $SD_{\text{age}} = 2.69$ ) who were German native speakers studying at or having graduated from a German university. They participated for course credit or monetary reward. Participants were randomly assigned to a condition and were distributed equally between the two experimental conditions. An age limit of 30 years kept both groups comparable.

#### MATERIALS

The learning materials consisted of 21 nouns (10 easy, 10 difficult, and one practice item of medium difficulty) that were pre-tested and compiled using the following procedure: Forty German native speakers rated 200 nouns in randomized order regarding their memorization

difficulty, concreteness, and emotional valence on 5-point scales in an online survey. As memorization difficulty ratings showed considerable overlap between words with one to two syllables and again for words with three to four syllables, we decided to form two difficulty groups. From the list, ten easy words (one to two syllables, mean difficulty rating = 1.39,  $SD = 0.22$ ) and ten difficult words (three to four syllables, mean difficulty rating = 3.18,  $SD = 0.23$ ) were selected as learning targets. In addition, one word of intermediate difficulty was chosen as a practice item. All words were matched for medium concreteness and medium emotional valence. The learning task was conducted using a tablet computer, with participants entering JOLs using slider controls from 0% to 100% in a browser-based survey.

## DESIGN AND PROCEDURE

At the beginning of each session, participants provided informed consent and gave information regarding their body weight. Participants were randomly assigned either to wear a backpack or no backpack (between-subjects) during the learning phase. In the backpack condition, based on participants' reported body weight, an experimenter filled the backpack to weigh 15% of the participants' weight (rounded to the nearest 0.5 kg) (a comparable procedure has been used in Experiment 1 of Bhalla & Proffitt, 1999). At this point, participants assigned to the effort condition put on the backpack. Next, the experimenter handed them the tablet and asked them to complete the learning task while standing. Participants held the tablet with their non-dominant hand and interacted with the experimental software using their dominant hand. On a welcome screen, participants read the instructions for the learning stage, describing that they would now learn 21 words consecutively and should enter a JOL for each word. They were informed that each word would only be displayed for 5 s, while there would be no time limit for entering their JOL. Since JOLs had to be entered on the tablet screen using a slider control, participants were now asked to familiarize themselves with this type of control by dragging a practice slider to a position of their choice. The instructions re-appeared and participants entered the learning stage. Participants started this stage with a trial featuring the practice word of medium difficulty before the 20 easy or difficult learning targets were displayed in randomized order. Since all participants learned all 20 learning targets, the factor difficulty was a within-subject factor. Each word appeared on the tablet screen for 5 s. Next, participants entered a corresponding JOL by using a slider control (0% to 100% probability of recall) and repeated this procedure with each of the remaining words. After the learning phase, participants were asked to sit down at a desk and were given 3 min to write down all words they could recall, followed by filling out a German translation of the NASA Task Load Index (NASA-TLX, Hart & Staveland, 1988; translation found in Pfendler, 1991), administered using 20-point scales as well as a demographic survey containing, among others, questions regarding their computer experience and tablet usage. We used the NASA-TLX as a manipulation check regarding mental and physical demands (analyzed without the weighting procedure as well as paired comparisons, see Hart, 2006). Each experimental session took 20 to 30 min.

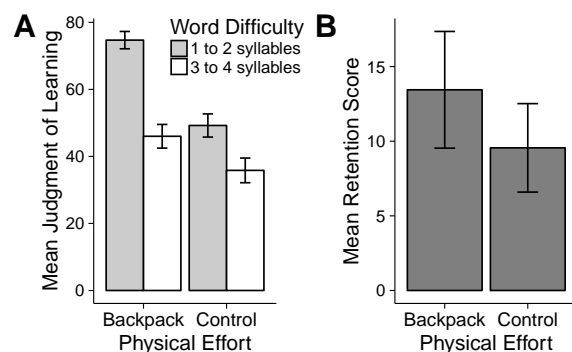
## SCORING

Only words that were recalled without any spelling mistakes were scored as correct and were awarded with one point. The practice item was not included in the learning performance score or any other analysis. Thus, a maximum of 20 points could be earned.

