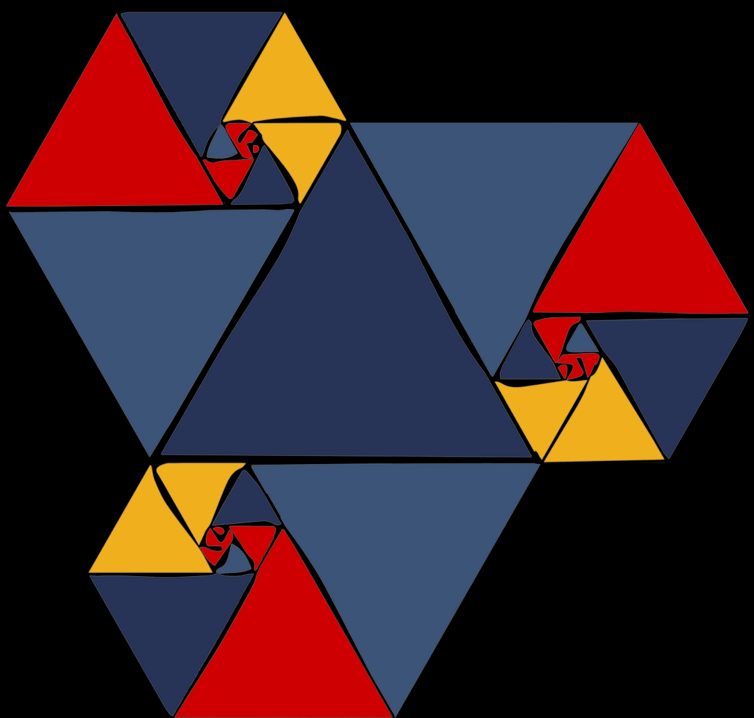
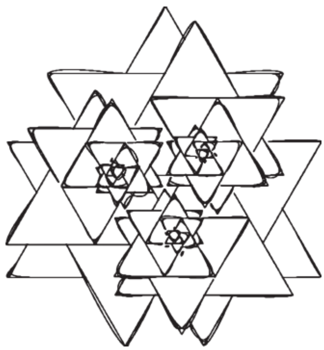
The Cognitive Function of Manual Activity in Learning and Problem Solving

WIM POUW



The Cognitive Function of Manual Activity in Learning and Problem Solving



Wim-Paul Theodorus Josephus Lodewyck Pouw

Cover design and art by M. Ruiter

The research presented on this dissertation was funded by Netherlands Organisation for Scientific Research (NWO project number: 411-10-908).

The Cognitive Function of Manual Activity in Learning and Problem Solving

De rol van handelend denken in leren en probleemoplossen

Proefschrift

ter verkrijging van de graad van doctor aan de

Erasmus Universiteit Rotterdam

op gezag van de

rector magnificus

prof.dr. H.A.P. Pols

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

Donderdag 16 Maart 2017 om 15:30 uur

Wim-Paul Theodorus Josephus Lodewyck Pouw

geboren te Murcia (Spanje)

Erasmus University Rotterdam

Promotiecommissie

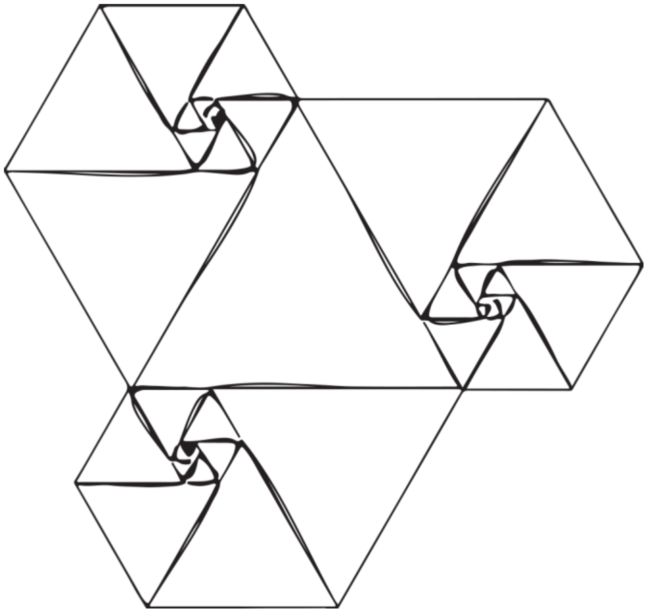
Promotoren	Prof.dr. T. van Gog
	Prof.dr. F. Paas
	Prof.dr. R. A. Zwaan
Overige leden	Dr. D. Pecher
	Prof.dr. H. Bekkering
	Prof.dr. S. Kita

Contents

1 2 3 4 5 6 7 8 9 10 R

Chapter 1

General Introduction



The human hand has been held to be of crucial importance for understanding (the evolution of) human intelligence (Arbib, 2005; Radman, 2013). Next to the unique morphological characteristics of the human hand (e.g., opposing thumbs relative to other digits) that allow humans to craft the environment in more flexible and sensitive ways than other animals, the *use* of the hands by humans is also truly unique (Tallis, 2003). Notably, humans often move their hands in signifying ways that mirror or complement the content of speech and thought (Kendon, 2004). To set these unique but ubiquitous types of hand-movements apart from the hands' most concrete manipulations, these hand-movements are referred to as gesticulation, or gestures in short. This dissertation is about the various ways that manual activity, such as gestures, but also actions on the environment, can support learning and problem solving.

Upon taking a closer look, the idea that manual actions "support" cognitive processes is by classic psychological interpretations of "action", quite a risky claim (Newell & Simon, 1972). That is, actions are - sometimes implicitely - held to be output (i.e., the result) of cognitive processes. This assumption involves the intuition that when a task-relevant movement is produced, say writing out a calculation on a piece of paper, the psychologically interesting stuff has already happened. More precisely, the physical movements of writing out the calculation are deemed *reflective* and an *output* of *mental* calculations that have occurred before the act of writing out the calculations. Cognition is as it were sandwiched between perception and action, and cognition should be logically distinguished from these peripheral input-output modules (Hurley, 2002).

In the recent decades the action-as-output assumption has come under scrutiny by approaches that sail under the banner of Embodied Cognitive Science (e.g., Hurley, 2002; Keijzer, 2001; for an overview see Shapiro, 2010). "Cognitive Science", as it involves insights from a wider range of disciplines, including philosophy (e.g., Clark, 1997, 2008; Dewey, 1896; Merleau-Ponty, 1956), psycholinguistics (e.g., Lakoff & Johnson, 1980; Zwaan, 2003), psychology (e.g., Barsalou, 1999; Gibson, 1969), and robotics (e.g., Brooks, 1991; Van Gelder, 1998). "Embodied" as these approaches assert, in more or less radical ways (e.g., Barsalou, 1999; Chemero, 2009; Hostetter & Alibali, 2008; Pouw, De Nooijer, Van Gog, Zwaan, & Paas, 2014), that cognitive processes are dependent on bodily processes, that is, the body's interaction with the environment. As such, mental inner processes, although necessary, are not sufficient for explaining cognition - the body and the environment need to be taken into account as well.

Given the exponential outgrowth of embodied cognition perspectives in the recent decade or two, "embodied cognition" has come to mean many different things to many different researchers (for reviews see Pouw & H. de Jong, 2015; Shapiro, 2010; Wilson & Golonka, 2013). To avoid confusion, therefore, some organization is necessary. In fact, at several places in this dissertation it is argued that failing to concretize the different underlying claims in embodied approaches may lead to inconsistencies that can reduce a theory's explanatory power (Pouw et al., 2014; Pouw, Van Gog, Zwaan, & Paas, in press).

An important, albeit still broad distinction between embodied approaches, is the direct versus indirect involvement of bodily processes in cognition. We will start with the more straightforward embodied approach, which assumes a direct cognitive function of bodily processes of the cognitive agent. In this dissertation, this kind of approach is often referred to as "embedded/extended cognition" (Pouw, De Nooijer et al., 2014). This approach can be introduced by a quote (recurring throughout this dissertation), concerning the act of drawing a picture:

> "One draws, responds to what one has drawn, draws more, and so on. The goals for the drawing change as the drawing evolves and different effects become possible, making the whole development a mutual affair rather than a matter of one-way determinism." (Bredo, 1994, pp. 28).

Key to this example is that drawing a picture does not imply a complete inner mental picture that is transformed into corresponding hand-movements (i.e., drawing). Although some inner mental process is necessary to understand the act of drawing, it is unsufficient for explaining how the drawing of a picture is accomplished. Instead, understanding the cognitive act of drawing, requires taking into account the ongoing loop of acting and perceiving its consequences. The implications of this drawing example are far from trivial. Researchers who are sensitive to this broader focus have shown, for instance, that children learn to understand fractions (1/5 of 10) by moving tiles around, and by doing so, discover that objects that are grouped together can be seen as part (fraction) of a greater whole (Martin & Schwartz, 2005). Without such manual exploratory actions, children are less likely to discover the concept of fractions. Notably, even when they are handed the partitioned solutions from the start, children still drastically underperform compared to children who explore the tiles manually. In similar vein, when we write out calculations (say 5+4x6), it has been shown (Landy & Goldstone, 2007), that we automatically space the symbols in a way that corresponds with the order of precedence, i.e., 5+4 x6 (note the spaces). Such spacing affects the way such calculations are performed. Many such examples will be reported in this dissertation.

Upping the ante, the simple example of drawing that was presented above can also be mapped onto the more complex bidirectional effects of action and cognition that occur on shorter time-scales. Take J. J. Gibson's famous assertion that perception is something we do (Gibson, 1979). According to this view, the perceiving organism does not construct a 3-D inner world through some cognitive transformations of static 2-D sense-data afforded by a single bi-ocular sample of the environment. Rather, visual information changes as soon as one moves and the motor-driven visual changes (and invariants present over those changes) provide the type of information to perceive the environment directly, as it is relevant for the organism, thus 3-D. According to Gibson, just as in the case of drawing, cognition must be understood as a temporally extended loop of perception and action - and this leads us to a different and less complex picture of the *internal* cognitive processes in such a wider organism-environment whole (e.g., Gibson, 1969; O'regan & Noë, 2001). That is, internal cognitive processes are *embedded* in immediate bodily processes, and these bodily processes may, at times, *extend* the cognitive capacities of the embodied organism.

Yet, of course, much of human cognition does not seem to be directly guided by action (e.g., thinking about tomorrow's chores; working out a problem mentally). In such cases it seems difficult to maintain that bodily processes are directly necessary for understanding such cognitive processes. However, according to a major branch of embodied cognition, often referred to as grounded cognition (Barsalou, 1999), cognition is still dependent on bodily processes in such instances, but in a more indirect way. That is, the history of bodily interactions with the environment shapes the way cognitive processes unfold, and cognitive processes unfold by recruiting the information structures (or brain-mechanisms) that were initially active for guiding interactions with the environment (Barsalou, 1999; Hostetter & Alibali, 2008; Lakoff & Johnson, 1990). An intuitive example demonstrating that the structure of human thought is dependent on bodily processes is the standard use of a ten-base decimal system in mathematics. This seemingly arbitrary (from a purely objective standpoint) system in all likelihood developed as *the* standard because basic arithmetic skills in humans emerged from using our fingers as arithmetic operators (Lakoff & Nuñez, 2000). As mentioned, according to indirect embodied approaches it is not only that our cognitive processes are constrained by previous bodily experiences, they also operate through functional reuse of those bodily experiences. Indeed, researchers sympathethic to this view suggest that activations of hand motor-areas in the brain that are commonly found when subjects engage in mental counting and calculations are evidence that such motor-activations are still functionally involved in mental artithmetic operations (e.g., Andres, Seron, & Oliver, 2007; Roux et al., 2003; Zago et al., 2001). Across the neuroscientific board, it is indeed well-established that during more abstract cognition such as reading and mental imagery, sensori-motor brain areas are consistently activated without apparent physical movements (e.g., Anderson, 2014). Such findings can be understood from the idea that the sensori-motor activations that initially evolved to guide interactions with the environment become constitutive for more abstract cognition. Yet, the common criticism to this interpretation is that activation does not necessitate functional involvement (e.g., Mahon & Caramazza, 2008), that is, such activation may be epiphenomenal, driven by basic association mechanisms. In light of this criticism, one of the major challenges for this type of embodied view is to show in which way the information that typically governs basic perception-action processes may come to be functionally employed in the service of more abstract cognitive processes. Thus, research in experimental psychology is identifying the *functional* mechanisms underlying the behavior under study. The theoretical and empirical studies presented in this dissertation aim to contribute to that goal.

Organization and Outline of this Dissertation

If cognition indeed depends directly or indirectly on bodily processes, then what are the implications for learning and problem solving? In other words, what type of behaviors and interventions would we expect to impact learning and problem solving if cognitive processes are dependent on bodily processes? These general questions formed the starting point for the more specific questions addressed in the theoretical and empirical studies presented in this dissertation. The dissertation is organized in three parts, and the chapters roughly follow a chronological order of completion during the PhD project.

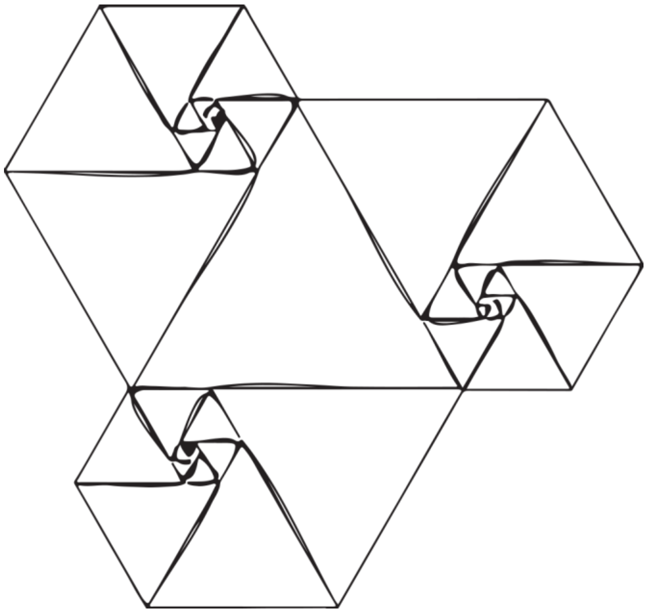
In **Part 1** the applicability of embodied cognition approaches for improving learning from instructional animations is explored. This part starts out with a theoretical chapter that lays the groundwork for the subsequent experimental studies. **Chapter 2** provides an analysis of current literature on the effectiveness of instructional manipulatives, and how these findings might relate to principles put forward by direct and indirect embodied approaches to cognition as compared to more traditional approaches. Instructional manipulatives are physical objects that invite the learner to interact with them, and through this interaction, they afford acquisition of knowledge that is difficult to obtain by passive observation alone. As such, a test of the effectiveness (or lack thereof) of these instructional manipulatives provides an ideal testbed for some principles of embodied cognition. **Chapter 3** essentially tests an assumption identified in Chapter 2 regarding the effectiveness of learning through physical interaction. It is often assumed, rather than directly tested, that improvements of learning through physical interaction can be attributed to the type of meaningful information that emerges in the interaction. Therefore we manipulated the meaningfulness of the information that could be obtained in interacting with an instructional animation about levers through the use of a Wii Board. As such we could assess whether physical interaction meaningful to the learning principle (i.e., movements needed to be performed that perfectly covaried with the mechanics of the seesaw) afforded information that is functionally reused in solving problems about levers after training. Such information was manipulated to be absent in the non-meaningful condition, where the required movements did only imperfectly covary with the mechanics of the seesaw. Finally, in **Chapter 4**, we report a study with primary school childrens' learning with a similar instructional animation about class 1 levers. Instead of providing learners with an opportunity to interact with the animation as in Chapter 3, we draw on indirect approaches to embodied cognition that predict that learning about some abstract principle might improve if it is related to knowledge obtained during daily interaction with the environment. As such we test whether childrens' learning about levers can be improved by mapping the physical forces that act on the child's own body (i.e., the arms) with the physical forces that occur in similar ways on a lever.

In **Part 2** the occurrence of non-communicative hand-gestures and their influence on problem-solving performance are investigated. Again, this part starts with a theoretical analysis, presented in **Chapter 5** that evaluates current theoretical approaches to cognitive function of gesture. It is suggested that there are some logical inconsistencies common to present perspectives on the embodied nature of gesticulation, as these assert that gestures are supporting cognition, whilst failing to characterize what cognitively potent consequences arise when moving the hands (i.e., gesturing). If gestures do not have cognitively potent consequences, they cannot (logically) support cognition. It is argued that our understanding of gesture and cognition should draw more from direct embedded/extended approaches to embodied cognition. In **Chapters 6 and 7,** drawing from hypotheses formulated in Chapter 5, it is investigated whether non-communicative gestures would spontaneously occur in the absence of speech (i.e., co-thought gestures) during mental problem solving, and whether gesturing (either spontaneously or when instructed to do so) during mental problem solving would improve subsequent physical problem solving under conditions of higher cognitive load (when working memory capacity is low, and task difficulty is high). This was investigated in adults (Chapter 6) as well as primary school children (Chapter7). In **Chapter 8** we report a study designed to investigate the possible functional role that gestures may have in mental problem solving by gauging the effect of gesture versus no gesture on gaze-behavior. We predicted that gestures bring about stable proprioceptive states that can be used to mark space in similar ways as has been found in how problem solvers use gaze-transitions.

In **Part 3** the main ideas in the dissertation are synthesized and interpreted in light of recent developments. In **Chapter 9** a perspective on gesture is presented that aims to reconcile the indirect versus direct juxtaposition made throughout the dissertation concerning embodied approaches to the role of manual activity in cognition. This is reconciliation is made by introducing a recent theoretical development in cognitive science, namely, Predictive Processing Perspectives. In the general discussion chapter (**Chapter 10)**, the main results of the dissertation are summarized and briefly discussed.

Chapter 2

An embedded and embodied cognition review of instructional manipulatives*



*This chapter has been published as Pouw, W. T. J. L., Van Gog, T., & Paas, F. (2014). An embedded and embodied cognition review of instructional manipulatives. *Educational Psychology Review*, *26*(1), 51-72. doi: 10.1007/s10648-014-9255-5

An embedded and embodied cognition review of instructional manipulatives

Recent literature on learning with instructional manipulatives seems to call for a moderate view on the effects of perceptual and interactive richness of instructional manipulatives on learning. This 'moderate view' holds that manipulatives' perceptual and interactive richness may compromise learning in two ways: 1) by imposing a very high cognitive load on the learner, and 2) by hindering drawing of symbolic inferences that are supposed to play a key role in transfer (i.e., application of knowledge to new situations in the absence of instructional manipulatives). This paper presents a contrasting view. Drawing on recent insights from Embedded Embodied perspectives on cognition, it is argued that: 1) perceptual and interactive richness may provide opportunities for *alleviating* cognitive load (Embedded Cognition), and 2) transfer of learning is not reliant on decontextualized knowledge but may draw on previous sensorimotor experiences of the kind afforded by perceptual and interactive richness of manipulatives (Embodied Cognition). By negotiating the Embedded Embodied Cognition view with the moderate view, implications for research are derived.

Introduction

In a seminal but critical paper on instructional manipulatives and their applications for the classroom, Ball (1992) stated "Understanding does not travel through the fingertips and up the arm" (p.3). This statement was meant to go against an overly simplistic view (or "magical hope"; Ball, 1992) prevalent in the literature concerning the effectiveness of learning with *physical* manipulatives. Many scholars today have followed suit (Brown, McNeil, & Glenberg, 2009; Kaminski, Sloutsky, & Heckler, 2009; McNeil & Jarvin 2007; Sarama & Clements, 2009; Sherman & Bisanz, 2009; Uttal, Scrudder & DeLoache, 1997), suggesting that "physicality is not important" and rather "their *manipulability* and *meaningfulness* make them [manipulatives] educationally effective" (original emphasis; Sarama & Clements, 2009, pp. 148). Indeed, it has been suggested that previously identified virtues of *physical* manipulatives - learning through concrete and perceptually rich physical practices - are not the drivers of learning (e.g., Triona & Klahr, 2003; Zacharia & Olympiou, 2011), and can even be detrimental to learning (e.g., DeLoache, 2000, 2004; Finkelstein et al., 2005; Sloutsky, Kaminski, & Heckler, 2005). This "moderate view" has led to a trend towards minimizing perceptual and interactive richness of manipulatives, as evidenced by the upsurge of *mouse-based virtual* manipulatives (e.g., Clements, 2000; Moyer, Bolyard, & Spikell, 2002) that compared to their physical counterparts differ in perceptual aspects (e.g., information gained from touching objects vs. manipulating them with a mouse) as well as interactive aspects (e.g., mouse-based interaction is constrained to one hand).

Drawing on insights from embedded embodied perspectives on cognition (Barsalou, 1999, 2008; Clark, 2005, 2008; De Vega, Glenberg, & Graesser, 2008; Hutchins, 1995; Kiefer & Trumpp, 2012; Lindgren & Johnson-Glenberg, 2013; Shapiro, 2011; Wilson, 2002; Winn, 2003) and state-of-the art research on physical and virtual manipulatives and including tangible-user-interfaces[1] (Manches & O'Malley, 2012) we suggest that some of the assumptions that underlie the moderate view are to some extent misguided, most centrally the assumptions that:

[1] Tangible User interfaces go beyond classical user interfaces (e.g., mouse, keyboard) and are designed to provide more natural or functional *physical* manipulation of virtual objects (Manches & O'Malley, 2012; O'Malley & Fraser, 2004; Shaer & Hornecker, 2010).

I. Higher perceptual and interactive richness of manipulatives imposes a high cognitive load on the learner, resulting in lower learning outcomes (Brown et al., 2009; McNeil & Jarvin, 2007; Sarama & Clements, 2009; Uttal et al., 1997)[2].
II. Transfer of learning from manipulatives involves a change in representation from concrete to symbolic, which is hindered by higher perceptual and interactive richness (Kaminski et al., 2009; Uttal et al., 1997; for an overview of this popular view in education see Nathan, 2012).