## Results

All analyses of variance (ANOVAs) were computed using aligned rank transformation (Salter & Fawcett, 1985) in order to perform nonparametric analyses. A  $2 \times 2$  mixed repeated-measures ANOVA (Physical Effort  $\times$  Word Difficulty) revealed a significant interaction between physical effort and word difficulty with respect to JOLs,  $F(1, 16) = 5.74, p = .029, d = 1.2$ , with physical effort increasing the JOLs for easy words more than for difficult ones (see Figure 1A), supporting our hypothesis that superficial processing triggered by easy words is more strongly affected by physical effort. In addition, there was a main effect of physical effort, with higher JOLs being entered by participants wearing a backpack,  $F(1, 16) = 10.71, p = .005, d = 1.64$ . Also, the more difficult words received significantly lower JOLs,  $F(1, 16) = 33.79, p < .001, d = 2.91$ .

Regarding recall performance, a  $2 \times 2$  mixed ANOVA (Physical Effort  $\times$  Word Difficulty) resulted in a significant main effect of physical effort,  $F(1, 16) = 6.23, p = .024, d = 1.25$  (see Figure 1B). As expected, participants wearing backpacks were able to remember more words correctly than those not wearing a heavy backpack. In addition, we found a significant main effect of word difficulty,  $F(1, 16) = 13.33, p = .002, d = 1.83$ , with lower recall performance for more difficult words ( $M = 5.06, SD = 2.55$ ) compared with easy words ( $M = 6.44, SD = 1.65$ ). The NASA-TLX results demonstrated that wearing a heavy backpack imposed a higher physical demand ( $M = 7.44, SD = 5.08$ ) compared with not wearing a backpack ( $M = 1.67, SD = 0.87$ ), using a Wilcoxon-Pratt Signed-Rank Test,  $Z = 2.52, p = .016$  (all other NASA-TLX  $ps \geq .273$ ).



**FIGURE 1.**

Mean judgments of learning (JOLs) for Experiment 1 (A) and mean retention score for Experiment 1 (B; maximum score of 20). JOL error bars depict the standard error of the mean and were normalized using the procedure described by Morey (2008). Retention error bars depict the standard deviation.

## EXPERIMENT 2

We performed a second experiment to assess the effects of physical effort and cognitive overload on metacognition and learning. Research on the Yerkes-Dodson law suggests that an optimal balance between physical and cognitive demands must be found in order to enhance cognitive performance (Teigen, 1994; Yerkes & Dodson, 1908). Based on results showing that a high cognitive load can counteract experimental influences on judgments (Jiang & Hong, 2014), we hypothesized that the effects of physical effort on metacognition and learning would decrease with higher cognitive demands.

### Method

#### PARTICIPANTS

As we expected that a higher overall cognitive load would weaken the effects of physical effort, we decided to lower our estimated effect size to  $\eta_p^2 = .07$ , with a power of .80, and accordingly planned to test 30 participants. Due to technical difficulties, the data from one participant had to be discarded and were replaced with data from an additional participant. We analyzed the data from 30 German native speakers (15 female, 15 male;  $M_{age} = 24.74$ ,  $SD_{age} = 3.3$ ) studying at or having graduated from a German university. They participated in the experiment for course credit or monetary reward. An age limit of 32 years kept both groups comparable. Condition assignment was randomized; 16 participants were assigned to wear a backpack.

#### MATERIALS

The same materials as in Experiment 1 were used, except for the word list. To test whether the effects found in Experiment 1 generalized to learning situations with more difficult words and hence a higher overall cognitive load, we compiled a new word list featuring nouns of higher difficulty than the previously used words. For the new word

list, the ten easy nouns from the previous experiment were used. We replaced the ten difficult words from the previous experiment with ten more difficult nouns. These words were compiled by pre-testing 45 nouns (five to six syllables) together with five words from the previous list that scored highest or lowest on one of the scales concreteness or emotional valence or lowest in perceived memorization difficulty. The additional words were added to provide context for the ratings. Nineteen German native speakers were tested online using the same procedure as described in the Materials section of Experiment 1.

#### DESIGN AND PROCEDURE

The experimental design and procedure were identical to Experiment 1.

#### SCORING

The scoring method was identical to Experiment 1.

### Results

As for Experiment 1, all non-Bayesian ANOVAs were computed using aligned rank transformation (Salter & Fawcett, 1985). Using a  $2 \times 2$  mixed repeated measures ANOVA (Physical Effort  $\times$  Word Difficulty), we found neither a main effect of physical effort on JOLs nor an interaction (both  $ps \geq .621$ ). However, there was a significant main effect of word difficulty,  $F(1, 28) = 135.01$ ,  $p < .001$ ,  $d = 4.4$ , with lower JOLs being entered for the more difficult words compared with the easy words (see Table 1).