Without dismissing the empirical evidence upon which the above assumptions are based, we suggest that a viable case can be made for a more Embedded Embodied perspective on learning with manipulatives, namely:

I'. Under certain conditions, perceptual and interactive richness can alleviate cognitive load imposed on working memory by effectively embedding the learner's cognitive activity in the environment (Embedded Cognition claim).

II'. Transfer of learning from manipulatives does not necessarily involve a change in representation from concrete to symbolic. Rather, learning from manipulatives often involves internalizing sensorimotor routines that draw on the perceptual and interactive richness of manipulatives (Embodied Cognition claim).

We hasten to note that we will not argue for an exclusively Embedded Embodied approach to learning from manipulatives; rather, this review attempts to negotiate the findings from the Embedded Embodied perspective with findings associated more with the moderate view. For example, while researchers holding the moderate view may suggest that when physicality is not important this is evidence against the Embedded Embodied view (e.g., Triona et al., 2005), our review will show that this would be an overly simplistic understanding of the relevance of Embedded Embodied Cognition to manipulatives. By combining findings from both streams of research, we aim to develop a more balanced view of how Em-

2 For example, "When children interact with manipulatives, their cognitive resources may be committed to representing and manipulating the objects and may be largely unavailable for other processes, such as accessing relevant concepts or implementing appropriate procedures" (McNeil & Jarvin, 2007, p. 313).

bedded Embodied Cognition might guide the design of instructional manipulatives.

In the next section, we focus on *Embedded Cognition*, which suggests that effective learning depends on how learners coordinate their cognitive activity in concert with bodily and environmental resources. Embedded Cognition amounts to the idea that cognition is afforded and constrained by ongoing interactions between body and environment, emphasizing an intimate relation between external artifacts and cognitive processes (Clark, 2008; Hutchins, 1995; Kirsch, 1995, 2010; Wilson, 2002). In the subsequent section, we focus on *Embodied Cognition*, which amounts to the claim that knowledge is grounded in sensorimotor routines and experiences (Barsalou, 1999, 2008; Lakoff & Johnson, 1999; Lakoff & Núñez, 2000)[3]. In this section, we discuss empirical evidence that suggests that transfer of learning does not necessarily involve a concrete-to-abstract shift.

Importantly, Embedded Cognition *and* Embodied Cognition are complementary in their analysis of the role of the body in cognition (Shapiro, 2011; Wilson, 2002). Whereas Embedded Cognition focuses on the continuous coupling or "on-line" interaction with the environment, Embodied Cognition focuses on the role of these previously acquired sensorimotor experiences in "off-line" cognitive activity (i.e., disembedded from the environment).

For this review we provide a selective overview of state-of-the-art research in cognitive and educational psychology. At the end of both the Embedded *and* Embodied Cognition section, we provide a short intermediate discussion, making a connection with empirical evidence that aligns with a more moderate view. In the conclusion we provide a brief overview of our main conclusions, research challenges, and educational implications in relation to learning with manipulatives.

Embedded Cognition

According to theories of Embedded Cognition, cognitive activity is not something that simply happens internally, but involves a continuous transaction between current states of the brain, body and the environment (Clark, 2008). As such, understanding cognition requires a broader level of analysis that considers how we use our body and the world during the unfolding of cognitive processes (Clark, 2008; Hutchins, 1995; Kirsh, 2010; Wheeler, 2007). Examples of Embed-

---

[3] Note that the literature does not always explicitly distinguish between Embodied Cognition and Embedded Cognition (e.g., Risko et al., 2013; cf. Wilson, 2001).

2

ded Cognition are readily apparent: reducing working memory load by making notes during a conversation, using fingers to keep track of counting, asking another person to remind you of something, or using a tall building for navigating your way home, which alleviates the need to retain street names or spatial maps. As these examples show, Embedded Cognition refers to the adaptive flexibility of cognitive processes during interaction with the environment.

Although we can make the case that cognition might at times be disembedded during activities such as mental arithmetic, thinking about tomorrow's chores, talking about something absent (et cetera), learning from manipulatives pertains to an embedded cognitive situation. That is, learning with manipulatives involves a tight coupling of external artifacts with perceptual and cognitive processes, in which the artifacts structure the learner's cognitive states (Clark, 2005). As such, learning from manipulatives does not differ in kind from examples we have provided above, such as finger-counting. As the finger counting example also makes clear, however, manipulatives may in some cases become ill-suited for supporting cognitive states, just as arithmetic with large numbers might become difficult to perform through finger-counting. Thus, learning from manipulatives is always an embedded phenomenon in which some cognitive processes are more easily maintained than others.

A theoretical implication of Embedded Cognition, is that the states of the body and the environment can be considered extra-neural contributors to (Hutchins, 1995; Norman, 1988), and in a more radical reading, external vehicles of cognition (Clark, 2008; Clark & Chalmers, 1998). Not only does Embedded Cognition hold the more widely accepted claim that the external environment may serve as external working-memory, the transactions between the environment and the learner might dramatically change the way in which cognitive processes unfold. For example, thinking with or without a notepad may have dramatically different cognitive profiles. Bredo (1994, pp. 28) provides an example of this dynamic coupling of cognition with the environment: "One draws, responds to what one has drawn, draws more, and so on. The goals for the drawing change as the drawing evolves and different effects become possible, making the whole development a mutual affair rather than a matter of one-way determinism". As such, in Embedded Cognition, the external environment has an important status for understanding the way cognitive processes unfold.

That we use aspects of our environment in order to reduce cognitive load was demonstrated in an influential study by Ballard, Hayhoe, and Pelz (1995; see also Haselen, Steen, & Frens, 2000; Hayhoe, Pook, & Rao, 1997). In this study participants were asked to reproduce a pattern of colored blocks from a model as quickly as possible by clicking-and-dragging randomly ordered colored blocks from a *resource space* and ordering them in a *workplace*. Eye-movements were monitored to provide insight into the strategies that are involved in solving this problem. It was found that participants opted for a 'minimal memory strategy' as indicated by the many switches of eye fixations between the model, resource and workplace area. That is, to minimize the components to be retained in memory, participants tended to gather information incrementally by first attending to the color and then the position, all just in time, instead of memorizing information all at once. Note that the use of a minimal memory strategy also emerged when participants physically manipulated real blocks (Ballard et al., 1995).

To give another famous example, in a study by Kirsh and Maglio (1994) it was found that effective problem-solving behavior in the game Tetris does not solely rely on action that brings one physically closer to one's goal, which they termed "pragmatic actions". Rather, problem-solving also relies on "epistemic actions" that effectively structure the environment so as to uncover information that is hidden or cognitively demanding to compute. For instance, they found that advanced players, more often than less-advanced players, tended to rotate zoids physically instead of mentally when determining whether a zoid would fit the already-placed zoids at the bottom. This study as well as others (Gaschler, Vaterrodt, Frensch, Eichler, & Haider, 2013; Stevenson & Carlson, 2003) show that the environment does not only allow for offloading, but that efficient problem-solving evolves over time during interaction with the environment, and is dependent on how the agent learns to effectively negotiate internal and external resources.

In both of these experimental demonstrations it seems that the cognitive system prefers to manage working-memory load by making use of external resources when available. However, whether external resources are used also seems to depend on how readily information can be re-achieved from them. In a study by Gray and Fu (2004) participants were confronted with a task wherein subtle retrieval costs of attaining external task-relevant information in the context of programming a simulated VCR were manipulated. Lowering the ease of retrieval of information from external resources, from a single glimpse or an additional mouse-click, changed the cognitive strategy of the subjects. When external information was directly accessible, participants leaned primarily on retrieving "perfect-knowledge-in-the-world". However, when this external information was only indirectly available through a mouse-click, participants were more likely to retrieve it from memory. Although the reliance on internal memory lead to a higher number of mistakes it was shown that this "imperfect-knowledge-in-the-head" was more quickly available compared to retrieving information externally. That is, based on computational modeling Gray and Fu (2004) estimated the retrieval effort of relevant information expressed in the amount of milliseconds it takes to retrieve or recall information and showed that participants opt for the quickest problem-solving strategy with no a-priori preference for internal or external resources. It seems therefore that the cognitive system "tends to recruit, on the spot, whatever mix of problem-solving resources will yield an acceptable result with a minimum of effort" (Clark, 2008, pp. 13; see also Borst, Buwalda, Rijn, & Taatgen, 2013; Fu, 2011).

To date, Embedded Cognition research has been primarily focused on how and when the environment is used in terms of memory distribution (see also Droll & Hayhoe, 2007; Gray, Sims, Fu, & Schoelles, 2006). Although current research is extending its applications (e.g., Risko, Medimorec, Chisholm, & Kingstone, 2013) it is still ill-understood how information is encoded during embedded cognitive situations, and whether different interactive possibilities for distributing internal and external resources result in different learning outcomes, that is, in different representations in long-term memory (Fu, 2011). Especially the latter question seems to be of central importance for understanding how perceptual and interactive properties of manipulatives may affect learning. As such, as it is currently left unanswered, we should be hesitant to accept any claims about effects of perceptual and interactive richness of manipulatives.

Embedded Cognition and Instructional Manipulatives

The theory of Physically Distributed Learning (Martin & Schwartz, 2005; Schwartz & Martin, 2006) suggests that the environment changes the way in which learning unfolds. According to this theory, the learning affordances of physical manipulation can be mapped onto four separate quadrants that roughly categorize physical learning in terms of the *stability* and the *adaptability* of the learner's *ideas* and the *environment*; the quadrants *Repurposing* and *Mutual Adaptation* being important for present purposes. The quadrant called Repurposing, pertains to a situation similar to the above mentioned Tetris-players who have learned to repurpose pragmatic actions that bring one closer to one's goals for epistemic actions that reduce computational load (Kirsh & Maglio, 1994). In this example the environment is adaptable but ideas remain largely unchanged.

Most interesting for present purposes, however, are such situations in which new ideas arise through physical adaptation of the environment, called Mutual Adaptation. Martin and Schwartz give an example of a young child asked to come up with a one-fourth share of eight candies. Children often focus on the *one* of one-fourth, which leads them to adopt "one candy" to be the right answer. However, in physical interaction with eight candies, the child might push two candies apart, which increases the likelihood of reinterpreting the new arrangement of two candies as one group, putting the child on "a trajectory" to learn that one-fourth of eight means attaining "four groups of two" first (Martin & Schwartz, 2005, pp. 590). Thus, mutual adaptation involves structuring the environment haphazardly without preconceived goals that affords new interpretations difficult to obtain by thought alone. As such, the theory of Physically Distributed Learning extends the current focus of Embedded Cognition, suggesting that the environment also changes the way learning unfolds.

Martin and Schwartz (2005) have empirically substantiated the theory of Physically Distributed Learning through multiple experiments (see also Martin, Lukong, & Reaves, 2007). In the first two experiments reported by Martin and Schwartz (2005), children of nine to ten years old solved fraction operator problems (e.g., one-fourth of eight) with physical pie or tile wedges using physical manipulation *and* pictorial line drawings of pie or tile pieces using a pen to highlight partitions. In the first two experiments it was found that children using physical manipulatives solved more problems correctly which was measured by the number of partitions created correctly and the number of correct answers that were provided verbally. More importantly, it was shown that physical self-guided partitioning was the driver of understanding rather than mere perception of the desirable end state, a correctly pre-partitioned organization. According to Martin and Schwartz (2005), physical open-ended interaction allows for exploration and search for new interpretations and structures, which benefits learning (see also Martin et al., 2007).

Complementary to these results, it has recently been shown that the beneficial role of physically manipulating the external environment enhances task performance in physics education (Stull, Hegarty, Dixon, & Stieff , 2012). In a set of experiments, university-level physics students had to translate one type of diagram into another, called a diagrammatic translation task, which requires spatially translating the model into the other model's particular perspective. In all three experiments it was found that students' translation accuracy of one 2D representation into another was promoted by active use of a concrete 3D model during the task (a classic ball and stick manipulative). Importantly, only the active physical use of the 3D model, as opposed to mere perception of the model, promoted task performance. In line with Kirsh and Maglio (1994), it was explained that the concrete model aids students in externalizing spatial rotation operations (Stull et al., 2012).

A critical note of concern, however, is that based on these data one cannot fully disentangle the role of self-guided physically manipulating objects from the visual input that this process also generates. Even though seeing the end state arrangement (Martin & Schwartz, 2005) or the model (Stull et al., 2012) was not beneficial for learning, it is possible that watching someone else dynamically structuring the materials would produce the same learning benefits (for a review of the effectiveness of observational learning in educational contexts, see Van Gog & Rummel, 2010; for the effectiveness of observing someone else exploring a problem space, see Osman, 2008). Nevertheless, active skillful manipulation of these materials might in itself form the basis for performing similar cognitive tasks in the absence of the manipulatives, but this remains an open empirical question.

Martin and Schwartz (2005; Experiment 3) further explored whether there is an interaction between prior knowledge and environmental structure in instances of physically distributed learning. As a highly structured learning environment in the context of solving fraction-problems pie wedges were used since these already have a part of whole partition, whereas tiles were used as unstructured materials. It was found that children performed more correct partitions in solving fraction addition problems for which they had high prior knowledge when materials were structured compared to unstructured materials. In contrast, performance on multiplication problems for which children had low prior knowledge was unaffected by structure of the environment. They suggested that this finding indicates that a more mature understanding of the task allows for repurposing the environment more flexibly, with performance on low familiar tasks being more dependent on the environment's stability for action. However, they also raise a very interesting concern. That is, although a highly structured environment can aid problem solving, it might prevent learners from developing their own interpretation of how to solve a problem.

Indeed with children learning fraction additions in three sessions over a period of a week, it was found that those who had learned with pie wedges showed a lower ability to transfer skills to other manipulatives than children who had used tiles (Experiment 4 and 5). Martin and Schwartz (2005) explain this finding in that pie wedges' structure gives the learner a part-of-wholes-interpretation "for free", presumably preventing children to learn how to make and interpret such groupings and whole structures by themselves. Simply put, externalized cognitive operations might in some instances reduce the necessity to understand its function (e.g., that pieces are part of a whole).

Although research cited above already offers considerable evidence that Embedded Cognition is an important factor for learning, a more recent example shows that possibilities for physical interaction indeed change the learning trajectory. In a set of experiments Manches, O'Malley, and Benford (2010), sought to find out whether qualitative differences in manipulation predicted children's problem-solving strategies in a numerical partitioning task. In this task children are asked to provide all the different ways in which a certain amount can be combined (e.g., the number of ways seven can be recombined [e.g., seven and zero, zero and seven, six and one, five and two, et cetera]). In the first study reported by Manches and colleagues (2010), children ranging from five to seven years old were first asked to solve a partitioning problem without manipulation of any material (no material condition), and to subsequently solve two additional partitioning problems with paper and pencil (paper condition) and physical blocks (physical condition; order of physical and paper condition was counterbalanced). It was found that children provided significantly more unique solutions in the physical condition as opposed to the no material- and the paper condition. Qualitative observations were made that could explain this difference in terms of particular affordances that physical manipulatives have. For example, bimanual manipulation allowed for moving multiple blocks at a time and/or keeping track of block's locations through haptic sensation, which was not possible in the other conditions.

In the second experiment it was investigated how the affordance of bimanual manipulation might have constrained particular use of strategies. It was predicted that when children ranging from four to seven years old are instructed to manipulate only one object at a time (constraint condition) it would lead to different strategies as compared to children in a no constraint condition. Indeed, it was found that strategies differed dependent on whether manipulation was constrained. For example, reversing combinations (e.g., five and two into two and five) is much easier to perform when manipulating multiple objects at once than serial one-by-one manipulation. In the third study this effect was replicated for a portion of the sample in a slightly different set-up. The constraint condition was now set-up as a virtual manipulative (children could click-and-drag only one virtual object on the screen). Taking the results together, this study suggested that with unconstrained physical manipulation come particular affordances that shape the trajectory of young children's learning of numerical partitioning.

Importantly, however, unconstrained physical manipulation has also been shown to be sub-optimal for learning (Stull, Barrett, & Hegarty, 2013). Stull and colleagues (2013) let students interact with a Tangible User Interface (TUI) that was designed to combine affordances of virtual and physical manipulatives. The TUI included sensorimotor features that are typically afforded by physical manipulatives, such as stereo-depth cues and a direct manipulation interface (see Stull et al., 2013, for details). The only features that differed from a physical model were I) the shape of the tangible interface and its virtual representation (molecular model) were not the same, and II) interactivity was constrained such that the students could only rotate the model around the axis of a single molecular bond. Note that physical manipulatives allows for rotations around an indefinite number of axes. In these experiments, learners had to perform a diagrammatic matching task, which involved manipulating the model to match the orientation of a particular 2D molecular diagram. Although accuracy levels were the same for both model types, the physical manipulative condition was significantly slower in completing the task (in comparison to the TUI). This higher efficiency in the TUI condition was ascribed to the constrained interactivity of the TUI which automatically focused students on the most task-relevant interactions. Indeed, additional analysis revealed that students who first worked with the TUI performed less irrelevant bond-rotations in comparison to students who had worked with physical manipulatives first. As such constrained interaction might aid in learning to efficiently solve problems in similar unconstrained situations.

A final example for the way in which interaction possibilities may change learning comes from a study reported by Antle (2012) and Antle, Droumeva, and Ha (2009). In this study interaction styles emerging from different manipulatives were investigated in the context of a Jigsaw Puzzle Task with dyads of children ranging from seven to ten years old, using either traditional physical manipulatives (PM), mouse-based virtual manipulatives, or a TUI. The TUI was a tabletop prototype with normal puzzle pieces; action was mapped through an infrared camera that allowed for audiovisual feedback when a piece was correctly placed. By calculating relative measures for interaction style to account for single versus multiple input differences (for details, see Antle, 2012; Antle et al., 2009) it was shown that the PM and TUI conditions, which allowed for bimanual manipulation, resulted in more time spent performing epistemic actions. For example, grouping corner-, edge-, or same-color pieces into piles. Furthermore, it was found that more direct actions were taken in the PM and TUI condition as opposed to mouse-based virtual manipulatives. Although the design of this study does not allow for empirically rigorous conclusions about performance or learning (as Antle, 2012 concurs), it does, together with findings from the previous studies, suggest that properties of the interaction may shape the way in which cognitive processes and learning might unfold. However, based on these studies it is hard to derive clear design guidelines regarding unconstrained interactive richness.

Intermediate Discussion: Embedded Cognition and Manipulatives

In the previous section it was shown that manipulatives afford possibilities for reducing internal computational load through interaction. Furthermore, such possibilities are quite easily and automatically incorporated into learning behaviors. Arguably the most important contribution of the Theory of Physically Distributed Learning and the empirical evidence that supports it, is that although learning environments that are pre-structured and thus constrained may reduce problem solving steps and improve task performance, this reduction of task-load does not necessarily benefit transfer of learning. Children who learned to solve fraction problems with pie-wedges, in comparison to learning with tiles, were less able to transfer this knowledge to other materials that did not already have this part-of-wholes interpretation in its structure. Schwartz and Martin (2006) make the analogy with research on Dienes's (1973) base-ten blocks; children who become increasingly efficient to operate base-10 blocks for problem solving become dependent on (or "symbiotically tuned" to) these materials for its efficiency, underperforming in transferring this skill in the absence of these materials (e.g., isomorphic symbolic tasks; see Resnick & Omanson, 1987). The tentative lesson we might draw from this is that design of manipulatives should at times allow for self-discovery rather than pre-constrained problem solving when transfer of learning is the goal. As such, embedded learning might at times unfold best when it is learner-centered as opposed to being completely accommodated by the environment.

A further implication is that specific perceptual and interactive properties of manipulatives that might afford embedded learning stand in relation to the kind of bodily actions the learner can perform (Gibson, 1979). In the studies by Martin and Schwartz (2005) and Manches and colleagues (2010) it was shown that physicality of materials solicited specific patterns of interaction that led children to discover interpretations necessary for understanding the particular problem. "Solicited", in that children were simply drawn to the affordance of re-arranging the blocks and not driven by a pre-conceived end-state in mind. This arguably shapes the learning trajectory in that it leads to what Schwartz and Martin (2005) call "mutual adaptation"; adaptations to the environment further influence adaptations to children's interpretation.

To elaborate on this, the role of perceptual properties in embedded learning might be indirectly related to possibilities for interaction. The research discussed above provides insights on how perceptual richness affects learners' perception of possibilities for action (e.g., objects being physical rather than virtual). This can be appreciated by a modified interpretation of what Gray and Fu (2004) call *hard-* and *soft constraints*. That is, manipulatives have specific properties that make only certain actions possible (hard constraints). For example, consider a mouse-based virtual interface that only allows for uni-manual manipulation, or a pie-wedge that only allows for re-arranging parts in pre-set wholes. However, manipulative perceptual properties also determine which behavior given the possibilities is likely to be solicited (*soft-constraints*). For example, Manches and colleagues (2010) reported that children who were instructed to manipulate physical blocks one at a time had difficulty not to use two hands or manipulate multiple blocks. This resonates with a host of behavioral and neurological evidence on motor-affordances that has shown that perceptual properties of objects unreflectively solicit particular action-repertoires (Gibson, 1979; Snow et al., 2011; Symes, Ellis, & Tucker 2007; Van Elk, Van Schie, & Bekkering, 2014). In sum, whether an object is perceived to be easily manipulable impinges on the natural behavior it solicits from the learner.

Therefore, there is a case to be made that perceptual richness might impinge on how learners typically interact with the learning environment. As Dourish (2001) notes, "because we have highly developed skills for physical interaction with objects in the world - skills for exploring, sensing, assessing, manipulating, and navigating - we can make interaction easier by building interfaces that exploit these skills" (pp. 206). Therefore, suggesting that "physicality is not important" in manipulatives and rather their "manipulability and meaningfulness make them educationally effective" (Sarama & Clements, 2009, p. 148) might be at times misguided and involves an artificial distinction; perceptual richness may drive perceptions of manipulability.

The tentative conclusion we like to make up to this point is that contrary to the "moderate view" emphasis that perceptual and interactive richness of manipulatives can hinder learning, it should also be considered as an important source of learning. That is, perceptual and interactive richness may invite learners to interact in a certain way with the environment and therefore effectively embed learners' cognitive activities. In the final discussion we connect these insights with those from the upcoming review on Embodied Cognition, and discuss implications and suggestions for future research.

Embodied Cognition

Embodied Cognition holds that the *format* of cognition is sensorimotor or modal based instead of symbol-based (i.e., amodal; Barsalou, 1999, 2008; for an overview see Svensson, 2007). Furthermore, whilst the cognitive system might be disembedded and primarily dependent on internal cognitive processes in some cases, Embodied Cognition suggests that sensorimotor information made available during previous interactions is reused for internal cognitive processing. Thus Embedded Cognition emphasizes an ongoing "on-line" interaction with the environment whereas Embodied Cognition primarily focuses on how the body shapes disembedded or "off-line" cognition.

Embodied Cognition is therefore especially suitable for explaining how learning with manipulatives might impinge on cognitive activity in the absence of manipulatives (e.g., mathematical notations; mental arithmetic). The classic perspective on cognition (Fodor, 1975; Newell & Simon, 1972) holds that transferring knowledge learned in one situation to another is dependent on establishing a set of complex symbolic rules. According to this view, knowledge resides in a rule-governed semantic system that needs to be decontextualized from immediate sensorimotor states and the environment. In contrast to this traditional approach, the Embodied Cognition framework attempts to provide a more continuous explanation of perception and action on the one hand, and cognition on the other, by suggesting that cognition is constituted in sensorimotor experiences. More specifically, knowledge is derived from sensorimotor coded routines stored within a generalized system that was originally developed to control an organism's motor behavior and perceive the world around it (Anderson, 2008; Barsalou, 1999, 2008; Svensson, 2007).

Currently there is a great deal of interest from educational psychology in the notion of Embodied Cognition (Black, 2011; Calvo & Gomila, 2008, ch. 18; De Vega, Glenberg, & Graesser, 2008; Goldstone & Barsalou, 2005; Kiefer & Trumpp, 2012; Lindgren & Johnson-Glenberg, 2013). Often cited in this literature is Barsalou's (1999, 2008) perspective on Embodied Cognition, the Perceptual Symbol Systems Account. This perspective provides a fine-grained account of how knowledge might be embodied. In this account concepts are *grounded* in the re-activation of specific neural patterns in multiple modalities (e.g., motor system, visual system, et cetera) that were activated during previous interactions with the environment. These activation patterns are suggested to be captured in a single multimodal representation: a Perceptual Symbol (Barsalou, 1999).

Perceptual Symbols are not holistic or necessarily conscious vehicles of thought. Rather, Perceptual Symbols can selectively capture schematic aspects of sensorimotor regularities occurring in interaction that become stored in long-term memory (Goldstone & Barsalou, 2005). This allows for schematic extractions of perceptual but also introspective states that can be recombined in imagination. As such, concepts that are not readily available in the environment (e.g., a hammer made of pudding) might still be grounded in sensorimotor states by mashing the sensorimotor concept of hammer and pudding. Furthermore, it is held that perceptual symbols of very abstract concepts (e.g., truth, love) still rely on complex combinatorics of perceptual states (see also Lakoff & Johnson, 1999; Lakoff & Núnez, 2000). Importantly, in lieu of the principle of activation spread, particular sensorimotor states induced during interaction can trigger activation of Perceptual Symbols that activate stored sensorimotor information from previous experiences such that for example seeing a hammer induces modality specific simulations of the weight of the hammer.

There is increasing evidence that cognitions are intimately tied to the sensorimotor system (e.g., Kiefer & Trumpp, 2012; Pecher & Zwaan, 2005; Svensson, 2007). Indeed, the sensorimotor system has been found to be implicated in thought processes as diverse as reading, mental arithmetic, problem-solving, and conversely, semantic areas are often implicated in sensor-motor interactions suggesting that both systems are intimately related (Barsalou, 1999, 2008; Glenberg et al., 2008; Martin, 2007; Nathan, 2008). To give an example, research shows that merely reading words that have olfactory, gustatory, or motor connotations (e.g., garlic, jasmine, salt, sour, kick, pick) as opposed to neutral words, activates brain regions that are involved in smelling, tasting and moving (Barrós-Loscertales et al., 2011; Gonzales et al., 2006; Hauk, Johnsrude, & Pulvermüller, 2004). Furthermore, when subjects are mentally processing numbers, activation of motor-areas associated with finger-movements is consistently found (Andres, Seron, & Oliver, 2007; Roux et al., 2005; Zago et al., 2001). In sum, the current state of the literature suggests that knowledge-representations are intimately tied to the sensorimotor system, which raises the need to understand how the cognitive system draws from sensorimotor information that emerges during interaction with the environment.

Embodied Cognition and Manipulatives

In this section we give a representative overview of research on manipulatives that specifically claims to be, or in our view seems to be, relevant to Embodied Cognition. We review three streams of research on transfer that provide varying degrees of support for either an Embodied Cognition perspective or the more moderate view mentioned in the introduction that seems to suggest that abstraction is hampered by perceptual and interactive richness.

Transfer by internalizing sensorimotor information

The first line of research is well aligned with the Embodied Cognition perspective and shows that transfer of learning is simply internalization of sensorimotor information that is initially provided by the manipulative. To give a striking example: moderately advanced abacus users maintain high arithmetic capabilities during mental calculation in the absence of an abacus by "manipulating" what seems to be a mentally projected abacus. Such users often apply finger manipulations as if the abacus is physically accessible. Interestingly, expert abacus users even perform better in mentally manipulating as opposed to physically manipulating the abacus (Hatano, Miyake, & Binks, 1977; Hatano & Osawa, 1983). This suggests that having had a very high number of sensorimotor experiences with the abacus, can instantiate fully mental simulations without external support needed to maintain it. Importantly, the contention that non-verbal sensorimotor representations underlie mental calculation of abacus-users has recently been strengthened: performance of mental calculation in normal subjects is inhibited by verbal interference, whereas for trained abacus-users no interference effects are found (Frank & Barner, 2012).

Furthermore, a recent study showed that participants who had learned with either a physical or virtual abacus performed equally well in recognizing number representations of an abacus presented on paper; however, virtual abacus-trained participants performed worse on a transfer task that required more complex arithmetic operations with a physical abacus in comparison to participants who had trained with a physical abacus (Flanagan, 2013). Relatedly, in a study by Flusberg and Boroditsky (2011) on mental rotation it has been found that sensorimotor experience with objects that are difficult to manipulate actually hindered effective mental rotation of those objects, whereas easily manipulable objects promoted mental rotation. These studies show that sensorimotor simulations that underlie cognition are very sensitive to the experiences afforded by manipulatives. As such, if we understand transfer of learning as learning to think without manipulatives it does not necessarily involve decontextualization, but rather internalization of sensorimotor routines.

This development of internalized embodied knowledge seems to be a gradual process; that is, learners slowly dis-embed their mental activity from the environment. An obvious example is when children stop using finger-gestures to count. Moreover, as mentioned earlier, abacus users that have an intermediate level of expertise often use gestures to support their thinking whilst experts do not need such support for their mental calculations, which suggest a kind of transition state between relying on purely external to internal recourses. In a similar vein intermediate chess players perform better at thinking through moves (without manipulating the pieces) when a chess-board is present. In contrast, chess masters do not need external support in their mental chess-playing (Chase & Simon, 1973). A relevant study that provides insight on when external support is of importance comes from Kirsh (2009) in which subjects played a mental tic-tac-toe game with the experimenter. It was found that external perceptual support of a sheet with a matrix depicted on it as opposed to providing no support, a blank sheet, aided performance. However, this external support was only beneficial when the tic-tac-toe game was complex (4x4 matrix), and especially for subjects who scored low on spatial ability. Thus, this study suggests that external support is especially helpful when computational load is high, and this depends on whether the subject is effective of performing those computations internally (e.g., spatial ability; Kirsh, 2009). This might characterize how novices become experts. External structures are gradually internalized, and internalization being dependent on the learners' "representational stability" (Hutchins, 2005), that is, the ability to mentally stand in for external structures. For example, as demonstrated above, being low on spatial cognition - signifying a difficulty to produce a stable representation - leads to a higher need to lean on external support (Kirsch, 2009). Interestingly, the use of hand gestures can also be seen as an instance of external support to maintain representational stability (Chu & Kita, 2011; Chu, Meyer, Foulkes, & Kita, 2013; Radman, 2013).For example, it has been found that frequency of spontaneous use of gestures is correlated with having low ability in spatial imagery and rotation (Chu et al., 2013).

Taking these results together, from an Embodied Cognition perspective it can be argued that actively learning with manipulatives, can establish sensorimotor routines that are internalized (i.e., embodied); without the need to invoke symbolic rules as expertise develops. Thus in these specific cases interaction vs. thinking with a manipulative does not rely on a concrete to abstract shift; both modes of cognitive performance rely on the same representational format (sensorimotor routines), wherein increasing computational load is put on the brain as the learner is required to dis-embed cognitive activity (e.g., mental calculations).

Transfer by actually or mentally simulating text or science scenarios

The second line of research focuses on attaining conceptual and narrative understanding of texts and science-materials through manipulatives. In these cases the role of *grounding* a concept in sensorimotor experiences has been studied (Glenberg, Goldberg, & Zhu, 2011; Glenberg, Gibson, Goldberg, & Zu, 2012; Glenberg, Gutierrez, Levin, Japuntich, & Kaschak, 2004; for a review see De Koning & Van der Schoot, 2013). For example, in several experiments by Glenberg and colleagues (2004) first- and second-grade children read a text and manipulated toy-figures that referred to, and offered a way to enact, the scenario of the text. It was found that children enacting the text scenarios (compared to only reading them), were better at story recall, making inferences from the story, and in their understanding of spatial relations mentioned in the story. Furthermore, having had practice with physical manipulation of toys, children who had to re-enact the scenario mentally through imagination showed similar improvements. Importantly however, it has been found that positive effect of manipulatives for text comprehension can be attained by virtual manipulatives as well (Glenberg et al., 2011) and simply watching someone else enact the story can equally benefit learning (Marley, Levin, & Glenberg, 2007; Marley, Szabo, Levin, & Glenberg, 2011).

As Glenberg and Kaschak (2002) argue, these results suggest that understanding of text arises through simulating the scenario's content. Manipulatives offer a way to ground the scenario's content directly, as such promoting simulation processes that underlie text comprehension (Glenberg et al., 2004). Similar findings are obtained in science education, in which the role of physical versus virtual manipulatives has been studied extensively (T. De Jong, Linn, & Zacharia, 2013; Olympiou & Zacharia, 2012; Triona & Klahr, 2003; Triona, Klahr, & Williams, 2005; Zacharia & Constantinou, 2008; Zacharia, Loizou, & Papaevripidou, 2012; Zacharia & Olympiou, 2011).

For example, Zacharia and Olympiou (2011) investigated experimentation with heat and temperature by undergraduate students who interacted with either physical or virtual materials or both and were tested for conceptual learning through assessment of pre-test and post-test. In the physical condition the materials consisted of normal beakers, waters, hotplate et cetera, whereas virtual materials consisted of 2D approximations of those materials that could be manipulated with a mouse. It was found that participants learned equally across conditions (for an earlier study obtaining similar results see Zacharia & Constantinou, 2008). In another study, undergraduate physics students learned the workings and conceptual underpinnings of simple electrical circuits, such as voltage, and parallel vs. series circuits. In the critical conditions, during a 15-week physics course students either learned through concrete physical materials or interactive computer simulations thereof (Finkelstein et al., 2005). It was found that students learning with physical versus virtual materials performed worse on a test of conceptual understanding, as well in their evaluations of a set-up with physical materials.

These and other studies (Klahr et al., 2007; Triona & Klahr, 2003; Triona et al., 2005) consistently show that in many cases physical manipulatives are replaceable by virtual ones without learning costs. Based on this research it has been suggested that null (and negative; e.g., Finkelstein et al., 2005) results concerning physicality seems to contradict "the embodied nature of cognition [that] would seem to suggest that type of materials [i.e., whether they are physical or virtual] would influence student's learning" (Triona et al., 2005, pp.1). In similar vein, in a recent review on the role of physical and virtual manipulatives in laboratory education it was suggested that physical laboratories should promote learning by offering "…tactile information that, according to theories of Embodied Cognition, fosters development of conceptual knowledge" (T. De Jong et al., 2013, pp.305).

Importantly, however, if one indeed takes the position that previous findings (including research on text-comprehension) contradict with Embodied Cognition[4], one must have a clear understanding of what Embodied Cognition would predict in a particular context. Unfortunately, in all the previously reported studies it is not clearly explained why physical as opposed to virtual beakers, short springs, or toys would aid conceptual understanding, besides the broad but sim-

[4] It is important to note this is *a possible* position that can be drawn from the results, not necessarily a position that all the authors of the previously reported studies take.

plistic claim that tactile or multimodal experiences should aid conceptual learning (T. De Jong et al., 2013; Triona & Klahr, 2003; Zacharia & Olympiou, 2011).

What would a proper or more moderate reading of Embodied Cognition predict in these research contexts? Firstly, it cannot be denied that learners understand a physical beaker differently if one has haptic vs. no haptic experience with it, perhaps even producing richer multi-modal simulations when thinking about it. However, there is no reason to assume that this information aids "learning" of the sort assessed in these experiments. In a recent study in physics learning it was shown that Embodied Cognition does allow one to make more fine-grained predictions concerning the role of physicality. The researchers predicted based on Embodied Cognition that physical rather than virtual manipulatives would positively affect kindergartners' learning, but only for those who had an incorrect preconception of mass before the learning phase, such as the incorrect conception that a heavier object would go up on a beam balance. This was further explained that if a concept of mass and its effect on a balance beam is incorrectly or simply not instantiated in experience, additional physical experience becomes more important. Children were pre-assigned on the basis of whether they knew what a beam balance does to either the incorrect or correct preconception group, and were then further subdivided in a physical manipulative condition with real weights and balance beam vs. virtual manipulative condition in which children learned with a computer simulation of weights and beam balance. In line with the predictions it was found that only children with an incorrect conception of mass in relation to the beam balance showed learning gains from physical materials. This can be explained by the fact that children with correct preconceptions already had a good understanding of mass (grounded in previous haptic experiences), and therefore had no additional relevant information to gain from learning from physical rather than virtual manipulatives. These and other results seem to show that sensorimotor information can indeed be important for learning (Black et al., 2011; Morris, Tan, Barbagli, Chang, & Salisbury, 2007; Reiner, 1999; for an overview see Sigrist, Rauter, Riener, & Wolf, 2013).

As such, these examples show that Embodied Cognition would only predict that learning a particular concept through sensorimotor experience is important for those concepts that draw on that information for understanding of a particular concept. For example, mass must be grounded multi-modally simply because mass cannot be easily determined by the visual modality alone (e.g., two objects might look the same but vary in weight). This directly aligns with the finding that text comprehension is promoted by virtual manipulatives to the same extent as by physical ones (Glenberg et al., 2011). For example, we would only predict an effect of physical and interactive richness when the text involves information not readily attainable through the visual modality alone. An example of such a scenario could be a protagonist that has to choose between two treasure cases that look the same, but weigh differently or one that is locked and the other can be opened. Furthermore, when increasing the quantity or complexity of these visually unattainable features (providing that they are not already grounded in previous experiences; Zacharia et al., 2012) one would predict that physical manipulation becomes beneficial to text-comprehension.

Transfer by replacement

The final stream of research we present here seems to entail greater problems for Embodied Cognition. This research involves manipulatives that hold the "task of figuring out that one thing is intended to represent another, that the meaning and importance of the symbol lie in its relation to its referent." (Uttal et al., 2009, pp. 157; see also Uttal et al., 1997).

This line of research has shown that perceptually rich physical objects can actually hinder performances in cases where the manipulative stands-in-for something else (DeLoache 1987, 1991, 2000). In these studies children ranging from two to three years old have to obtain a toy hidden in a room. Children must do this by watching the experimenter hide a toy in a 3D reconstructed model of the room accompanied with the instruction that the real toy is hidden at the same place as in the model. It has consistently been found that children perform worse with 2d representations rather than perceptually rich and realistic 3D mock-ups at retrieving the toy in the real room (DeLoache, 1987, 1991). Furthermore, a glass plate put in front of the child - which prevents solicitations of acting on the model - actually improves inferential performance in contrast to a model that can be interacted with (DeLoache, 2000).

Although not about manipulatives directly, but often presented as relevant to the domain of manipulatives, are findings from studies that show that learning abstract (mathematical) relations and extending them onto novel but principally isomorphic situations is promoted when it is instantiated in a more abstract form as opposed to a concrete, or perceptually rich form (De Bock, Deprez, Van Dooren, Roelens, & Verschaffel, 2011; Goldstone & Sakamoto, 2003; Goldstone & Son, 2005; Johnson, Reisslein, & Reisslein, 2014; Kaminski et al., 2009; Kaminski, Sloutsky, & Heckler 2008; Kaminski, Sloutsky, & Heckler, 2013; Sloutsky et al., 2005). For example, although concrete (cupcakes) in comparison to abstract (circles) instantiations were better for learning a mathematical relation (fractions), transfer of learning was higher for kindergartners who learned with the arbitrary symbolic instantiations (Kaminski et al., 2009). In another well-known study of Kaminski and colleagues (2008) similar results were found showing that although concrete instantiation resulted in the highest performance on problem-solving, transfer of learning in which the same mathematical relation had to be deduced was hampered by concrete instantiations (also see Kaminski, Sloutsky, & Heckler, 2013). Importantly, although concrete-to-abstract might prove to be a leap too far when there is too much emphasis on the concrete, it has recently been shown that such a symbolic leap may sometimes best unfold in steps, fading concreteness into abstract forms (Fyfe, McNeil, Son, & Goldstone, 2014; Goldstone & Son, 2005; cf. Johnson et al., 2014; McNeil & Fyfe, 2012; Scheiter, Gerjets, & Shuh, 2010). For example, in a problem-solving task in which the proportion of different trees had to be discovered it was found that by gradually morphing realistically visualized trees into less detailed green squares led to reliable learning benefits as compared to a generic text based format (Scheiter et al., 2010).

According to the Dual Representation hypothesis (DeLoache, 2000), inhibited performance with concrete objects can be explained by the fact that subjects have to attain a dual representation: the concrete object in its own right, and its referent. Perceptual richness, therefore, may simply incline participants towards treating the situation as one single concrete instance, as one representation. Indeed for some researchers this explains why learning from some manipulatives (e.g., Dienes base-ten blocks; Resnick & Omanson, 1987) is notoriously difficult to translate into formalized forms (Uttal et al., 1997). Thus, it is suggested that manipulatives should be designed to be like symbols when they refer to some higher-order else, avoiding perceptually rich and real world characteristics (Uttal et al., 1997).

Intermediate Discussion: Embodied Cognition

In this section we made the case that transfer of learning does not necessarily rely on a concrete-to-abstract shift. We have presented three lines of research on transfer with manipulatives that seem to lead to different results on whether such a claim can be maintained.

Firstly, there are those situations in which thinking in the absence of manipulatives remains true to a sensorimotor format, which seems to be the case with mental calculation with abacus-trained users. We suggest that learning in this respect depends on the gradual internalization of sensorimotor routines. It is often gradual in that the learner slowly loses its dependence on external props (from full dependence of the environment, to projection with bodily resources [gestures], solely visual projection et cetera). As such, in lieu of the "concreteness fading" (Goldstone & Son, 2005), it can be argued that transfer of learning from manipulatives often involves gradual fading of the interaction with the environment, making place for internal sensorimotor simulations, the speed of internalization being dependent on the learners capability of having "representational stability" to stand in for external goings-on (Hutchins, 2005; Kirsh, 2009).

The second line of research, with evidence from science education and reading comprehension, showed that the tenet of Embodied Cognition that concepts are grounded in sensorimotor experiences has been implicitly and unduly interpreted as learning should benefit from grounding concepts and procedures in the kind of perceptual richness that unfolds in real-world practices (cf. T. De Jong et al., 2013; Klahr et al., 2007; Triona & Klahr, 2003; Triona et al., 2005). Moreover, the kind of assessment of learning in studies that we have cited above are not, and perhaps should not be, sensitive to the kind of perceptual richness-differences between physical and virtual environments that Embodied Cognition does acknowledge. Indeed, for Embodied Cognition learning from a physical laboratory or virtual laboratory would result in different lived experiences and as such would have different multi-modal associations when thinking about the learned context. However, for Embodied Cognition-driven experimental educational psychologists the challenge is to be sensitive to sensorimotor information that allows the Perceptual Symbol (i.e., concept) "to do its work" in the context of a task. Needless to say, physical and virtual manipulatives' properties vary deeply in the sensorimotor information they can provide. Important to note, is that much of the research on science education as discussed above has been agnostic to studying the affordances that come with Embedded Cognition which are undoubtedly relevant for the science education domain (e.g., ordering objects in 3-D space as to determine which procedure comes first; for example see Kastens, Liben, & Agrawal, 2008). In sum, this line of research seems to suggest that a more moderate reading of Embodied Cognition would be appropriate, wherein perceptual and interactive richness in and of itself is not something that promotes learning, but is contextually dependent on the learning content being constituted on multi-modal information.