As for recall performance, a  $2 \times 2$  mixed ANOVA (Physical Effort  $\times$  Word Difficulty) showed no significant effects (all  $ps \geq .258$ ), except for a main effect of word difficulty,  $F(1, 28) = 88.66$ ,  $p < .001$ ,  $d = 3.57$ , with lower recall performance for the more difficult words compared with the easy words (see Table 2). Again, the NASA-TLX indicated a higher physical demand when wearing the backpack ( $M = 6$ ,  $SD = 4.58$ ) compared with not wearing a backpack ( $M = 2.57$ ,  $SD = 1.74$ ),

**TABLE 1.**  
Mean Judgments of Learning (JOLs) in Experiment 2

	Group							
	BP easy		BP difficult		Control easy		Control difficult	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
JOLs	66.6	2.26	37.86	2.63	68.28	2.44	37.65	2.87

Note. BP = backpack group; easy = one to two syllable words, difficult = five to six syllable words. Standard errors were normalized using the procedure described by Morey (2008).

**TABLE 2.**  
Mean recall performance in Experiment 2

	Group							
	BP easy		BP difficult		Control easy		Control difficult	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Retention	6.88	1.51	3.31	1.51	6.29	1.32	2.86	1.32

Note. BP = backpack group; easy = one to two syllable words, difficult = five to six syllable words. Standard deviations were normalized using the procedure described by Morey (2008).

using a Wilcoxon rank sum test,  $W = 170.5$ ,  $p = .015$  (all other NASA-TLX  $ps \geq .173$ ).

To provide conclusive evidence for the claim that the cognitive overload induced by the high word difficulty nullified the effects of physical effort on JOLs and retention performance, we conducted Bayesian ANOVAs using the same models specified in the previous analyses. For the influence of physical effort on JOLs, we found a  $BF_{10}$  of 0.18 (error =  $\pm 3.23\%$ ), indicating that the null hypothesis is 5.6 times more likely than the alternative hypothesis. Regarding retention, the factor physical effort revealed a  $BF_{10}$  of 0.38 (error =  $\pm 1.76\%$ ), demonstrating that the null hypothesis is 2.63 times more likely than the alternative hypothesis. According to the categories presented in Table 1 of Wetzels and Wagenmakers (2012; based on Jeffreys, 1961), these results indicate “substantial evidence” and “anecdotal evidence” for the null hypothesis regarding JOLs and retention, respectively.

## GENERAL DISCUSSION

Physical effort induced by carrying weight strongly influences learning and metacognition, but this effect is limited to superficial information processing and low overall cognitive load. While previous research has often attempted to explain a wide variety of effects of bodily cues on cognition as metaphor-related influences on cognitive processes, such as weight being perceived as a cue for importance (e.g., Jostmann et al., 2009), we do not consider such models as suitable to account for our results.

The results of Experiment 1 suggest that physical effort enhances learning for both easy and more difficult words and affects JOLs dependent upon word difficulty. While there was a large increase in predicted recall probability for easy words for those participants who wore a heavy backpack, little difference ensued for difficult words. The result of differential effects of physical effort on metacognition depending upon word difficulty can be taken as evidence for a dual-process model of metacognition in which bodily cues primarily affect superficial thought processes triggered by easy learning contents, whereas processing routes activated through higher task difficulty remain unaffected. A similar embodiment model with a focus on judgment processes was put forth by Maglio and Trope (2012). A dual-process model of metacognition proposed by Alter et al. (2007), in which difficulty triggers more thorough modes of thought, offers an explanation for the weak effect of bodily cues on JOLs for the more difficult words in Experiment 1.

In contrast to Alban and Kelley (2013), who found no significant effects of weight on learning results, we discovered a large effect of physical effort induced by carrying weight on retention performance. As previously suggested (Alban & Kelley, 2013; Jostmann et al., 2009), strong weight differences may direct attention towards stimuli and consequently enhance learning performance. The JOL increases for easy words that we found as a result of the physical effort manipulation in Experiment 1 are in line with the results of Koriat and Nussinson (2009; Experiment 2) and provide further evidence for the claim that physical demands play a key role in determining how cognitive re-

sources are distributed (Jostmann et al., 2009). Experiment 2 expands the findings regarding the limits of embodiment effects by demonstrating that a high overall cognitive load induced by highly difficult words diminishes the effects of physical effort on learning and metacognition. The absence of an embodiment effect due to cognitive overload further supports the models of Maglio and Trope (2012) and Alter et al. (2007). Due to the large effects on metacognition found in Experiment 1, we have strong grounds to accept the null hypothesis for Experiment 2. Our Bayesian analyses provided further support for this interpretation.