The third stream of research seems to be on par with the moderate view that transfer of learning is hampered by perceptual richness (De Bock et al. 2011;

Goldstone & Sakamoto 2003; Goldstone & Son 2005; Kaminski et al., 2008, 2009, 2013; Sloutsky et al., 2005). Although such research might not fall in the domain of manipulatives (e.g., Kaminski et al., 2008), it has been argued that manipulatives have similar disadvantages if the adage of maximal perceptual richness is maintained (e.g., Uttal et al., 1997). We think this line of research cannot easily be dismissed and weakens in these particular cases the more simplistic reading of Embodied Cognition wherein *more* sensorimotor information *is better*.

Indeed we would suggest that with abstract learning goals we should treat manipulatives as what Andy Clark calls *Surrogate Situations* (Clark, 2005, 2008). In Surrogate Situations cognition is to some extent decontextualized from the environment since it goes beyond the immediate environment, but not disembedded[5], since the environment still provides a concrete surface that allows for deploying sensorimotor routines (e.g., just-in-time sensing; Ballard et al., 2001). Indeed, Nathan (2012) recently suggested that the research by Kaminski and colleagues (2008) does not show that interaction with materials hampers symbolic inferences (see however Deloache, 2010). Clark (2005) similarly argues that it is important to retain possibilities for interaction, but keep non-essential detail low as to avoid "gravitational pull" of sensorimotor distractions (e.g., automatic visual attention cues). For example, it has recently been shown with children who have to judge relations of sameness and difference are best able to do this when labels and objects are used that have an "optimal vagueness" (Son, Smith, Goldstone, & Leslie, 2012). Optimally vague to be recognized as something familiar but not too perceptually rich to avoid what Andy Clark might call the "gravitational pull of perception-action routines" (see also Markman & Gentner, 1997). That is, a vague or schematic as opposed to a concrete instantiation of objects that have a sameness relation are more easily generalized to other objects that share this sameness relation.

Thus we might speculate that manipulatives for abstract thinking should be considered as "manipulable symbols" that still allows for the affordances that are related to Embedded Cognition but are minimally rich in perceptual detail. Indeed, it has been found that even in highly symbolic environments learners draw on perceptual features, such as (self-induced) spacing in algebraic expression, that guide their problem-solving strategies (see Landy & Goldstone, 2007). For example, in an expression of 8 x 4 + 6, it is found in line with the syntactic structure that "8 x 4" is often written with less space between the symbols in

[5] Note that Clark (2005) uses "disembodied" here. We use disembedded as to consistently make a distinction between embeddedness and embodiment.

comparison to 4 + 6 as to denote a grouping order. As such Landy and Goldstone (2007, p. 2038) suggest that spatial relations in the algebraic expression serve "to ground the abstract relationships they express in more immediately available sensorimotor relationships". Interestingly, this use of space is highly similar to epistemic actions performed by Tetris players (Kirsh & Maglio, 1994) as spacing allows the task to be structured as to reduce computational load.

Conclusion

Most scientific discourse shows cyclical and reactionary patterns of progress - continually recycling and tempering theories in light of new findings. While early promoters of manipulatives, such as Montessori or Pestalozzi, held that unconstrained, self-guided, manipulation of physical objects would automatically impress complex ideas upon the mind (Page, 1991), in more recent literature such views are equated with "magical hopes" (Ball, 1992) or "folk psychology or vague theory" (Triona & Klahr, 2003, pp. 171). Indeed, research seems to indicate that more moderate claims about the role of perceptual and interactive richness are warranted, which has been important for furthering our understanding of learning with manipulatives (Brown et al., 2009; Kaminski et al.,, 2009; McNeil & Jarvin 2007; Sarama & Clements, 2009; Sherman & Bisanz, 2009; Uttal et al., 1997). However, in this paper we have in turn made the claim in light of Embedded Embodied Cognition that the current moderate view is also to some extent misguided if it is not negotiated with findings we have provided in this review.

The research reviewed here from an Embedded Cognition perspective firstly suggests that learners quite naturally draw on external support from the environment to alleviate cognitive load (e.g., Ballard et al., 2011; Kirsh & Maglio, 1994). Secondly, learners are affected by subtle changes in the environment that influence the ease of attaining information either internally or externally (e.g., Gray & Fu, 2004; Risko et al., 2013). Thirdly, embedded learning can be constrained by manipulatives that impose a certain course of action (e.g., Martin & Schwartz, 2005; Stull et al., 2012) whereas self-guided problem solving strategies can be effective, but seem to be moderated by the perception of possibilities for action on manipulatives (e.g., Antle, 2012; O'Malley et al., 2010; Stull et al., 2013).

However, it is not yet clear how manipulatives can be designed in such a way that these different processes are optimally supported, especially in relation to each other. In other words, based on the evidence reviewed here, it is not yet possible to derive clear instructional design guidelines. As such, one of the challenges for research on Embedded Cognition and manipulatives is to determine how perceptual and interactive properties alter both the way interaction can occur (hard constraints) as well as how these properties impinge on learners' likely course of action given the possibilities (soft constraints). Tangible User Interfaces seem well-suited for addressing such questions, as they provide a plethora of possibilities in maintaining physical interaction that can be related to perceptual properties in digital learning environments (Manches & O'Malley, 2012; O'Malley & Fraser, 2004; Shaer & Hornecker, 2010). Another important research question is how differing numbers of affordances that elicit external as opposed to internal learning strategies relate to long-term memory representations that are the source of transferring knowledge in the absence of manipulatives.

Current research reviewed here from an Embodied Cognition perspective seems to indicate that successful transfer of learning, in which the goal of manipulatives is to structure thinking in the absence of those manipulatives, does not necessarily involve decontextualization from perceptual and interactive constraints of manipulatives (e.g., Frank & Barner, 2012). In the research discussed here, it became clear that embedded interactions become embodied and aid in off-line thinking. We have further made the case that this often occurs gradually, wherein external support is faded out when expertise develops[6] and is dependent on the *internal* representational stability that the learner can maintain (e.g., Hatano et al., 1977; Hatano & Osawa, 1983; Kirsh 2009). As such an interesting prediction to be tested in future research would be that it is important for transfer of learning that learners have enough sensorimotor experience with a manipulative to be able to think without it. Interestingly, research seems to indicate that internal representational stability is promoted when interaction is easy (Flanagan & Boroditsky, 2013), suggesting that ease of manipulability affects ease of internalization.

Nevertheless, it has also become clear that whether Embedded Embodied Cognition can help make relevant predictions also depends on the learning goals and the assessment of whether these have been attained. In line with a moderate view, perceptual richness is not beneficial to learning when the assessed learning outcomes do not depend on multimodal information (e.g., Glenberg et al., 2011; Triona & Klahr, 2003). In fact it can be argued that much of the research reviewed here, actually shows that perceptual richness might hamper making abstract inferences (e.g., Kaminski et al., 2009). On a speculative note we have argued that manipulatives might still be important for learning abstract rela-

[6] Important to note, this depends on whether expertise is defined as a disembedded cognitive capability.

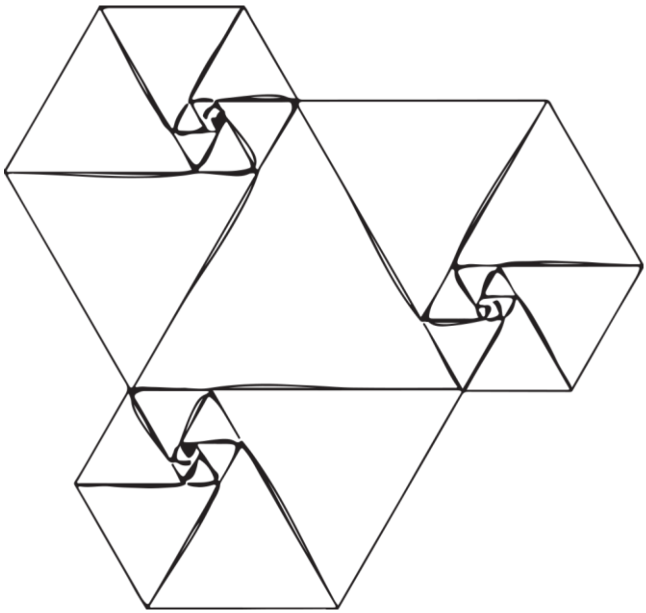
tions since they provide the learner with external support, and that current research should focus on how embedding learners in manipulable but not perceptually rich learning environments (i.e., surrogate situations).

Although theories of Embedded Embodied Cognition might be a suitable starting point for research on these open questions and enjoys empirical support to weaken moderate claims presented in the introduction, an important shortcoming of the current perspective in terms of educational implications is that given the current state of the literature it is difficult to provide guidelines for how manipulatives should be designed. Yet, an important educational implication we can take home from the Embedded Embodied perspective is that mouse-based virtual manipulatives, which reduce perceptual and interactive richness compared to physical manipulatives or tangible user interfaces, do not necessarily optimize the learner's cognitive load either. Furthermore, it has recently been argued by Nathan (2012) that research that shows that perceptually rich representations might not be suitable for bringing across abstract symbolic relations, should not lead educators to adopt the view that learning should go "without exposure to perceptually rich stimuli" since it "robs learners of opportunities to learn how to recognize deep structure and filter out irrelevancies" (Nathan, 2012, pp. 137). We would make a similar argument that educational design and research should focus on ways to expose learners to a range of interactive possibilities from which efficient externally mediated problem-solving strategies might arise.

To end with a theoretical note, the Embedded Embodied perspective, as opposed to a moderate view, attempts to provide an account of how the central aspect of manipulatives, that is, what sensorimotor information they provide, is beneficial for learning. Learning from manipulatives is always sensorimotor in nature - i.e., it always involves some degree of bodily interaction of the learner with the environment, if not, it ceases to be a manipulative. Indeed, when "subtracting" learning with manipulatives from learning with other materials such as texts or non-interactive instructional animations, we will always be left with perceptual and interactive richness as the key residual difference at the side of manipulatives. Thus, any perspective that seeks to guide instructional design of manipulatives should specify how the *body in action* affords processing of information not easily maintained with other learning materials and how this relates to long-term knowledge representations. In our opinion, the research reviewed here suggests that while more research is clearly necessary, the Embedded Embodied Cognition perspective provides a more promising starting point than a moderate view for furthering our understanding of how perceptual and interactive richness might aid learning.

Chapter 3

Does (non-)meaningful sensori-motor engagement promote learning with animated physical systems?*



*This chapter has been accepted for publication as Pouw, W. T. J. L., Eielts, E. C., Van Gog, T., Zwaan, R. A., & Paas, F. (in press). Does (non-)meaningful sensori-motor engagement promote learning with animated physical systems? *Mind, Brain & Education.*

Does (non-)meaningful sensori-motor engagement promote learning with animated physical systems?

Previous research indicates that sensori-motor experience with physical systems can have a positive effect on learning. However, it is not clear whether this effect is caused by mere bodily engagement or the intrinsically meaningful information that such interaction affords in performing the learning task. We investigated (*N* = 74), through the use of a Wii Balance Board, whether different forms of physical engagement that was either meaningfully, non-meaningfully or minimally related to the learning content would be beneficial (or detrimental) to learning about the workings of seesaws from instructional animations. The results were inconclusive, indicating that motoric competency on lever problem solving did not significantly differ between conditions, nor were response speed and transfer performance affected. These findings suggest that adult's implicit and explicit knowledge about physical systems is stable and not easily affected by (contradictory) sensori-motor experiences. Implications for embodied learning are discussed.

Introduction

How does practical experience with physical systems (e.g., gear systems, levers) affect learning about its mechanisms? For example, does our *experience* with riding a bicycle contribute to our understanding of the working of mechanical parts (e.g., gear systems) of a bicycle? The answer to this question is relevant for educational practices in science education, as it dictates whether learning about physical systems should be grounded in concrete physical experiences next to abstract formalisms (Nathan, 2012; Pouw, Van Gog, & Paas, 2014). According to embodied learning theories, understanding of abstract principles relies upon the structural relations that emerge in bodily interaction with the environment (e.g., Goldstone & Barsalou, 1998; Lakoff & Nunez, 2000; Pouw et al., 2014). If this is correct, effective design of digital learning environments at times involves providing possibilities for bodily interaction.

Previous research indeed indicates that bodily interaction while learning or working with physical systems may in some cases promote understanding (e.g., Han & Black, 2011; Schönborn, Bivall, & Tibell, 2011; Zacharia Loizou, & Papaevripidou, 2012). Such findings may prove informative for guiding applications of computer-based technology in education, such as Tangible User Interfaces (TUIs; e.g., Manches & O'Malley, 2012; Marshall, Price, & Rogers, 2003). TUIs are characterized by the combination of physical and virtual objects, running in real-time, and allowing for physical interactions between the users and virtual objects that are typically afforded by interactions with real non-virtual objects (Daponte, De Vito, Picariello, & Riccio, 2014). For example, the Nintendo Wii Balance Board can be used for continuous full-body *physical* interaction with *virtual* objects that can simulate complex affordances with non-virtual objects (e.g., snowboarding). However, it is as yet unclear whether positive effects of bodily interactions with physical systems on understanding are promoted by the particular structural relations between agent and environment that emerge during physical interaction, or by the motivational processes that are affected by physical engagement (e.g., Bivall, Ainsworth, & Tibell, 2011; Han & Black, 2011; Wiebe, Minogue, Jones, Cowley, & Krebs, 2009).

In this study participants learned about a class 1 lever (a lever where the fulcrum is in the middle and the effort and resistance on opposite sides; e.g., a seesaw) through physical engagement with a Nintendo Wii-Balance Board (hereafter: Wii Board)[1]. Their physical engagement with the Wii Board was either minimally, meaningfully, or non-meaningfully related to the underlying principles of the physical system. This allows for studying not only *whether*, but also *how* bodily interaction supports learning of mechanical concepts.

Physical Engagement and Learning

There is increasing empirical evidence that physical engagement with learning materials can be an effective learning practice in for example mathematics, reading comprehension, and science education (e.g., Fyfe, Mcneil, Son, & Goldsone, 2014; Glenberg, Gutierrez, Levin, Japuntich, & Kaschak, 2000; Kiefer & Trump, 2012). This research is informed by theories of embodied learning which suggest that learning and applying knowledge involves the effective re-use, simulation, or reactivation of sensori-motor experiences (for an overview see Pouw et al., 2014).

Strong evidence for embodied learning comes from the field of gesture research, which has shown that actively producing, imitating, or enacting gestures during word-learning (and retrieval) enhances memory (-retrieval) as opposed to more passive control conditions (e.g., Kelly, McDevitt, & Esch, 2009; Macedonia & Klimesch, 2014; for an overview see Macedonia & Kriegstein, 2012). These findings are explained by the idea that the use of gestures during word-learning enriches the conceptual understanding with multi-modal information. This enrichment of the conceptual understanding is held to consist of a higher degree of associations with the concept's relevant modality-specific information (motor, haptic, spatial, etc.), which aids in the prevention of memory-decay and the retrievability of the concept (e.g., Macedonia & Kriegstein, 2012). Going beyond word-learning, it has been found that gesturing (vs. not gesturing) during learning of science-related texts improves learners' ability to make inferences about the learning content (Cutica & Bucciarelli, 2013; Cutica, Iani, & Bucciarelli, 2014).

A deeper analysis of the gesture literature shows that gesturing during learning is not effective merely by virtue of activating the sensori-motor system, but the meaningfulness of gestures seems to be important too. For example, it has been found that gesturing during word-learning is only beneficial to memory when these gestures bear an iconic relation with the meaning of the word (e.g., moving hand up and down to depict 'hammering') as opposed to gestures consisting of movements that are not concretely related to the semantic content of the word (e.g., Macedonia & Knösche, 2011; Kelly et al., 2009; see also Cook, Yip, & Goldin-Meadow, 2012). Thus, the gesture literature suggests that bodily activity might only aid learning when it is meaningfully related to the learning content.

Embodied Learning and Science Education

We are interested in how active bodily activity might aid learning of principles underlying physical systems, a central learning topic in science education. To date, there are only few quantitative experimental studies on the precise role of bodily engagement in this context (e.g., Bivall, Ainsworth, & Tibell, 2011; Han & Black, 2011; Johnson-Glenberg et al., 2014; Olympiou & Zacharia 2012; Kontra, Lyons, Fischer, & Beilock, 2015; Schönborn, Bivall, & Tibell, 2011; Triona & Klahr 2003; Wiebe et al., 2009; Zacharia & Constantinou, 2008; Zacharia, Loizou, & Papaevripidou, 2012; Zacharia & Olympiou, 2011). Next we will only address findings that focus on concepts such as force and mass as these are central concepts for understanding the dynamics of a class 1 lever.

A demonstration of the benefit of recruiting sensori-motor processes in science concept learning is offered by a study of Bivall and colleagues (2011). They show that conceptual understanding of the structure of a bio-molecular model improved when learners were offered haptic feedback during the training-phase. More precisely, learners were either engaging with a haptic device that simulated the repulsive and attractive forces of the molecules, or engaging with the same haptic device but with the haptic force-feedback disabled. Engaging with the haptic device with force feedback bolstered the learning outcomes from pre- to post-test (for similar results see Schönborn et al., 2011). It was suggested that haptic feedback during instruction afforded learners the opportunity to off-load visual working memory onto the sensori-motor system (Bival et al., 2011; Schönborn et al., 2011). Furthermore, haptic feedback provided the learner with information about repellant and attractive forces *directly*, while that had to be visually inferred in the no-haptic feedback condition.

In a comparable study, fifth-graders learned the workings of simple mechanical gear devices through different degrees of sensori-motor engagement (Han & Black, 2011). Subjects in the control condition observed the unfolding of the simulation, whereas in the other two conditions participants controlled the spinning of the gears with a joystick (kinesthetic condition); in the third condition the joystick control was augmented with force-feedback (force-kinesthetic). Participants in the two kinesthetic conditions shower higher learning gains than participants in the control condition. Han and Black (2011) suggested that the kinesthetic experience allowed participants to re-enact the relevant haptic information related to force as to actively compare it to visual information presented during the task.

In another study concerning workings of levers it was investigated whether providing learners (ages 11-14) with haptic feedback during training would benefit learning performances as opposed to learners who only received visual information during training (Wiebe et al., 2009). In the first part of the training participants set up the position of the lever's fulcrum and applied a number of weights. Subsequently the program generated a second lever (with different position of the fulcrum and different number of weights). Participants were to judge which of the two levers (self-constructed vs. program-generated) required the highest amount of force to lift the weights. The participants in the haptic condition were also allowed to "feel" the amount of weight needed to lift the weights by using a device which produced haptic feedback. It was found that learning performance in terms of declarative or conceptual knowledge did not differ between the haptic and visual condition. In fact, participants in the visual condition outperformed those in the haptic condition in judging which lever required the highest amount of force.

There are indications, however, that the learning impact of physical engagement with objects or interfaces might be dependent on prior knowledge. For instance, in a study by Zacharia and colleagues (2012) kindergartners learned about the role of mass and its effects on a balance beam (class 1 lever) by either physically interacting with a balance beam or a virtual equivalent programmed on a computer. Prior to the training it was assessed whether children already possessed the correct conception that heavier objects placed on one side will pivot the balance beam. It was found that only children with an incorrect preconception benefited in terms of learning outcomes from physically interacting with a balance beam. This finding suggests that if learners already have an understanding of how mass relates to the balance beam they can assess mass based on perception alone and no additional sensori-motor information is needed to allow them to perceive mass directly (through kinesthetic feedback).

The previous results suggest that sensori-motor activity can be beneficial to learning underlying principles of physical systems. However, it should be noted that some of these studies did not find beneficial effects (Wiebe et al., 2009), and that some of the studies were very low powered (Bivall et al., 2011; Schönborn et al., 2011; Wiebe et al., 2009) to moderately powered (Zacharia et al., 2012); the study by Han and Black (2011) was an exception, it included a high number of participants. Even if we sidestep the issue of robustness of some of the previous findings, the design of the previous studies cannot rule out that bodily engagement only affects the motivational processes of the learner (e.g., Jones et al., 2006). Most studies leave open the possibility that sensori-motor activity affects learning performance indirectly through affecting motivation and experiences of immersion, instead of by providing *meaningful* information about the learning content. As research on gesture and learning shows, it seems likely that only meaningful physical engagement would promote learning, but it cannot be ruled out that it is the physical engagement as such (and the structural relations that are picked up) had indeed benefited learning in those studies. Therefore, the aim of the present study was to address *how* enriching the learning content with sensori-motor information affects learning.

The Present Study

To study *how* enriching the learning content with sensori-motor information affects learning, we manipulated the meaningfulness of the structural relations between physical actions on the Wii Board and the instructional animations of the learning content (i.e., mechanical principles of class 1 levers; a seesaw). Participants were assigned to one of three conditions. In the first, subjects were given a meaningful embodied training, in which they learned to balance a seesaw across several trials, by applying force on the Wii Board that matched the number and position of the weights that acted as counterforce on the seesaw (*meaningful condition)*. In the second condition, subjects were given a similar training, but in this condition the forces that needed to be applied on the Wii Board to balance the seesaw were non-meaningfully correlated with the number and position of the weights that acted as counterforce (*non-meaningful condition*). Yet, in this condition participants did apply force on the congruent side of the seesaw. Thus while participants pushed a seesaw down on the congruent side, the force needed to push the seesaw into balance was not consistently (i.e., non-meaningfully) related with the number of weights placed on the seesaw. These conditions were compared with a third, *minimal condition,* in which participants merely provided a small push that started an animation of a seesaw balancing out. Thus, importantly, participants in all three conditions are using the Wii-Board to interact with the instructional animation, which allows us to eliminate some of the motivational effects on performance that might arise from the use of the Wii-Board and from having the animation respond to an action by the learner.

Not only do we explore whether meaningful physical experiences may support learning, we also assess whether non-meaningful physical experiences (i.e., acting with incorrect relations with the learning principle) hamper learning. After all, if knowledge is indeed grounded in action as embodied theories of learning have it, then we might also predict the opposite, namely physical experiences that are incongruent with the learning principle should hamper learning. This is a novel question that allows us to further gauge the degree to which knowledge of physical systems (i.e., levers) is affected by sensori-motor experience.

To assess the broad aspects of learning afforded by sensori-motor interaction, we used three different performance measures: a reaction-time task (henceforth RT-Task), a transfer task, and a motor task. The RT-task relied heavily on visual perceptual experiences; it assessed speed and accuracy of judging whether a depicted seesaw should balance out or pivot given the weights and their position. The transfer task relied more on deliberate reasoning; it measured the accuracy of judgments about more complex class I lever concepts (e.g., interconnecting seesaws, varied fulcrum positions). The motor task relied purely on motoric knowledge; it assessed whether participants were able to physically enact the correct amount of force to balance a seesaw when provided with a non-interactive picture of a seesaw.

This motor task provides us with a novel and exploratory way to assess whether knowledge of mechanical systems can be partly assessed in the way subjects enact the solution of the problem as opposed to tasks that are procedurally very different in nature (pushing a button; i.e., RT-task and the transfer task). Essentially, it allows us to assess whether our learning manipulations differentially affect whether participants *know how to* physically balance a seesaw (motor task) as opposed to *knowing that* a seesaw balances out under particular conditions (RT- and transfer task; Ryle, 1945).

To assess cognitive load and motivation differences, we also included subjective attitudes (mental effort, interest, difficulty) towards the learning-phase and test-phases as to check for possible mediating effects of motivation (interest) and experienced difficulty. Participants' reports of the interest of the learning phase are of special concern to the present study, as it provides a way to assess whether there were motivational differences across conditions.

We hypothesized that participants in the meaningful condition would outperform participants in the minimal and non-meaningful condition on all performance tests. We also hypothesized that the non-meaningfully embodied instructional animation (i.e., non-meaningful condition) would actually hinder performance on these tasks as compared to the other conditions, as it provides interfering sensori-motor information.

Method

Participants and Design

A total of 92 Dutch university students participated in the present study for course credit or 10 euros. Unfortunately, due to a programming error for 15 participants the Wii-Board data were lost (meaningful = 4, non-meaningful = 5, minimal = 6). Additionally, one participant (non-meaningful) was excluded from the analyses for not following the instructions correctly (participant employed two hands instead of one to push on one side of the WiiBoard). This resulted in data of 76 participants for the analyses (37 males [48.68%]; age range = 18 to 25, *M* = 21.32, *SD* = 2.112; 93.4% right handed, as determined by Oldfield, 1971), who were randomly distributed among 3 conditions in a between-subjects design: meaningful (*N*= 26), non-meaningful (*N* = 25) or minimal (*N* = 25).

Materials

Instructional animations

The voice-over and textual instructions and self-report questions were programmed in ActionScript 3.0 and the animations were designed in Adobe Flash Professional CS 5.5 (see http://www.charlyeielts.nl/wbb/materials.html or https://osf.io/ebjvm/). The Wii Board communication was handled by the WiiFlash Actionscript API and WiiFlash Server developed by Joa Ebert and Thibault Imbert (http://wiiflash.bytearray.org/).

Prior to this study we assessed whether adults were affected in performance in one of our main learning measures (reaction time task) by comparing the effect of only observing the instructional animation as opposed to receiving no instructional animation. This was to ensure that adults are still receptive to training about class-1 levers. In this pilot-study with adults (*N* = 78; 52.6% female; Age *M* = 33.47, *SD* = 12.29, with 83.4 % reporting having had college experience) using Amazon's Mechanical Turk we used the exact instructional materials designed for this study but without possibilities for physical interaction. This pilst showed that the animations were effective for learning (57.24 %, *SD* = 19.4% accuracy on the reaction-time task) as compared to no instruction (69.26%, *SD* = 20.4%), t (76) = -2.644, p = .010, Cohen's d = .602 [large effect]). No effects were obtained for solving speed on the RT-task, t (76) = -0.945, p = .348, Cohen's d = 0.218.

Introductory instructional animation

Before the manipulation phase, each participant viewed a short non-interactive instructional animation of 190 seconds with a Dutch female voice-over, in which the different concepts involved in the operation of a lever were introduced (introduction phase). The introduction phase presented the seesaw and its components (fulcrum, left arm, right arm), and the concepts of load, force and balance. This introduction phase further focused on the mechanical principle of levers. The mechanical advantage principle explained in this animation involved the concept that force can be amplified by increasing the distance from the fulcrum.

Manipulation: Interactive instructional animation

In the manipulation phase, participants had to perform 24 interactive study trials in which they had to return a tilted seesaw to a state of balance using the Wii Board[7].

Before each of these trials a fixation cross was displayed and subjects were instructed not to apply any force on the Wii Board. The experiment would automatically start when the subjects employed force that did not deviate more than 0.2 lbs from the calibration values for longer than 500 ms, which ensured that every study trial started from a rest position. At the beginning of each trial a seesaw was presented that could be divided into 9 even sized parts with the fulcrum placed in the middle. In each trial, the seesaw was either tilted left or right with one weight (either small [one cube] or large [two cubes]) placed on one side of the seesaw. In half of the trials a load of one blue cube was tilting the seesaw and in the other half, a larger load of two blue cubes stacked on top of each other tilted the seesaw. The animation was designed such that the large weight was exactly two times the volume of the small weight. Participants were instructed to return the seesaw to balance by employing a required amount of force on

[7] Before the experiment the participants engaged in a calibration session: After a 3 second countdown, pressure data of sensors on both sides were recorded for the duration of 5 seconds at a sampling rate of 60 Hz. For each side of the Wii Board, the average recorded force on that side was subtracted from force values resulting in a new calibrated force value for each sensor (minus the weight of the hands in rest-state). This ensured that the interface only reacted when participants actively engaged with the Wii Board.

the opposite arm of the seesaw. The location on the seesaw where the required force should be applied was marked by a yellow highlight around the edges of the area. When subjects applied the correct amount of pressure on the Wii Board, the seesaw would react and the arm carrying the counter-weight would be lifted from the ground and the cube representing the participant's administered force would grow to the correct size to establish balance. The animation would stop if a state of balance was reached.

The required force to balance a seesaw differed across conditions. For the meaningful condition, the required pressure for small weights was 5 lbs, with a range of 4 to 6 lbs and 10 lbs for large weights with a range of 9 lbs to 11 lbs. In the non-meaningful condition, the force requirements of 5 lbs and 10 lbs were *randomized* for the small and large weights, so that there was no structural correlation between amount of weight and counterweight to achieve balance across trials. In both the meaningful and non-meaningful condition, the seesaw would go out of balance at the force side if the upper bound of the accepted range of employed force was exceeded. If the applied force was lower than the required minimum, the seesaw would return to its initial state. In the minimal condition, the animation would simply play if participants prompted it to start by shortly applying a small amount of pressure (>0.3 lbs) on both sides of the Wii Board. Importantly, when the seesaw was in balance, participants in the meaningful and non-meaningful motor conditions had to continue employing the appropriate amount of force for 2 seconds before the experiment proceeded to the next trial. In the minimal condition, the experiment would automatically proceed to the next trial 2 seconds after the seesaw reached a state of balance.

Test Tasks

Reaction-time task

We developed a three-choice reaction-time task programmed in E-Prime (henceforth RT-task) to assess participants' accuracy (number of correct responses) and efficiency (reaction-time) in assessing class 1 lever's mechanics. In this RT-task participants were shown a seesaw that was either in balance or tilted to the left or right. In each trial one or two blocks are presented on each side of the seesaw on deferring distances from the fulcrum. The size and location of the weights varied across the 45 trials. Subjects had to determine which way the seesaw should be tilted given the presented weights, regardless of the current state of the seesaw (i.e., pivoted to left/right or balanced). Subjects responded with a keyboard by pressing "P" if the seesaw should be tilted to the right, "Q" if it should be tilted to the left and SPACE if the seesaw should be in balance. Subjects were instructed to respond as fast as possible. Thirty-two trials of the forty-five consisted of a situation where the principle of mechanical advantage was relevant, meaning a weight was closer or further from the fulcrum than the opposite weight (see Figure 1 for an example).

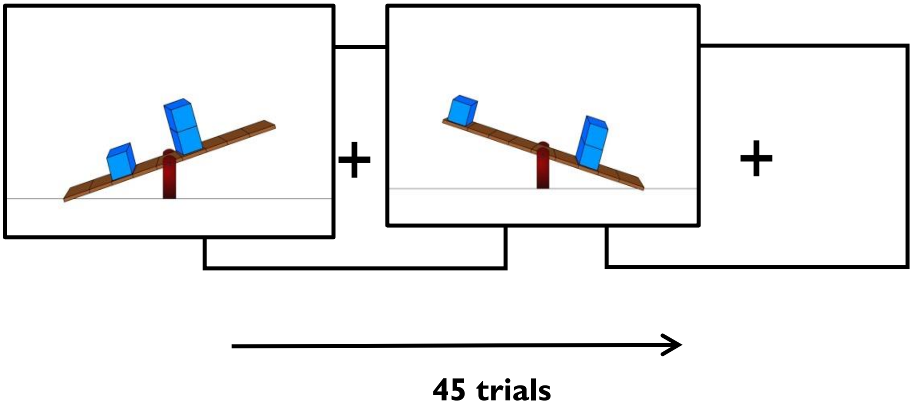


Figure 1. Example of two reaction time trials, after each response a fixation cross would appear intermittently.

Transfer task

The transfer task consisted of a total of 12 trials consisting of a 3-choice judgment task. Participants were prompted to think as long as they needed to produce the correct answer. The trials required participants to judge whether a seesaw in a set of several interconnected see-saws and differing positions of the seesaws' fulcrum, would pivot to the right, to the left or would stay balanced (see Figure 2). Also 4 trials involved the judgment of the amount of force needed to balance two seesaws in which participants had to judge which arm of the avatar needed to exert the most amount of force to balance the seesaws.

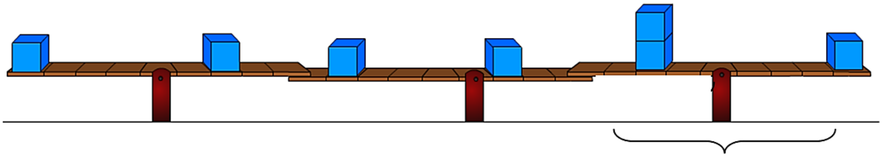


*Figure 2***.** Example of one transfer task trial. Participants were asked to "Judge whether the right seesaw ('rechterwip') pivots to the left, remains in balance, or pivots to the right".

Motor task

In the *motor* task participants had to determine the amount of force that needed to be employed to balance a seesaw that was statically presented, by applying the force on the Wii Board (22 trials). These trials were identical to the practice trials in the study phase, with the exception that the seesaw could not be controlled via the Wii Board (i.e., it remained a static picture). During each trial participants would employ force for 1500 ms. During the trials we assessed the amount of weight applied by the participant over a 1000 ms period (sampling: 60Hz) for 22 trials and sampled force-values from 500 ms onward (thus 500 to 1500 ms). We decided not to sample the first 500 ms of force-employment as we were interested in the moment participants reached a stable force-employment. This allowed us to gauge participants' ability to correctly judge the different levels of force that should be employed for balancing the seesaw.

Self-Report Questions

Mental effort, difficulty, interest

As an indication of experienced mental effort, perceived difficulty, and experienced interest during the learning phase and after each of the test phases (RT-task, transfer task, motor memory task) participants answered on a 5-point scale "How much mental effort did you invest during the learning phase [or RT-task, transfer task, motor-Task]" (mental effort; 1 = 'very low mental effort', to 5 = 'very high mental effort'), "How difficult did you find this task" (difficulty; 1 = 'not difficult, to 5 = 'highly difficult'), and "How interesting did you find this task" (interest; 1 = 'not interesting', to 5 = 'very interesting').

Physical effort

Amount of physical effort invested in completing the interactive animation trials ("how much physical effort did you exert during this task" on a 5-point scale (1 = 'very low effort' to 5 = 'very high effort').

Handedness

Using a modification of the Oldfield (1971) Handedness questionnaire participants reported hand-dominance for several manipulative situations (e.g., writing, brushing teeth, etc.) on a 5-point scale (1 = very left hand dominant, to 5 = very right-hand dominant). We computed the mean responses and categorized left (right) handedness for means lower (higher) than 3.

Prior knowledge self-report

Prior knowledge of the learning material ("Before this experiment I was knowledgeable about levers"; 1 = not knowledgeable, to 5 = very knowledgeable). We also checked whether participants had a physics background obtained in secondary school (0 = no, 1 = yes).

Demographics

Participants reported after the experiment their age, gender, and study program, and were allowed to comment on the nature of the experiment.

Procedure

Participants were informed that they would start with a training phase and would subsequently perform several learning-tests. First, participants were seated at a table on which a Wii Board was mounted. The chair's height and elbow supports were adjusted such that participants' hands rested on the Wii-board and the elbows had a 90-degree angle. The Wii Board was first calibrated to the participant's resting state. During the calibration procedure, participants positioned their left and right hand on the corresponding side of the Wii Board with their hand placed on a marker that represented the location of the pressure sensors in the Wii Board for either side. It was stressed that the subjects should only rest their hands on the Wii Board and should not apply any pressure on the Wii Board during the calibration. In order to familiarize participants with the Wii Board controls, they performed a short training sequence before the experiment. The training consisted of four different cubes that changed color if the participant gave the correct amount of pressure on the Wii Board. The different levels of pressure corresponded to the ones used in the experimental procedure. If the pressure exceeded the required force, the color of the cube would overflow and participants were instructed to apply less pressure. Participants then watched the introductory instructional animation and subsequently proceeded to the interactive instructional animation, which they interacted with in a meaningful-, non-meaningful, or minimal manner depending on their assigned condition. During the training-phase, the study time (time needed to balance the seesaw in the 24 trials) was recorded by the software, as this was likely to vary across conditions as a result of the experimental manipulation. After this interactive training-phase participants reported exerted physical effort. Subsequently participants performed the RT-task, Transfer task, and Motor task (in that order). After the training phase and test phases participants reported their experienced mental effort, difficulty of task, and interest in task. Finally, participants answered questions concerning gender, age, prior-knowledge, handedness, physics-background as reported above.

Data Analyses

Accuracy and RT-scores for the transfer task and RT-task lying outside 2.5 *SD* of the overall-mean were replaced with the overall mean (will be reported in the results if applicable).

Reaction-time task

The number of correct answers on 45 trials (performance range: 0-45) was taken as a measure of accuracy and the mean reaction-time (in ms) on correct trials as a measure of speed.

Transfer task

The number of correct answers on 12 trials was taken (performance range: 0-12).

Motor task

We obtained two outcome measures from the motor task. Firstly we provide the different trajectories for the applied force during 1000 ms for the two different levels of force (one cube vs. two cubes to balance a seesaw). This should give us exploratory information about whether the conditions indeed performed differently; as can be expected since participants learned to balance a seesaw with diferring weights. Since we are interested whether participants' motor-performance reflects understanding of the mechanics of a seesaw we used an additional *ratio-measure* which reflects whether participants could correctly differentiate between one versus two cubes, that is, one cube should be half the force of two cubes. This was done by dividing the mean amount of force given for one-cube trials (11 trials) by the mean amount of force for two-cube trials (11 trials); when participants indeed were able to correctly differentiate between one versus two cubes the ratio would give ½ value (i.e., .5). The final measure is therefore the absolute difference of the correct ratio of .5 and the ratio attained by the participants; $[\frac{\text{Mean Force Cube 1 Trials}}{\text{Mean Force Cube 2 Trials}} - .5]$; this yields .0 as a perfect score (i.e., lower score is better).

Unfortunately, due to technical issues we failed to administer Wii-board data for this particular task for an additional 7 participants, yielding a sample of 68 participants (meaningful [*N* = 23] vs. non-meaningful [*N* = 24] vs. minimal condition [*N* = 21]).

Results

Prior knowledge and physics background

No significant differences (see Table 1 for means) were found across conditions (ANOVA) for prior knowledge, $F(2, 73) = 2.27$, $p = .110$. This was also the case for physics background, $F(2, 73) = 1.078$, $p = .346$.

Wii-board training phase

Training duration

The duration of the training phase differed, such that the non-meaningful condition (*M* = 184.50 seconds, *SD* = 107.15) was longer in duration than the meaningful condition (*M* = 133.79 seconds, *SD* = 27.56) and the minimal condition (*M* = 95.12 seconds, *SD* = 73.26). As is evident, variances were not equal across groups (Levene's $\alpha < .001$). To test whether the differences in training-phase duration were significant we performed a Kruskal-Wallis analysis with pairwise comparisons. There was a significant overall effect of condition on duration, $\chi2(2) = 37.519$, $p < .001$. Pairwise comparison showed that the non-meaningful condition and the meaningful condition took longer than the minimal condition (minimal vs. non-meaningful condition, $\chi2\ [1] = 5.991$, $p < .001$; minimal- vs. meaningful condition, $\chi2[2] = 4.126$, $p < .00$). However, the meaningful condition did not differ from the non-meaningful condition, $\chi2(1) = -1.924$, $p = .163$.

Condition		Training Phase		Reaction-Time Task		Transfer Task		Motor Task		Prior Knowledge		Physics Background	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Meaningful condition	Mental Effort	3.15	0.68	3.19	0.85	3.73	0.92	2.73	1.22				
	Difficulty	1.96	0.92	3.73	0.72	3.88	0.86	3.73	1.185	3.27	1.19	.231	0.07
	Interest	1.62	0.85	3.62	0.80	3.81	0.80	3.35	0.80				
Non-meaningful condition	Mental Effort	2.80	0.96	3.32	0.95	3.80	0.76	2.72	1.17				
	Difficulty	2.68	1.03	3.56	0.96	3.76	1.05	3.16	1.179	2.60	1.12	.160	.07
	Interest	1.64	.64	3.28	0.84	3.81	0.81	2.96	0.98				
Minimal condition	Mental Effort	2.68	0.99	3.04	0.94	3.28	1.06	2.20	0.96				
	Difficulty	1.48	0.65	3.56	1.08	3.56	1.08	3.20	1.291	2.80	1.16	.08	.07
	Interest	1.36	0.70	3.52	0.77	3.84	0.69	2.96	0.98				

Table 1. Means and standard deviations for mental effort, difficulty, and interest for training phase and performance tasks across conditions.

Task load

See Table 1, column 1 for the means and standard deviations for the reported mental effort, difficulty, interest for the training phase across conditions[8]. One-way ANOVAs only showed a significant effect of condition on difficulty, $F(2, 73) = 11.754$, $p < .001$, $\eta_p^2 = .24$. Post-hoc comparisons (Bonferroni) showed that the non-meaningful training-phase ($M = 2.68$) was reported to be significantly more difficult than the minimal training-phase ($M_{difference} = -1.20$, $p < .001$), and the the meaningful training-phase ($M_{difference} = -.72$, $p = .014$). The meaningful training phase did not differ on difficulty from the minimal training phase ($M_{difference} = -.48$, $p = .164$).

Physical effort

Reported physical effort during the training-phase in the minimal condition ($M = 1.88$, $SD = 0.93$) the meaningful ($M = 2.35$, $SD = 1.06$), and the non-meaningful ($M = 2.56$, $SD = 1.16$) did not differ, $F(2, 75) = 2.738$, $p = .071$).

RT-task

Accuracy

We replaced outliers (outside 2.5 *SD* range from the mean) with the overall mean ($n = 2$). Overall accuracy was 80.04% ($M = 36.02$ correct responses out of 45, $SD = 3.26$), with the meaningful condition scoring 80.05% ($M = 36.23$ [/45], $SD = 2.83$), the non-meaningful condition scoring 79.56% ($M = 35.80$ [/45], $SD = 3.30$), and the minimal condition scoring 80.00% ($M = 36.00$, $SD = 3.73$); also see Figure 3a. A one-way ANOVA yielded no significant differences across conditions, $F(2, 75) = .109$, $p = .897$, $\eta_p^2 = .003$. An additional Bayesian analysis for the effect of motor-involvement condition on accuracy yielded$p_{BIC}(H_0|D) = .986$ (Masson, 2011). This probability indicates a 98.6% likelihood that motor-involvement condition (meaningful, non-meaningful, minimal) does not affect accuracy on the RT-task. Following guidelines by Kass and Raftery (1995), this information criterion is strong evidence for the absence of an effect of motor-involvement on accuracy.

Furthermore, in our pilot study with adults on Mechanical Turk, those participants ($N = 43$) who did not view an instructional animation had a considerably lower accuracy score on the RT-task (57.24 %) than participants (N = 35) who did view animations (but without opportunities for interaction) in the pilot study (69.26 %) and participants in the present study where overall accuracy was

---

[8] These analyses were performed on the complete sample of 90 participants, as they did not include Wii-Board data of the training phase.

80.04%. In sum, the pilot study data suggest that the instructional animations used here contribute to learning.

Reaction Time

No outliers outside the 2.5 SD range from the mean were found. The average reaction-time in *ms* for correct trials (see Figure 3b) for the meaningful condition (M = 2152.83, SD = 489.45), the minimal condition (M = 2396.02, SD = 857.29) and the non-meaningful condition (M = 2717.07, SD = 1202.200) showed unequal variances across condition (Levene's $\alpha < .001$). We performed a Kruskal-Wallis analysis with pairwise comparisons which yielded no significant overall effect of condition on reaction-times, $\chi2(2) = 1.860$, $p = .395$.

Task load

See Table 1, column 2 for the means and standard deviations for the reported mental effort, difficulty, and interest on the reaction-time task across conditions. ANOVAs showed no significant overall main effects of condition on these self-report measures regarding the RT-task. Furthermore, there were no significant correlations of self-report measures with performance on the RT-task.