Crucially, the differential effects of physical effort on metacognition depending upon word difficulty cannot be adequately explained by metaphor accounts of embodied cognition, such as conceptual metaphor theory. Furthermore, such models cannot account for the results of Experiment 2, as they provide no reason for why a high overall cognitive load should have nullified embodiment effects. As some researchers have attempted to explain more complex embodiment effects using extended models of conceptual metaphor theory that require certain preconditions to be met in order for an embodiment effect to ensue (e.g., Chandler et al., 2012), one might hold the view that conceptual metaphor theory can be combined with a dual-process model of embodiment based on Maglio and Trope (2012) described above. However, we argue that the embodiment manipulation in our studies may be more properly categorized as being unrelated to metaphors within the taxonomy put forth by Lee and Schwarz (2014). Similarly, Kaufmann and Allen (2014) described a study design involving a backpack manipulation as a weight cue bearing no relation to a social judgment task, supporting the assumption that the induction of physical effort using a heavy backpack should not be regarded as invoking metaphorically related mental contents (see Kaspar & Vennekötter, 2015, for a related critique of metaphor-based embodiment explanations).

Further research is needed to conclusively determine the underlying mechanisms that have led to our results. Based on our experiments, we argue that the exertion of physical effort resulted in a stronger focus of attentional resources that enhanced learning performance (a similar mechanism has been proposed by Alban & Kelley, 2013). This interpretation is compatible with the idea that physical effort during learning may act as a type of desirable difficulty (see E. L. Bjork & Bjork, 2011). Furthermore, our results are in line with perspectives dealing with the effects of arousal on learning. Higher physical effort and a (presumed) higher arousal appear to lead to an increased performance for easy stimuli, lending further support to the hypothesis that there is an optimum point of arousal for specific levels of mental effort (Teigen, 1994; Yerkes & Dodson, 1908). Follow-up studies should more closely investigate how attention allocation during learning is affected by physical effort.

There are some limitations that need to be noted in our study. As in all studies for which university students and graduates are recruited as participants, certain restrictions regarding generalizability apply. In our experiments, students may have had more experience with memorizing words by means of rote learning than other people. As a result,

their average learning performance may have been higher, which may have been a necessary condition for the results found in Experiment 1. Further research with different samples should be performed to assess the generalizability of our findings. At the same time, the effects of physical effort on complex cognitive processes could be investigated using a variety of cognitive tasks that possibly are even more susceptible to the influence of effort and arousal. An example would be learning or problem-solving tasks that feature an emotional component. Another important aspect to consider is that our results do not imply that every cognitive task benefits from exerting physical effort. Rather, our results demonstrated that physical effort may under certain circumstances support cognitive processes, presumably by increasing physiological arousal. Moreover, being asked to carry a heavy backpack during a word learning task may have appeared to participants as a somewhat unnatural situation. Backpack manipulations have been considered to be problematic in some instances because of obvious demand characteristics potentially biasing the responses of participants (see Durgin et al., 2012, for a discussion). However, we assume that demand characteristics cannot be the cause of the increases in recall performance that we have found in Experiment 1. Yet, different methods of inducing physical effort should be employed to assess the external validity of our results. Although the NASA-TLX questionnaire contains a question item measuring mental demand, our experimental manipulation did not lead to a significant difference concerning this variable, casting some doubt on whether participants who wore a backpack subjectively experienced more cognitive demand as a result of higher physical demands. Furthermore, our findings are limited by our decision to form distinct levels of difficulty instead of using words of continuously varying difficulty. It needs to be noted that using a greater number of words than we used in our experiments could possibly lead to different results due to potential fatigue resulting from carrying a heavy backpack for a longer time.

In sum, our results suggested that learning and metacognition can be enhanced by physical effort depending on which processing pathway is triggered by learning targets. At the same time, these effects appeared to be limited to learning tasks involving only low overall cognitive load. More generally, our results highlighted the need for continued work on theoretical accounts of embodied cognition. While conceptual metaphor theory is severely limited in explaining our results, dual-process models offer a more adequate alternative. Still, the underlying processes and their relations need to be further explored in order to develop a more comprehensive account of embodiment.

#### AUTHOR NOTE

This work was supported by the German Research Foundation (DFG) within the Research Training Group 1780/1. The authors thank Yannik Augustin for his help with data collection and scoring.

#### REFERENCES

- Ackerman, J. M., Nocera, C. C., & Bargh, J. A. (2010, June). Incidental haptic sensations influence social judgments. *Science*, 328(5986), 1712–1715. doi: 10.1126/science.1189993
- Alban, M. W., & Kelley, C. M. (2013). Embodiment meets metamemory: Weight as a cue for metacognitive judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39, 1628–1634. doi: 10.1037/a0032420
- Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General*, 136, 569–576. doi: 10.1037/0096-3445.136.4.569
- Barsalou, L. W. (1999). Perceptions of perceptual symbols. *Behavioral and Brain Sciences*, 22, 637–660. doi: 10.1017/S0140525X99532147
- Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, 59, 617–645. doi: 10.1146/annurev.psych.59.103006.093639
- Bhalla, M., & Proffitt, D. R. (1999). Visual–motor recalibration in geographical slant perception. *Journal of Experimental Psychology: Human Perception and Performance*, 25, 1076–1096. doi: 10.1037/0096-1523.25.4.1076
- Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). Duffield, UK: Worth Publishing.
- Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: MIT Press.
- Block, R. A., Hancock, P. A., & Zakay, D. (2016). Physical load affects duration judgments: A meta-analytic review. *Acta Psychologica*, 165, 43–47. doi: 10.1016/j.actpsy.2016.01.002
- Chandler, J. J., Reinhard, D., & Schwarz, N. (2012). To judge a book by its weight you need to know its content: Knowledge moderates the use of embodied cues. *Journal of Experimental Social Psychology*, 48, 948–952. doi: 10.1016/j.jesp.2012.03.003
- Choi, H. H., Van Merriënboer, J. J., & Paas, F. (2014). Effects of the physical environment on cognitive load and learning: Towards a new model of cognitive load. *Educational Psychology Review*, 26, 225–244. doi: 10.1007/s10648-014-9262-6
- Durgin, F. H., Klein, B., Spiegel, A., Strawser, C. J., & Williams, M. (2012). The social psychology of perception experiments: Hills, backpacks, glucose, and the problem of generalizability. *Journal of Experimental Psychology: Human Perception and Performance*, 38, 1582–1595. doi: 10.1037/a0027805
- Eitel, A., Kühn, T., Scheiter, K., & Gerjets, P. (2014). Disfluency Meets Cognitive Load in Multimedia Learning: Does Harder-to-Read Mean Better-to-Understand? *Applied Cognitive Psychology*, 28, 488–501. doi: 10.1002/acp.3004
- Eysenck, M. W. (1976). Arousal, learning, and memory. *Psychological Bulletin*, 83, 389–404. doi: 10.1037/0033-2909.83.3.389