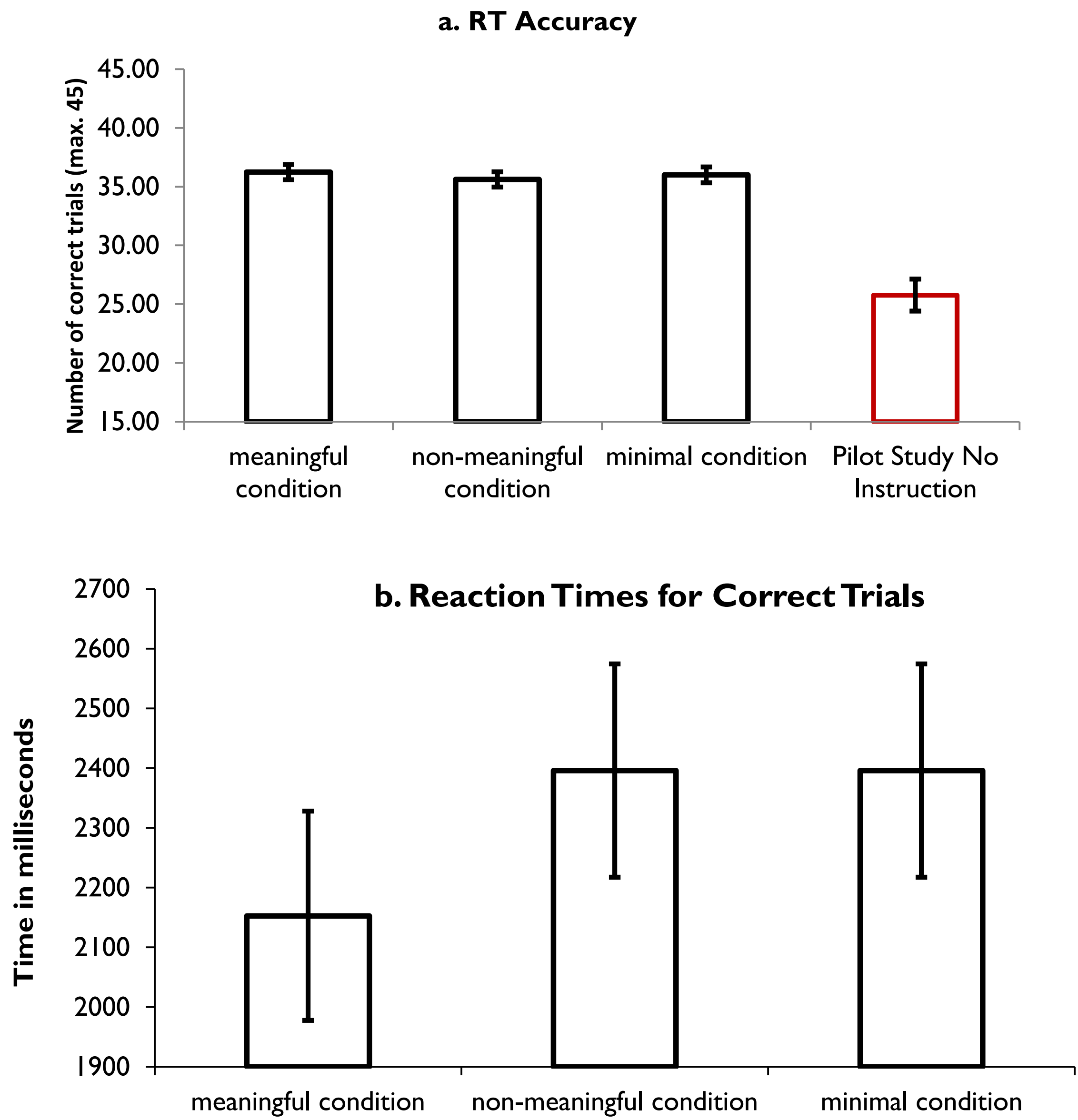


Figure 3a, b. Number of correct trials and reaction times for correct trials (ms) for the meaningful, non-meaningful, and minimal condition. Error-bars indicate standard errors. For number of correct trials (RT-accuracy, Figure 3a) we have added the results of a pilot study performed on Mechanical Turk where participants did not receive instructions.

Transfer Task

Accuracy

No outliers outside 2.5 SD range from the mean were found. Overall mean accuracy (see Figure 4) was 52% (i.e., a mean of 6.24 correct responses out of 12, *SD* = 1.6); meaningful (*M* = 6.54 [%], *SD* = 1.363), non-meaningful (*M* = 5.84 [50.00%], *SD* = 1.625), and minimal condition (*M* = 6.32 [50.00%], *SD* = 1.77). A one-way ANOVA yielded no significant differences across condition, $F(2, 73) = 3.24$, $p = .258$, partial $\eta_p^2 = .034$. An additional Bayesian analysis for the effect of motor-involvement condition on transfer task accuracy yielded $p_{BIC}(H_0|D) = .954$ (Masson, 2011). This probability indicates a 95.4% likelihood that motor-involvement condition (meaningful, non-neaningful, minimal) does not affect accuracy on the transfer task; which can be considered strong evidence for the absence of an effect of condition (Kass & Raftery, 1995).

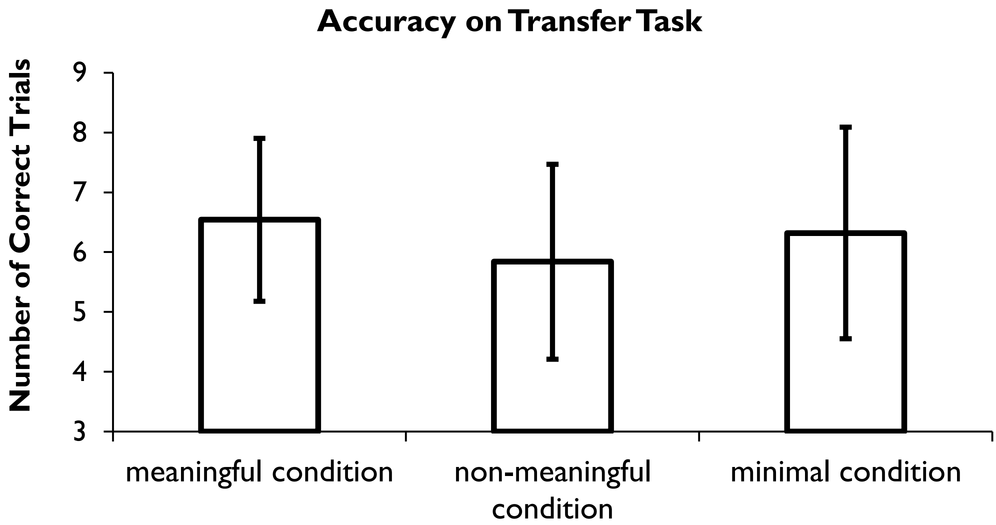


Figure 4. Number of correct responses on the transfer task for the meaningful, non-meaningful, and minimal condition. Error-bars indicate standard errors.

Task Load

See Table 1, column 3 for the means and standard deviations for the reported mental effort, difficulty, and interest on the transfer task across conditions. ANOVAs showed no significant overall main effects of condition on these self-report measures regarding the transfer-task. With regard to overall correlations between self-report measures and performance on the transfer task we only found a significant correlation between experienced interest and accuracy on the transfer task, such that more reported interest resulted in higher performance, $r = .292$, $p = .011$.

Motor-memory Task

Figure 5 shows the mean force-responses for one-cube (grey) and two cube trials (lock) plotted over time (1 second) with 95% CI's. As can be qualitatively inferred from the graphs, there is a considerable difference in participants motor-responses in the minimal condition as compared other conditions. This is not surprising as participants in the minimal condition were trained to give a short push on the Wii-Board to balance the seesaw, which is reflected in this task as well. Namely, participants in the minimal condition gave a short but large force response as compared to the other conditions.

However a more interesting pattern seems to have emerged if we consider that only participants in the meaningful condition were trained to motorically differentiate between forces of one vs. two blocks as the forces in the non-meaningful condition were not consistently related to the weights. Further consider that, participants in the control condition only gave one force-response with both hands that did not covary with one vs. two blocks. Interestingly, the figures seem to indicate that, indeed, the non-meaningful condition motorically differentiated less between one versus two blocks as compared to the meaningful condition. Moreover, the participants in the minimal condition - although not motorically trained to differentiate between weights - did seem to transfer their knowledge motorically, as indicated by the distances between curves.

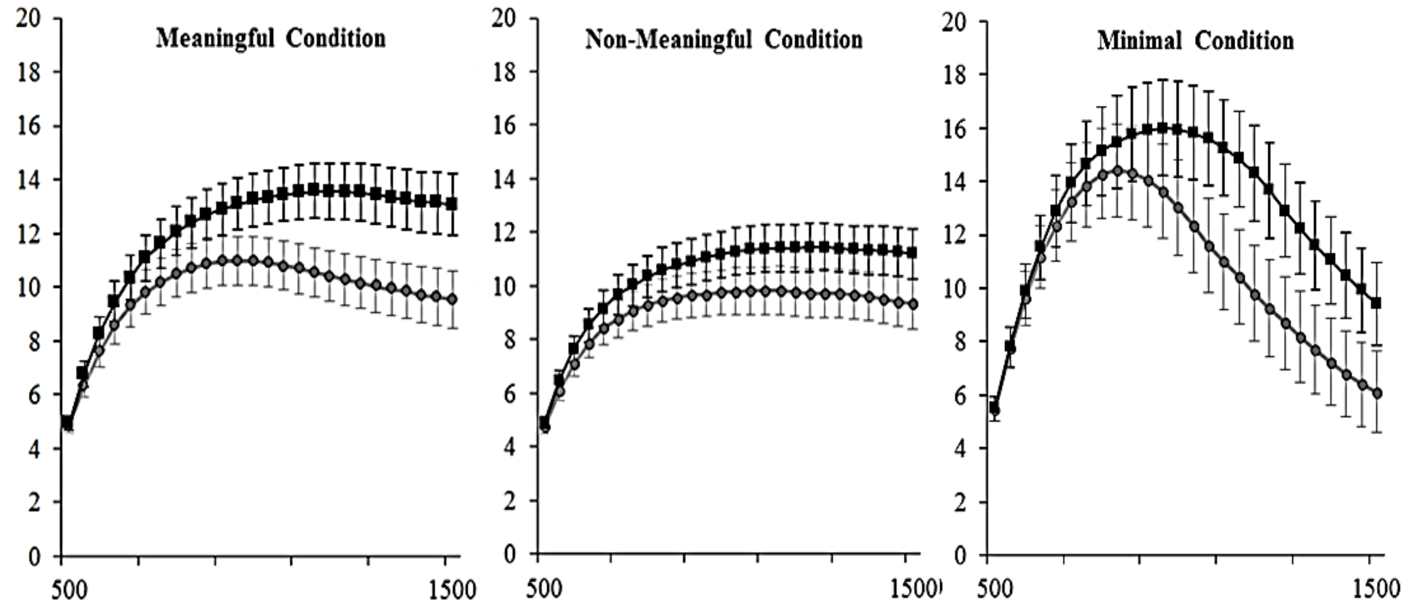


Figure 5. The force-responses for the motor-memory task over time (1000ms) for trials (differentiated by force-for-one-cube [in Grey] and force-for-two-blocks [in Black]) per condition. Error-bars indicate 95% CIs.

Ratio Measure

To test whether these differentiations for force-responses for one vs. two-cubes trials were significant we obtained a ratio-measure as described in the method section. As is shown in Figure 6 the participants in the Meaningfully-EC(M = .3026, SD = .156) performed better (a score of 0 being perfect) in differentiating between one-cube- versus two-cube forces as compared to non-meaningful(M = .362, SD = .125), and minimal condition (M = .354, SD = .163). However, a one-way ANOVA yielded no significant differences across condition, $F(2, 78) = 1.091$, $p = .342$, partial $\eta_p^2 = .032$. An additional Bayesian analysis for the effect of motor-involvement condition on ratio measure on the motor-memory task yielded $p_{BIC}(H_0|D) = .956$ (Masson, 2011). This probability indicates a 95.6% likelihood that motor-involvement condition (meaningful, non-meaningful, minimal) does not affect motor-knowledge as reflected by the ratio measure (which can be considered strong evidence for the absence of an effect of condition, Kass and Raftery, 1995).

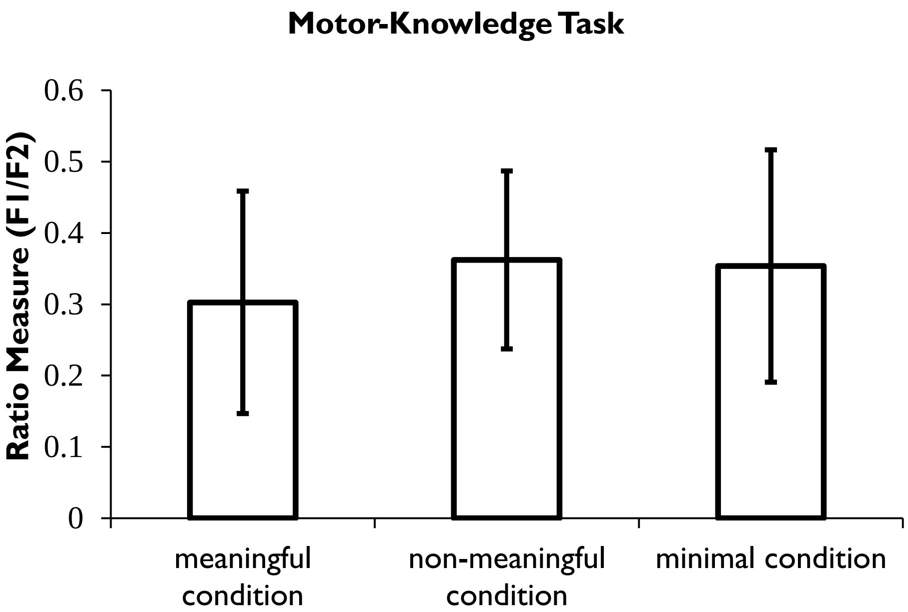


Figure 6. Ratio-measures as a measure of motor-competency to differentiate between force-for-one-cube vs. force-for-two-cubes trials. A score of 0 means a perfect score, meaning that participants' mean force given for two cube trials was twice the force compared to one cube trials.

Task Load

See Table 1, column 4 for the means and standard deviations for the reported mental effort, difficulty, and interest of the motor task across conditions. ANOVAs showed no significant overall main effects of condition on these self-report measures regarding the motor task. Overall correlations between self-report measures and performance, showed that those with that found the task more difficult (r = -.267, p = .028) and more interesting performed (r = -.263, p = .030) better on the motor-task.

Discussion

We investigated whether different meaningful- and non-meaningful forms of physical engagement with instructional animations concerning the workings of class one levers affects unreflective (RT-task), reflective (transfer task), and motoric (motor task) competency on problem-solving. The results showed that either a training-phase in which participants learned how to physically balance a virtual seesaw (meaningful condition), a training in which participants physically balance a seesaw but with inconsistent weight mapping (non-meaningful condition), or a training phase in which participant merely activated the playing of an instructional animation through a minimal physical engagement (minimal condition), did not differently affect performance on RT-task or the transfer task.

Participants in the minimal condition did have different motoric judgments of the force that needed to be applied to balance a seesaw. This result is not surprising since participants in the minimal condition learned to balance a seesaw only through minimal physical engagement (3lbs) whereas the other participants consistently or inconsistently learned to balance a seesaw around 5 and 10 lbs for one-cube and two-cube forces, respectively. Nevertheless, this result confirms that there was *some* implicit embodied memory of the correct sensori-motor dynamics with the seesaw during the training phase. Yet no significant differences were found between conditions for the ratio measure, which was designed to assess motoric competence in correctly differentiating between one- vs. two-cube forces over an interval of 500-1500 milliseconds. Interestingly, visual evaluation of Figure 6 shows, participants in the minimal condition show *some* motoric competence (as indicated by differentiation of forces between one versus two blocks between 1000-1500 milliseconds after response-onset), which suggests that knowledge about the mechanisms of the seesaw learned through non-motoric means may transfer to motoric competence.

Yet, this study has some limitations that might have prevented us to find the hypothesized beneficial (negative) effect of the meaningful (non-meaningful) training-phase as compared to the minimal condition. Firstly, although the current paradigm was explicitly designed to pick up potential small effects of training of a non-reflective and automatic sort, it might be the case that the manipulation was simply too short to imbue effects of the different training-phases (approximately two minutes). Indeed, it could be argued that perhaps especially in learning sensori-motor routines repetition is important to achieve a certain level of competence (e.g., Marley & Carbonneau, 2014). This can be appreciated by the fact that, in contrast to understanding a propositional rule, motor-competence does not follow an *either-or* transition of understanding (cf. Ryle, 1949). Thus, for the learner, practice might be a very important factor to pick up information that is constituted by the structural correlations that emerge during interaction, or in simpler words, embodied learning takes time.

Another limitation of the present design is that we could have obtained a more sensitive measurement by including a pre-test. For example, Zacharia and colleagues (2012) showed that children that had correct conceptions of mass and its effect on a balance-beam were not benefiting from physically engaging with learning materials. It is thus possible that the learners' degree of competence affect whether physical engagement is beneficial for learning; unfortunately the present design fails to take this into account.

Additionally, it might be argued that the null-findings actually show that the more passive training (i.e., minimal condition) was more efficient for learning than the other forms of physical engagement. After all, participants in the minimal condition had a significantly shorter study time as compared to the meaningful and non-meaningful condition. Unfortunately this is difficult to assess. However, the reason why the embodied instructional animations took longer is that participants had to acquire competence in wielding the Wii-board (for example, during the training phase participants often over-pushed and then stopped pushing altogether to begin all over again). As such it can be argued that participants in the physically engaged conditions were actually performing several tasks at once, and were thus in another respect hindered to study the materials.

Methodological issues aside, given that previous research (with children and adults) does not consistently find a potential beneficial role of augmenting instructional animations with sensori-motor information (Bivall et al., 2011; Schönborn et al., 2011; Wiebe et al., 2009; Zacharia et al., 2012), it might be the case that learning *how* to do something physically is not always necessary to know *that* a mechanical device works such and so. In other words, perhaps learning the workings of levers can be done entirely through visual information alone in the current task (i.e., the actions participants performed in the meaningful condition were not relevant). In fact, it might have worked the other way around. Knowing-that informs how to motorically balance a seesaw, as indicated by the apparent motor competence of participants in the minimal condition. Indeed it has been argued, that when visual information is present and usable to understand a particular task at hand, haptic information - even when it provides extra information - will not necessarily be used next to visual information (Driver & Spence, 1998; Klatzky & Lederman, 2002). Moreover, it may be the case that integrating haptic information with visual information produces additional cognitive load which counteracts potential beneficial effects of extra-visual information provided by haptic interaction (Skulmowski, Pradel, Kühnert, Brunnett, & Rey, 2016). Yet, it is important to note that on our reading of most theories on embodied learning (for an overview of such theories see Pouw et al., 2014) the current actions performed with the virtual seesaw would be relevant for further reasoning with seesaws. Namely, learning to judge the force needed to balance a seesaw motorically corresponded lawfully (in the meaningful condition) with the visual information (i.e., the number of blocks, in combination with the position of the blocks on the arms of the seesaw, lawfully corresponded to the force that needed to be applied by the participant on the relevant arm). Although of course, it cannot be excluded that a more natural correspondence of action and perception (say interacting with an actual seesaw) would have provided different results. Nevertheless, the visual information presented in the motoric training sessions directly corresponded with the visual information provided in the subsequent performance RT- and transfer tasks. Embodied learning theories prescribe that after motoric experiences further visual encounters with similar situations are laden with previous multimodal associations and become in fact part of reasoning with such visual information (e.g., Barsalou, 1999). If the current data reflect a true null-effect, it thus seems to suggest (for the present context) that these multimodal associations predicted by embodied learning theories are a) either not established after a short motoric training or/and thus b) not used for further reasoning.

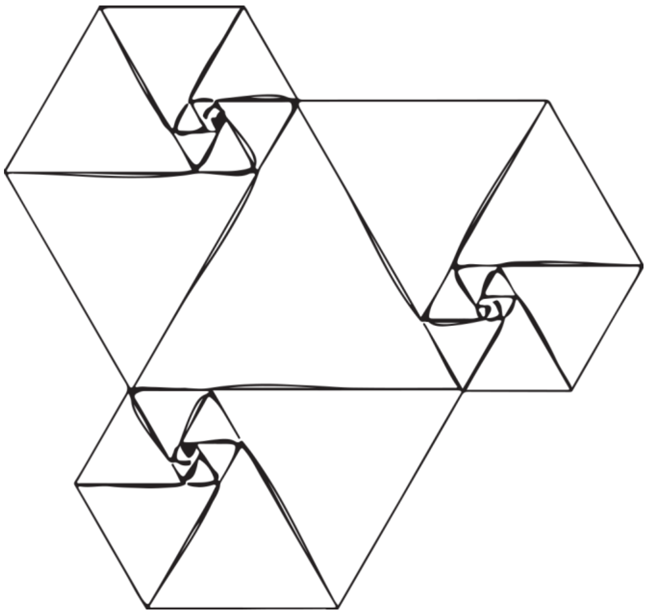
Additionally, it might be that basic concepts such as weight and mass are learned early on in childhood, and thus need not be provided with extra information anymore. In other words, certain basic concepts are already grounded in physical experiences. For example, in the current task, since participants were able to differentiate between one cube and two-blocks separately, as well as the position of the blocks through visual information alone, it might have rendered the motoric information redundant for the participant. As such, grounding science content in physical experiences is only necessary if it adds something otherwise unknown to the learner (e.g., Pouw et al., 2014; Zacharia, 2012). However, the idea that participants' knowledge is still grounded in physical experiences is less informative to address the current results given the finding that participants were not affected by contradictory physical experiences provided in the non-meaningful training-phase. Furthermore, we should highlight (as reported in the method section and results) that we performed a pilot test using instructional animations about class-1 lever problems wherein we did find large performance effects of animation versus no animation on the RT-task with adults. Additionally, as reported in the results section, in the pilot study it was already established that providing participants with a similar but non-interactive instructional animation leads to better performance (69.26%) on the RT-task than no animation (57.24%). If we consider that as a baseline for the current study (see Figure 3a) in which participants interacted with the animation, we see that the present sample performs much better on that same task (80.04%). This suggests that the instructional animations are effective for improving performance compared to no training, and therefore, that the current lack of differences between conditions cannot be explained merely by poor learning effectiveness (i.e., floor effect) of the instructional animations. Furthermore, with regard to the fast decision making that was required in performing the RT-task competence (as compared to the transfer task), competence is likely to be a matter of small degrees which we believe would be affected by our manipulation in the present context (if there were an actual effect).