- Firestone, C. (2013). How "paternalistic" is spatial perception? Why wearing a heavy backpack doesn't—and couldn't—make hills look steeper. *Perspectives on Psychological Science*, 8, 455–473. doi: 10.1177/1745691613489835
- Glenberg, A. M. (2010). Embodiment as a unifying perspective for psychology. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1, 586–596. doi: 10.1002/wcs.55
- Hart, S. G. (2006). Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting October 2006*, 50, 904–908. doi: 10.1177/154193120605000909
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology*, 52, 139–183. doi: 10.1016/S0166-4115(08)62386-9
- Jeffreys, H. (1961). *Theory of probability* (3rd ed.). New York: Oxford University Press.
- Jiang, Y., & Hong, J. (2014). It feels fluent, but not right: The interactive effect of expected and experienced processing fluency on evaluative judgment. *Journal of Experimental Social Psychology*, 54, 147–152. doi: 10.1016/j.jesp.2014.05.004
- Jostmann, N. B., Lakens, D., & Schubert, T. W. (2009). Weight as an embodiment of importance. *Psychological Science*, 20, 1169–1174. doi: 10.1111/j.1467-9280.2009.02426.x
- Kaspar, K., & Vennekötter, A. (2015). Embodied information in cognitive tasks: Haptic weight sensations affect task performance and processing style. *Advances in Cognitive Psychology*, 11, 64–76. doi: 10.5709/acp-0172-0
- Kaufmann, L. M., & Allen, S. (2014). Adding weight to judgments: The role of stimulus focality on weight-related embodied cognition. *Sensoria: A Journal of Mind, Brain & Culture*, 10, 41–48. doi: 10.7790/sa.v10i1.390
- Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, 126, 349–370. doi: 10.1037/0096-3445.126.4.349
- Koriat, A., & Nussinson, R. (2009). Attributing study effort to data-driven and goal-driven effects: Implications for metacognitive judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35, 1338–1343. doi: 10.1037/a0016374
- Kouchaki, M., Gino, F., & Jami, A. (2014). The burden of guilt: Heavy backpacks, light snacks, and enhanced morality. *Journal of Experimental Psychology: General*, 143, 414–424. doi: 10.1037/a0031769
- Lakoff, G., & Johnson, M. (1980). The metaphorical structure of the human conceptual system. *Cognitive Science*, 4, 195–208. doi: 10.1016/S0364-0213(80)80017-6
- Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to western thought*. New York, NY: Basic Books.
- Lee, S. W. S., & Schwarz, N. (2014). Metaphors in judgment and decision making. In M. J. Landau, M. D. Robinson, & B. P. Meier (Eds.), *The power of metaphor: Examining its influence on social life* (pp. 85–108). Washington, DC: APA.
- Maglio, S. J., & Trope, Y. (2012). Disembodiment: abstract construal attenuates the influence of contextual bodily state in judgment. *Journal of Experimental Psychology: General*, 141, 211–216. doi: 10.1037/a0024520
- Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, 4, 61–64. doi: 10.20982/tqmp.04.2.p061
- Murphy, G. L. (1996). On metaphoric representation. *Cognition*, 60, 173–204. doi: 10.1016/0010-0277(96)00711-1
- Pfendler, C. (1991). *Vergleichende Bewertung der NASA-TLX Skala und der ZEIS-Skala bei der Erfassung von Lernprozessen* (Bericht Nr. 92) [Comparative assessment of the NASA-TLX scale and the ZEIS scale during the recording of learning processes (Report No. 92)]. Wachtberg, Germany: Forschungsinstitut für Anthropotechnik.
- Proffitt, D. R., Stefanucci, J., Banton, T., & Epstein, W. (2003). The role of effort in perceiving distance. *Psychological Science*, 14, 106–112. doi: 10.1111/1467-9280.t01-1-01427
- Rabelo, A. L., Keller, V. N., Pilati, R., & Wicherts, J. M. (2015). No effect of weight on judgments of importance in the moral domain and evidence of publication bias from a meta-analysis. *PLoS ONE*, 10, e0134808. doi: 10.1371/journal.pone.0134808
- Salter, K. C., & Fawcett, R. F. (1985). A robust and powerful rank test of treatment effects in balanced incomplete block designs. *Communications in Statistics-Simulation and Computation*, 14, 807–828. doi: 10.1080/03610918508812475
- Schwabe, L., & Wolf, O. T. (2013). Stress and multiple memory systems: From 'thinking' to 'doing'. *Trends in Cognitive Sciences*, 17, 60–68. doi: 10.1080/03610918508812475
- Slepian, M. L., Masicampo, E. J., Toosi, N. R., & Ambady, N. (2012). The physical burdens of secrecy. *Journal of Experimental Psychology: General*, 141, 619–624. doi: 10.1037/a0027598
- Teigen, K. H. (1994). Yerkes-Dodson: A law for all seasons. *Theory & Psychology*, 4, 525–547. doi: 10.1177/0959354394044004
- Wetzels, R., & Wagenmakers, E. J. (2012). A default Bayesian hypothesis test for correlations and partial correlations. *Psychonomic Bulletin & Review*, 19, 1057–1064. doi: 10.3758/s13423-012-0295-x
- Witt, J. K. (2011). Action's effect on perception. *Current Directions in Psychological Science*, 20, 201–206. doi: 10.1177/096372141140877
- Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, 18, 459–482. doi: 10.1002/cne.920180503

RECEIVED 25.06.2016 | ACCEPTED 21.10.2016