How should we relate the present null-findings to other positive findings in the literature (e.g., Han & Black, 2011; Kontra et al., 2015)? We believe the key difference is that the present study differs from previous studies on learning science concepts through physical interaction as we manipulated the lawful information that physical interaction affords, as opposed to contrasting different modes of physical interaction (e.g., mouse-based versus haptic manipulation, e.g., Skulmowski et al., 2016; Zacharia et al., (2012); no manipulation vs. haptic manipulation, e.g., Han & Black, 2011, Kontra et al., 2015). As such we aimed to exclude effects that can be attributed to different modes of physical interaction. Of course, this is not to say that previous studies that revealed an effect of different modes of physical interaction cannot be attributed to the lawful information that is afforded by these different modes of physical interaction. In fact, if embodied learning theories are correct, positive effects of physical interaction should be explained in terms of meaningful correspondences with the learning content (Pouw, van Gog, & Paas, 2014). Yet, if our interpretation of the current (unexpected) results is on track, caution is advised when attributing effects of physical interaction based on meaningful correspondences that may exist between action and science concepts. Therefore, future research could focus more on manipulating structural information that emerges out of perception and action loops (by loosening or tightening the correspondence between action and its perceptual correlates) rather than manipulating the perception-action loop altogether (i.e., manipulating the mode of interaction).

In sum, the current findings are interesting as it shows that physical experiences in adults are not readily or easily integrated with the knowledge schemas of the kind that allows one to solve the performance tasks reported here. This resonates well with findings regarding physics misconceptions, which show that incorrect knowledge schemas are not easily altered by concrete counterevidence (Duit & Treagust, 2012). Thus future research could focus more on longer bodily training, and more specifically how this affords learners meaningful information (rather than mere physical engagement) that is *not* provided by the visual modality alone.

Chapter 4

Augmenting instructional animations with a body analogy to help children learn about physical systems*



*This chapter has been This chapter has been accepted for publication as Pouw, W. T. J. L., Van Gog, T., Zwaan, R. A., & Paas, F. (2016). Augmenting instructional animations with a body analogy to help children learn about physical systems. *Frontiers in Psychology,* 7, 860.

Augmenting instructional animations with a body analogy to help children learn about physical systems

We investigated whether augmenting instructional animations with a body analogy would improve 10-13 year old children's learning about class 1 levers. Children with a lower level of general math skill who learned with an instructional animation that provided a bodily analogy of the physical system, showed higher accuracy on a lever problem-solving reaction time task than children studying the instructional animation without this body analogy. Learning with a body analogy led to slower reaction times, but did not affect accuracy and solving speed on a transfer task as compared to learning without this body analogy. These results suggest that providing children with a body analogy during animation study provides a stepping-stone for understanding abstract principles of a physical system.

Introduction

Instructional animations (from hereon: IA) are increasingly implemented in educational environments (Chandler, 2009). The value of animated over static visualizations for instruction can be intuitively grasped: IA offer the learner *direc*t pick-up of process related information (i.e., information that interacts with time, such as causality and motion), which must be *inferred* from static visualizations (Spanjers, Van Gog, & Van Merriënboer, 2010). Surprisingly, empirical results concerning the effectiveness of IA are not as encouraging as these intuitions would predict. For example, in the instructional domain of physical systems (e.g., gears, electrical systems etc.), although visual presentation benefits learning overall (as opposed to non-graphical instructions), findings regarding the effectiveness of animated versus static visualizations are mixed (Hegarty, Kriz, & Cate, 2003).

Based on the mixed results Tversky, Bauer-Morrison, and Bétrancourt's (2002) concluded: "The many failures to find benefits of animation …calls for deeper inquiry into information processing of animation" (p. 255). This was taken to heart, and later studies have suggested that the main problem with learning from dynamic visualizations is that it imposes a high cognitive load on working memory from the learner due to information transience inherent to dynamically changing visualizations (Ayres & Paas, 2007a, b). To be effective, it is argued, the negative effects of transience in IA need to be counteracted, for instance, by means of cueing, or segmentation (Spanjers et al., 2010).

There is one type of task, however, for which IA consistently seem beneficial for learning compared to static visualizations even without measures to counteract transience. Namely, a meta-analysis (Höffler & Leutner, 2007) showed a small effect size of learning gains in animated vs. static visualizations under the condition that the instructional content involves learning bodily routines (e.g., origami, assembly, knot tying). It has been suggested that because human movement is automatically and efficiently processed by the cognitive system (we will return to this in the next section), the transience inherent in IA depicting such tasks may be counteracted (Van Gog, Paas, Marcus, Ayres, & Sweller, 2009).

Indeed, evidence is accumulating that the human cognitive system is distinctively attuned to the body, the body of others, and its possibilities for interactions (e.g., Amorim, Isableu, & Jarraya, 2006; Rizzolatti, Fadiga, Gallese, & Fogassi, 1996). For example, neuropsychological evidence suggests that perceived human body parts are distinctively processed in particular areas of the brain (extrastriate body area; Peelen & Downing, 2007) as compared to perceived body parts of non-human animals (Peelen & Downing, 2007). Moreover, human bodies are readily mapped onto one's own body schema (Semenza & Goodglass, 1985; Van Gog et al., 2009). For instance, mental rotation of shapes represented as a body is performed faster than mental rotation of inanimate objects (Amorim et al., 2006).

Therefore, in the present paper we investigate whether augmenting IA with a body analogy improves learning about non-human movement content (originally proposed by De Koning & Tabbers, 2011). Specifically, we investigate whether the effectiveness of IA might be improved by augmenting the learning content (in this study: class I lever problems) with a body analogy. We hypothesize: by meaningfully mapping a physical body on a physical system during instruction, a less cognitively demanding route of knowledge-transfer might be created (as opposed to learning about inanimate objects). "Less demanding", as learners readily map bodily actions on their own body schema. Moreover, learners are very familiar with forces acting on the body, which can be used as an analogy for forces acting on physical systems.

There is evidence already that the body can be mapped on physical systems. For example, when children or adults convey their knowledge about a particular topic they often use gestures that are meaningfully related to the topic's content (e.g., Garber & Goldin-Meadow, 2002; Goldin-Meadow & Alibali, 2002; Goldin-Meadow, & Momeni-Sandhofer, 1999; Hutchins & Nomura, 2007). Importantly, gestures do not simply mirror what is expressed in speech. Rather, gestures can accommodate and complement what is expressed verbally with idiosyncratic information expressed in gesture alone. For instance, in a study by Pine, Lufkin, and Messer (2004) co-speech gestures that emerged when children explained the workings of a class I lever (balance beam) were analyzed (see also Pine & Messer, 2000). To solve lever (e.g., a balance-beam) problems children must attain knowledge about the effects of I) weights, II) distance of the weight from the fulcrum, and III) the positioning of the fulcrum. About one-third of the children (5 to 9 years) explaining the solution to a balance-beam problem produced gesture-speech mismatches. Children verbally explained the solution to the problem in terms of one property (e.g., I; talking about the weights on the beam), while concurrently expressing another (more advanced property) in gesture (e.g., III; expressing the position of the fulcrum in gesture). Even more remarkably, those children that produced mismatches as compared to those that did not, were more likely to improve on pre- to post-test measures of learning. If knowledge about physical systems develops in sensori-motor modalities as research on gesture suggests, augmenting the learning content with sensori-motor stimuli might improve learning (Höffler & Leutner, 2007; Van Gog et al., 2009; Pouw, Van Gog, & Paas, 2014).

Yet, it seems to be the case that augmenting IA about physical systems with sensori-motor information may be suitable for some but not for others (Zacharia, Loizou, & Papaevripidou, 2012; for an overview see Pouw et al., 2014). For example, kindergartners' learning about balance beams improved when they were given opportunities to physically interact with a balance beam (class 1 lever), but only when they possessed an incorrect preconception of how a balance beam works (Zacharia et al., 2012). This suggests that especially those with incomplete understanding of a physical system are aided by additional body-analogous information. Therefore, it is important to take into account learners cognitive predispositions when investigating the instructional potency.

Present study

In the present study primary school children learned from IA about a class 1 lever (a seesaw). The workings of levers can be considered as a classic context to test children's conceptual and procedural learning processes about physical systems (Dixon & Dohn, 2003; Karmiloff-Smith, & Inhelder, 1974; Pine et al., 2004). We designed an IA (duration 6.5 minutes) in which relevant concepts for understanding the working of a seesaw were demonstrated, such as weights, balance, fulcrum, and mechanical advantage. Half of the sample was confronted with a 'body analogy IA' in which a transparent body was projected onto the seesaw (see Figure 1: body-analogy condition) and the other half were given the same IA without this body analogy (control condition). The body provided an analogy of the concept of mechanical advantage: objects placed further from the fulcrum (analogy: joint) will exert more force than objects placed closer to the fulcrum. Furthermore, if similar weights are put at similar places on the arm they will feel equally heavy (balance) or when they are located at different places, they will not feel equally heavy (disbalance).

Learning performance was assessed through a three choice reaction-time task that assessed accuracy and speed of determining whether a seesaw will pivot to the left or the right, or will balance out, given different configurations of the weights, and the positions of the weight relative to the fulcrum. Additionally we confronted children with a similar three-choice transfer task that consisted of new concepts, such as interconnecting seesaws, or replacement of the fulcrum.

We hypothesized that the body analogy (BA) condition as compared to the control condition would show better learning overall (i.e., higher accuracy, faster solving speed on the test tasks). Importantly, to minimize individual cognitive differences between conditions we semi-randomly assigned conditions based on general math scores of the children. We used children's math scores as they are closely related to learning about physical systems, and have been found to strongly correlate with their visuospatial working memory capacity (e.g., Van der Ven, Straatmeier, & Jansen, 2013), which directly relates to issues of cognitive load associated with instructional animations (Ayres & Paas, 2007a, b). Per exploration we also investigate whether general math skill interacted with the effectiveness of the conditions, as it might be an important cognitive predisposition for learning in the current domain. We also measured subjective experiences of cognitive load, by asking children to rate how much mental effort they invested and how difficult they found the tasks. In addition, we asked them to rate how interesting they found the tasks, which could give an indication of differences in cognitive engagement.

Method

Participants and Design

This study was conducted in accordance with the guidelines of the ethical committee of the institute of psychology at Erasmus University Rotterdam. All children participated based on parental informed consent, where information about the study was provided two weeks prior to the experiment and parents were given the opportunity to withdraw their child from participating. A total of 74 Dutch primary school children (3 classrooms from 2 separate schools) were tested (mean age, 12.49, $SD = 0.54$; Range 10-13; 51.4 % female). The two IA-conditions were: control ($N = 36$, 52.8% female) vs. body analogy (BA; $N = 38$, 50 % female). Children were pseudo randomly assigned (see Table 1 for frequencies) to condition by matching for level of general math skill as measured by the national standardized Cito math test or (in one school) an equivalent standardized test that assigns the children to comparable levels of skill as the Cito test does. From highest to lowest, these are: A (highest 25%), B (next 25%), C (next 25%), D (next 15%) and E (lowest 10%). This test was taken within the school-semester year in which the experiment took place, and the children's scores were provided by the schools.

	Control Condition	BA condition
A	7	8
B	12	12
C	7	8
D	8	7
E	2	3
Total	36	38

Table 1. Number of participants per condition and general math skill.

Materials

Instructional animations

The IA[9] were designed in Adobe Flash Professional CS 5.5. The voice-over and textual instructions were programmed in ActionScript 3.0 (IA's can be downloaded at http://charlyeielts.nl/bodyanalogy/materials.html). The IA consisted of an introduction to the basic concepts of class 1 levers narrated by a female voiceover and explained with a dynamic visualization of a seesaw. In the first part of the IA (3.5 minutes), basic concepts such fulcrum, left and right arm of the seesaw, (dis)balance, weights, and mechanical advantage was introduced. Throughout the instruction no explicit information was provided about formulas related to the constructs. For example, mechanical advantage was explained by showing a balanced seesaw in a mechanical advantage state, with the voiceover instruction informing learners that: "The heavy weight is twice as heavy as the lighter weight, but the seesaw is still in balance! This is because the distance of the heavy weight is two times closer to the fulcrum than the lighter weight" (for further instructions see http://charlyeielts.nl/bodyanalogy/materials.html). The second part of the IA was not narrated and consisted of 24 trials (3 minutes) that

[9] The learning effectiveness of the animations was tested with the reaction time task in a pilot-study with adults (N = 78) using Amazon's Mechanical Turk. We translated the exact instructional materials designed for Dutch children for the English speaking adult sample. This pilot test showed that the animations were effective for learning (accuracy on the reaction-time task) as compared to no instruction (t (76) = -2.644, p = .010, Cohen's d = .602 [large effect]). No effects were obtained for solving speed on the RT-task.

showed different configurations of weights on varying positions from the fulcrum and its effect on the seesaw (tilt left, right, or balance).

For the BA condition the only difference in the IA as compared to the control condition was that a transparent human body was additionally projected over the seesaw (i.e., no differences in narrated instruction). Importantly, the arms of the projected body moved together with the movement of the seesaw (see Figure 1). Only once in the narration (but in both conditions) a reference was made to how it would feel to have weights on one's actual arms. This reference was made after the explanation of mechanical advantage, which showed a seesaw balancing out with unequal amount of weights (see Figure 1). This was done to ensure that children in the BA condition would be more likely to see the relevance of the body projected over the seesaw.

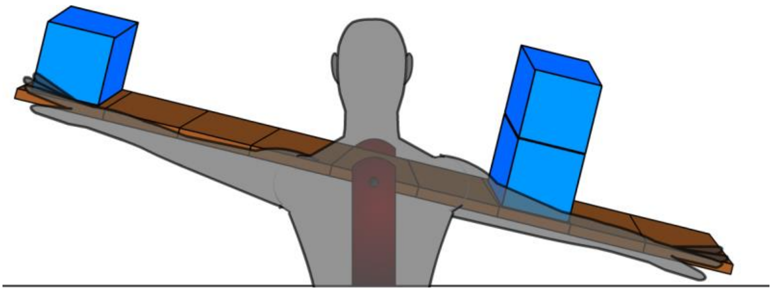


Figure 1. A snapshot of the instructional animation in the BA condition (the seesaw will balance out in this example).

Reaction-time task

A three choice reaction-time task was developed (programmed in E-prime) to assess children's accuracy (number of correct responses) and speed (reaction-time) in solving class-1 lever problems. The RT task consisted of 45 trials (and three practice trials) in which children had to judge whether a seesaw would balance, or tilt down to the left or to the right. Each trial showed a seesaw with one or two blocks on either side of the arms of the seesaw on deferring distances from the fulcrum (see Figure 2 for an example). The number and location of the weight varied for these 45 trials. Children were required to determine which way the seesaw would tilt, or whether it would attain balance, regardless of the current state of the seesaw (i.e., tilted to left/right or balanced). We varied the initial state of the seesaw randomly as to prevent any spurious effects of the initial state of the seesaw on accuracy and speed. Children responded by pressing on a QWERTY keyboard, "P" if the seesaw would tilt to the right, "Q" if it would tilt to the left and SPACE if the seesaw would be in balance.

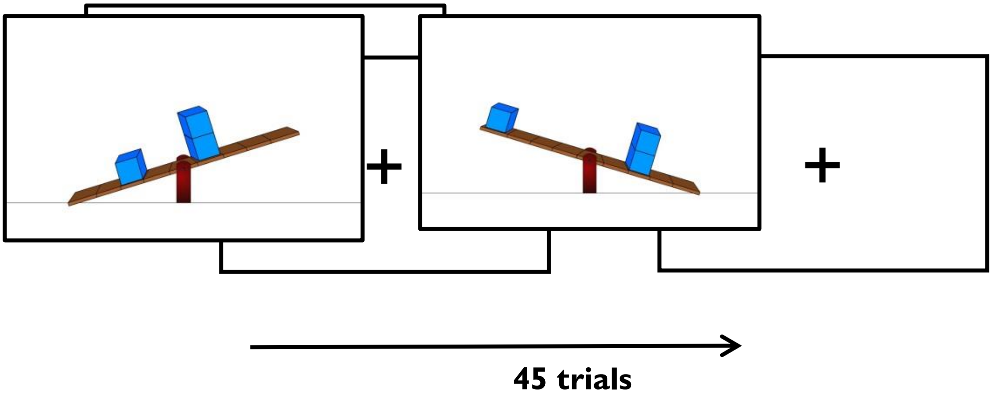


Figure 2. Example of two reaction time trials. Note that trials were given the initial state of the seesaw randomly and the children answered with button presses what the correct state of the seesaw would be (pivot left, balance, pivot right).

Transfer task

The transfer task, consisting of 15 lever-problems, aimed to assess children's ability and solving speed to further apply the principle of mechanical advantage on new or more complex problems. Twelve problems required children to judge what the end-state would be (tilt left, right, or balance) of a particular seesaw in a set of two interconnected seesaws, in four of those trials the fulcrum was not placed in the center (see Figure 3). The last three problems required children to predict how these forces would act on the body (e.g., how heavy a block would feel when placed on the arms, or which seesaw needed to be pushed down the hardest given a number of weights on the seesaws).

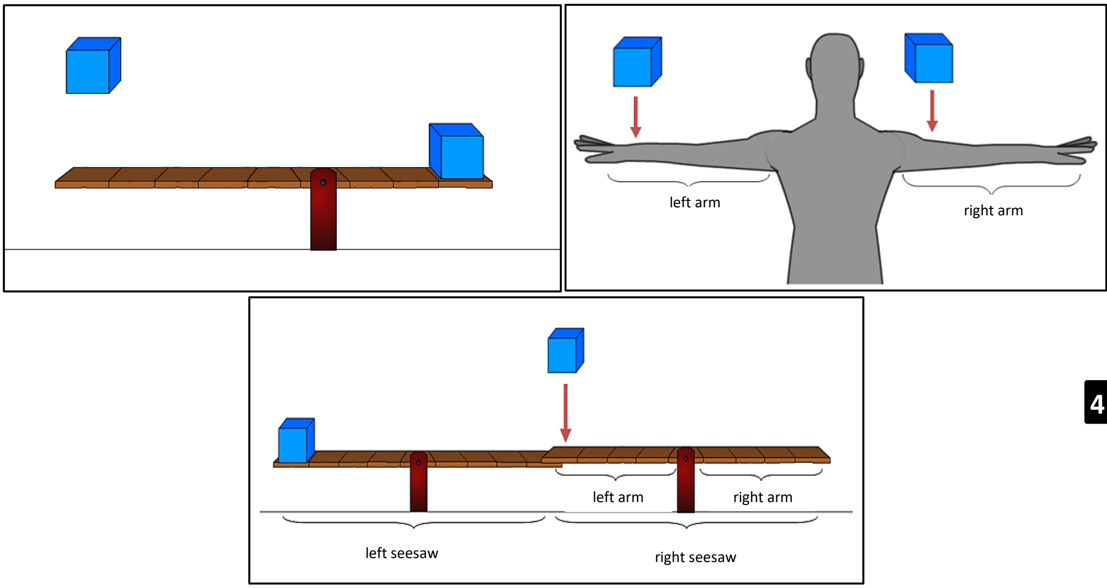


*Figure 3***.** Example of three transfer task problems. In the first example (left above), children were asked to "Judge whether the seesaw pivots to the left, remains in balance, or pivots to the right". This was the same for the second example (right above), but then for the right seesaw. In the third example children were asked to "which weight will feel the heaviest for this person, or will the weights feel just as heavy?".

Mental effort, difficulty, interest

We obtained ratings of experienced mental effort, interest, and perceived difficulty of the instructional animation, RT-task, and the transfer task directly after completion. Children answered on a 5-point scale "How hard did you need to think to understand the previous video/task" (mental effort; 1 = 'not hard', to 5 = 'very hard'), "How interesting did you find this previous video/task" (interest; 1 = 'not interesting', to 5 = 'very interesting') and "How difficult did you find this previous video/task" (difficulty; 1 = 'not difficult', to 5 = 'highly difficult').

Demographics

Information on age, sex, and Cito test score of general math skill of the children were provided by the schools.

Procedure

Children were tested one or two at a time, in a quiet room at their school. If children were tested at the same time the two experimenters ensured that children did not face each other directly and that there was enough distance between them so that they were not disturbed in any way. Children were seated in front of a laptop and were informed that they would watch an instructional video and perform two tasks to assess what they had learned. They were subsequently asked to put on the headphones so that the experimenter could start the video. Subsequently, children performed the reaction-time task and were instructed to do so "as fast and accurate as possible". Beforehand, children were given three easy practice trials which the experimenter could repeat if needed to ensure they understood the task. Subsequently, children were confronted with the transfer task that was provided in a booklet and they could solve at their own pace (i.e., speed was not emphasized as in the RT test task). The experimenter used a stopwatch to assess overall solving speed. Immediately after watching the IA, performing the RT, and solving the transfer task, children completed the subjective ratings of effort, interest and difficulty that were printed on a sheet of A4 paper per task. All children received a small present for their participation (handed out in class on the last day of testing).

Data Analyses

Accuracy and RT-scores for the transfer task and RT-task more than 2 SD from the overall-mean were treated as outliers and were excluded from the analysis (reported in the Results section when applicable).

Reaction-time task

Performance accuracy was measured by summing the correct answers on 45 trials (range: 0-45) and speed was measured by computing the mean reaction time (in ms) on correct trials.

Transfer task

Performance was measured by summing the correct answers on 15 trials (range: 0-15) as well as overall solving speed in seconds.

Results

Mental Effort, Difficulty, Interest

Data are presented in Table 2. T-tests showed no significant differences between conditions in self-reported mental effort, difficulty, or interest, on the IA, RT-task, or transfer task.

Condition		Instructional Animation		Reaction-Time Task		Transfer Task	
		M	SD	M	SD	M	SD
Control Condition	Mental Effort	1.81	1.064	2.25	1.05	2.50	1.00
	Interest	3.56	1.319	3.58	1.36	3.69	1.33
	Difficulty	1.92	1.13	2.58	1.16	2.64	1.10
BA Condition	Mental Effort	1.79	.935	2.26	1.155	2.27	1.03
	Interest	3.45	1.350	3.74	1.178	3.26	1.35
	Difficulty	2.11	1.23	2.37	1.08	2.71	1.04

Table 2. Means and *SD*'s per condition and task-phase for mental effort, interest, and difficulty.

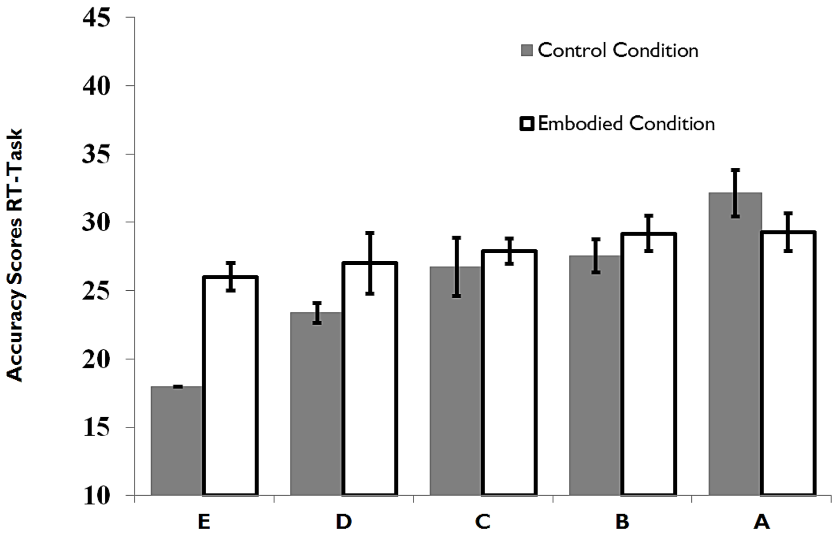


Figure 4. Accuracy scores and standard error per condition and general math skills (E = lower and A = higher general math skill).

RT-task Performance

Accuracy

The overall accuracy score on the RT-task was 59.91% (*M* = 26.96, SD = 5.56). Four participants scored < 2 SDs below the mean (i.e., < 15) and were therefore excluded from the analyses (no participants scored > 2 SD). This resulted in an analysis on data of 70 participants, with *N* = 34 in the control condition (*N* = 7 on math skill level A, *N* = 11 on level B, *N* = 7 on level C, *N* = 8 on level D, and *N* = 1 on level E), and *N* = 36 in the BA condition (*N* = 8 scoring on math skill level A, *N* = 12 on level B, *N* = 8 on lebel C, *N* = 6 on level D, and *N* = 2 on level E). Math skill was a significant predictor, $F(1, 68) = 17.256$, $p < .001$, explaining 19.1% of the variance (based on $R^2_{adjusted}$), with higher math skill resulting in higher accuracy, $\beta = .450$, $t(68) = 4.154$, $p < .001$.

The effect of condition was assessed by adding condition as a predictor for RT accuracy into a stepwise hierarchical regression after math skill. Condition was coded as 0 for the control condition and 1 for the BA condition. The overall model remained significant, $F(2, 67) = 9.417$, $p < .001$, explaining 19.6% of the variance in RT-accuracy. Condition was a positive but non-significant predictor for RT-accuracy, $\beta = .130$, $t(67) = 1.208$, $p = .231$. Math skill remained a significant predictor, $\beta = .429$, $t(68) = 4.136$, $p < .001$.

We further assessed whether general math skill moderated the effect of condition by adding an interaction term of condition and math skill into the regression model. This resulted in significant model-fit, $F(3, 66) = 8.533$, $p < .001$, explaining 24.7% of the variance in RT accuracy. General math skill remained a significant predictor, $\beta = .704$, $t(66) = 4.645$, $p < .001$, and now condition was significantly positively related with RT accuracy, $\beta = .230$, $t(66) = 2.040$, $p = .045$, $R_{Partial} = .244$. Furthermore, there was a significant interaction, $\beta = -.371$, $t(66) = 2.346$, $p = .022$, $R_{Partial} = -.277$, indicating that children with lower math skill were more likely to be positively affected by the BA condition (in terms of RT-accuracy) than those with higher math skill (see Figure 4).

Speed

The overall mean reaction time on correct trials was 2791ms (*SD* = 1331). Three additional participants were excluded from the analyses as their data fell over 2 SDs above the mean (> 5453 ms; no participants scored < 2 SD). This resulted in an analysis on data of 67 participants, with *N* = 33 in the control condition (*N* = 7 scoring on math skill level A, *N* = 10 on level B, *N* = 7 on level C, *N* = 8 on level D, and *N* = 1 on level E), and *N* = 34 in the BA condition (*N* = 8 scoring on math skill level A, *N* = 12 on level B, *N* = 6 on level C, *N* = 6 on level

D, and $N = 2$ on level E). Math skill was not a significant predictor, $F(1, 65) = 0.327$, $p = .569$, showing a non-significant relation with speed on correct RT trials $\beta = -.071$, $t(65) = -.572886$, $p = .569$. We added condition together with general math skill as a predictor for speed on correct trials into the hierarchical regression model. The overall model-fit was non-significant, $F(2, 64) = 2.878$, $p = .064$, math skill remained a non-significant predictor, $\beta = -.083$, $t(64) = -.0938$, $p = .352$, and condition was a positive significant predictor, with children in the BA condition being slower on correct trials overall, $\beta = .279$, $t(64) = 2.325$ $p = .023$. To assess a possible interaction effect we entered the interaction term of condition and math skill into the regression model, this yielded no significant results, nor a greater fit of the model.

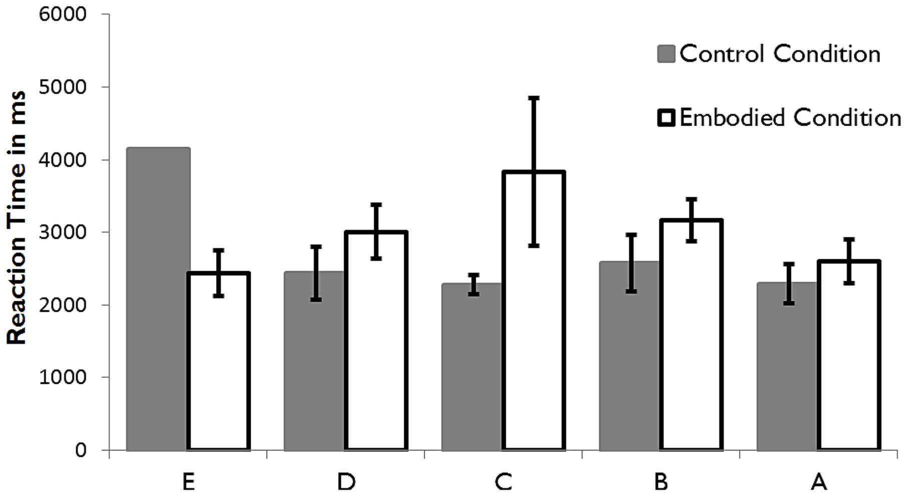


Figure 5. Mean reaction times and standard error for the RT-task per condition and general math skills (E lowest score, A highest score on general math skills).

Transfer Task Performance

Accuracy

The overall accuracy on the transfer task was 49.62% ($M = 7.38$, $SD = 1.90$. Two participants performed < 2 SDs below the mean (< 3.58; no participants scored > 2SD) and were therefore excluded from the analyses. This resulted in an analysis on data of 72 participants, with $N = 36$ in the control condition ($N = 7$ scoring on math skill level A, $N = 12$ on level B, $N = 7$ on level C, $N = 8$ on level D, and $N = 2$ on level E), and $N = 36$ in the BA condition ($N = 8$ scoring on math skill level A, $N = 12$ on level B, $N = 8$ on level C, $N = 6$ on level D, and

N = 2 on level E). A regression analysis showed that math skill was a significant predictor of transfer task performance, $F(1, 70) = 0.732$, $p < .01$, explaining 8.2% of the variance, showing a positive relation with performance $\beta = .308$, $t\ (70) = -.2.706$, $p < .01$.

We added condition after math skill as a predictor for transfer task performance into the hierarchical regression model. The overall model-fit remained significant, $F(1, 69) = 3.697$, $p < .05$, now explaining 7.1 % of the variance of performance. Math skill remained a significant predictor, $\beta = .310$, $t(69) = 2.705$, $p < .01$. Condition was not a significant predictor, $\beta = -.403$, $t(68) = -.403$, $p = .688$. We further added an interaction term of condition and general math skills into the hierarchical regression model, but this resulted in a model with only non-significant predictors ($p > .246$).

Speed

The overall mean solution speed on the transfer task was 308 seconds (*SD* = 77.11). Two additional participants were excluded from the analyses as their data fell 2 SD's from the mean (> 462 seconds; no participants scored < 2SD). This resulted in an analysis on data of 72 participants, with *N* = 34 in the control condition (*N* = 6 scoring on math skill level A, *N* = 11 on level B, *N* = 7 on level C, *N* = 8 on level D, and *N* = 2 on level E), and *N* = 38 in the BA condition (*N* = 8 scoring on math skill level A, *N* = 12 on level B, *N* = 8 on level C, *N* = 7 on level D, and *N* = 3 on level E). We first assessed whether math skill predicted overall speed on Transfer task in a regression analysis. Math skill was not a significant predictor, $F(1, 70) = .203$, $p = .653$, math skill $\beta = -.054$, $t(70) = -0.451$, $p = .653$. We added condition next to general math skill as a predictor for speed on transfer task into the hierarchical regression model. The overall model-fit was not significant, $F(2, 69) = 1.11$, $p = .335$, $R^2_{adjusted} = .003$. Math skill remained a non-significant predictor, $\beta = -.059$, $t\ (69) = -0.496$, $p = .352$, and condition was a non-significant predictor on solving speed on the transfer task, $\beta = .168$, $t\ (69) = 1.419$, $p = .160$. We obtained no significant results when entering an interaction term after math skill and condition.

Discussion

We investigated whether children's learning benefited from augmenting an instructional animation (IA) about class I levers with a body analogy. It was found that when taking general math skill into account as a moderator, this BA condition was positively affecting lever problem-solving accuracy on the RT-test as compared to the control condition, in which the same instructional animation was shown without the body analogy. However, this effect was qualified by an interaction, showing that the BA condition improved accuracy on the RT-test for children with a lower level of general math skill, and was absent (if not reversed) for children with a higher math skill. Finally, no evidence was obtained for performance benefits on the transfer task.

As the results are mixed, the question arises whether the body analogy was "analogous enough" to be informative for learning. Indeed, there are important differences between a seesaw and the body analogy. Most notably, the body analogy is imperfect, as the body has two joints with independent moving arms whereas the seesaw has one fulcrum with movement of the arms that are co-dependent. Such (and possibly other) differences might interfere with properly understanding mechanics of seesaws. However, there is some information in the body analogy that directly corresponds with the mechanics of the seesaw. Namely, there is a one to one correspondence to the difference in weight that would be felt when placing blocks on one's arm with that of the direction of pivot of the seesaw. For example, placing 1 block on the left arm of the seesaw near the fulcrum and 1 block on the right arm away from the fulcrum will result in a pivot to the right due to mechanical advantage. This directly corresponds with the relative difference in weight that would be felt when placing 1 block on the left arm near the fulcrum (joint) and one block on the right arm away from the fulcrum (also due to mechanical advantage). Indeed, in the voiceover of the instructional animations we emphasized to the learner that this was a relevant correspondence.

Therefore we speculate that the body-analogous information that was present provided a possible means to process the learning content by activating implicit motor knowledge, which provided those children that are least receptive to learning about abstract content (i.e., those with a lower general math skill) a way to ground unfamiliar force-dynamics of the seesaw in familiar force-dynamics of the body. In line with observations made by Van Gog and colleagues (2009), this grounding would be established through automatic mapping of the model's body onto one's own body. Indeed, it seems that when a rule or process is already understood, additional grounding in concrete experiences is unnecessary

(Zacharia et al., 2012). We further speculate that in the current case children benefited from the BA condition when performing the reaction time task because they were *simulating* the force dynamics related to the body (see Van Gog et al., 2009). This observation is consistent with the reaction time data. Although these data need to be interpreted with caution, given that the overall effect of the model was not significant, the finding that children in the BA condition were slower to respond might suggest that they performed mental simulations during the test. Finally, that we did not find a similar effect of condition on transfer task accuracy as we did in on the reaction time-task, signals that an efficient strategy on one task does not always readily transfer to another. As shown by Dixon and Dohn (2007), when solving problems with interconnecting balance beams (also included in our transfer task), problem solvers may use more abstract strategies (i.e., alternating strategy; see Dixon & Dohn, 2003 for details) than simply judging each state of each seesaw to judge its effect on the next connected seesaw. Perhaps, while judging the forces of a single seesaw like in the RT-task is aided by a body analogy through some simulation or inference, this strategy might prove inefficient for solving the interconnected seesaw problem as more abstract strategies are more efficient and discovery of these abstract strategies might actually be hampered by using a strategy solicited by having learned with a bodily analogy. In sum, future research should be sensitive to the kind of strategy a particular body analogy solicits, and on which tasks that strategy could be expected to help learning.

Furthermore, in line with findings on the expertise reversal effect (Kalyuga, 2007), the accuracy results show that this mental simulation on the reaction time test was only helpful for children with lower math ability (lower visuospatial working memory capacity) but not helpful, or potentially even detrimental, to those with higher ability (working memory capacity). Perhaps those with a higher ability did not require additional help to induce rules from physical systems so that for them mental simulation during the test task evoked by the body analogy is superfluous and possibly distracting process. Perhaps this explains why no effects on the transfer test were found, as it was more difficult to use one's own body as an analogy on most of the test items that involved multiple balance beams.

It should be noted that the present study provides more of a *demonstration* than an *elaboration* of how a body analogy can affect learning. Indeed, the current design has some shortcomings that prevent such elaboration. For instance, although this task was not one taught in school, the possibility cannot be excluded that some children had more prior knowledge than others; and our current design did not allow for assessing learning gains, as we did not provide children with a pre-test. Furthermore, the current results do not allow us to determine whether higher learning outcomes of children with lower math scores in the BA condition were indeed achieved because cognitive load related to transience was counteracted by more efficient processing due to the body analogy. There were no differences in mental effort or difficulty ratings between the conditions, but this does not necessarily mean that cognitive load imposed by transience was not reduced. Perhaps the cognitive capacity that was freed-up by reducing the load imposed by transience, was used for processes that were effective for learning, thereby resulting in an similar experience in cognitive load. Future studies might investigate the underlying cognitive mechanisms in more detail, for instance by using continuous and objective cognitive load measures that can be connected to events in the animation such as dual task measures that do not interfere with animation processing (e.g., Park & Brünken, 2014) or EEG measures (e.g., Antonenko, Paas, Grabner, & Van Gog, 2010).

Future research should further focus on a) the potential difference in receptivity of children with different individual cognitive capacities for learning with body analogies, b) the scope of the effectiveness of bodily analogies on other physical systems (e.g., gear systems, electrical circuits), or even more abstract learning domains such as grammatical or language learning (e.g., Lu, 2011), and c) finally the precise cognitive processes underlying this type of learning (e.g., Brucker et al., 2014).

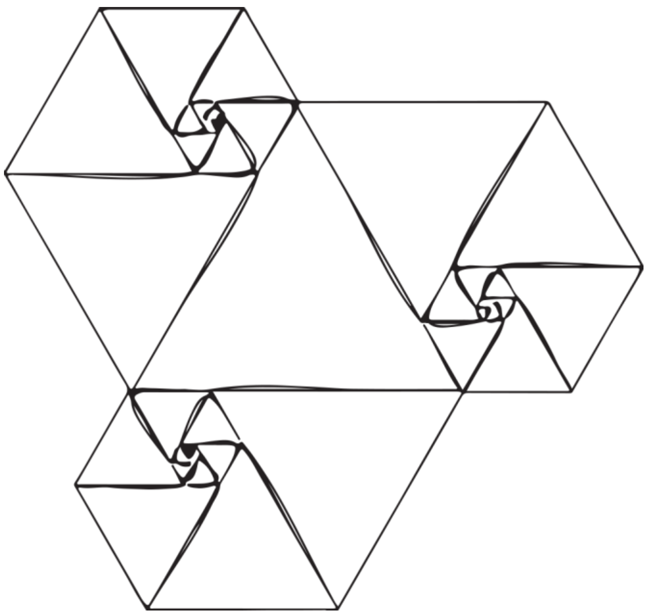
To conclude, despite some limitations, our finding that a relatively simple modification of the instructional animation via a body analogy imbued a positive effect on performance, especially for those with lower general math skill, is a very promising result for future applications in educational practice.

Part II

Effects of Manual Activity on Problem Solving

Chapter 5

Toward a more embedded/extended perspective on the cognitive function of gestures*



5

*This chapter has been published as Pouw, W. T. J. L., De Nooijer, J. A., Van Gog, T., Zwaan, R. A., & Paas, F. (2014a). Toward a more embedded/extended perspective on the cognitive function of gestures. *Frontiers in Psychology*, *5*, 359. doi: 10.3389/fpsyg.2014.00359

Toward a more embedded/extended perspective on the cognitive function of gestures

Gestures are often considered to be demonstrative of the embodied nature of the mind (Hostetter & Alibali, 2008). In this article we review current theories and research targeted at the intra-cognitive role of gestures. We ask the question how can gestures support internal cognitive processes of the gesturer? We suggest that extant theories are in a sense *disembodied*, because they focus solely on embodiment in terms of the sensorimotor neural precursors of gestures. As a result, current theories on the intra-cognitive role of gestures are lacking in explanatory scope to address how gestures-as-bodily-acts fulfill a cognitive function. On the basis of recent theoretical appeals that focus on the possibly embedded/extended cognitive role of gestures (Clark, 2013), we suggest that gestures are external physical tools of the cognitive system that *replace and support* otherwise solely internal cognitive processes. That is gestures provide the cognitive system with a stable external physical and visual presence that can provide a platform to think on. We show that there is a considerable amount of overlap between the way the human cognitive system has been found to use its environment, and how gestures are used during cognitive processes. Lastly, we provide several suggestions of how to investigate the embedded/extended perspective of the cognitive function of gestures.

Introduction

Gestures reflect internal cognitive processes. This is arguably the most fundamental, uncontroversial, and straightforward assumption in the current literature concerning gesticulation. Gestures provide a "window on the mind" (Goldin-Meadow, 2003), which provides a peek into the "embodied nature of the mind" (Hostetter & Alibali, 2008). In less metaphorical terms, it is argued that gestures are direct outcomes of multimodal, sensorimotor or embodied representations that constitute thought processes and speech production. Although not all theoretical perspectives on the function and underpinnings of gestures suggest a purely sensorimotor based approach to mental representations (see Kita, 2000; Krauss, 1998 for alternative views), it is commonly held that activation of the motor-system supports speech production and thought, at least when the conceptual content is visuospatial in nature (Alibali, 2005). Several perspectives on gesticulation (e.g., Kita, 2000; McNeill, 1992; Wesp, Hesse, Keutmann, & Wheaton, 2001) have abandoned the view that gestures are merely communicative tools that are elicited *after* central cognitive processes (e.g., lexical retrieval, conceptualization) have taken place (Graham & Argyle, 1975; Kendon, 1994). Instead, in these perspectives the motor-system has been upgraded from a mere output system to a constitutive system for (some of the) central processes underlying thought and speech production. This resonates well with a wider movement in embodied cognitive science (Shapiro, 2010; Wilson, 2002) in which mental representations are thought to be multimodal (Barsalou, 1999, 2008; Svensson, 2007) and coupled to the body's current state (Glenberg & Kaschak, 2002).

In this article we focus on the possible intra-cognitive function of gestures, as opposed to their inter-cognitive or communicative function, which we will touch upon only briefly. That is, gestures seem to support internal cognitive processes of the gesturer (e.g., Goldin-Meadow, Nusbaum, Kelly, & Wagner, 2001; Marstaller & Burianová, 2013; Morsella & Krauss, 2004; Rauscher, Krauss, & Chen, 1996). We argue that the current theoretical "embodied" movement in gesture research has fueled the upsurge of inquiry into the beneficial role of gestures in cognitive processes such as speech and visuospatial cognition, but that this line of thought is underspecified with regard to explaining how gestures *as bodily movements* aid cognitive processing. In a sense, current perspectives on gestures are still *disembodied* and too *internalistic* because they seem to implicitly reduce gestures to *cognitively trivial* bodily outputs of (sensorimotor) neural precursors.

We seek to provide a more embodied account of gesticulation on the basis of recent philosophical and theoretical appeals within embodied cognitive science (e.g., Wilson, 2002) that focus on the possibly embedded/extended role of gestures (Clark, 2008, 2013; Kirsh, 1995; Wheeler, 2013), and a review of related empirical literature (e.g., Gray & Fu, 2004; Kirsh, 2009). This account is "more embodied" because embedded/extended perspectives traditionally seek to provide an anti-internalist perspective on cognition (e.g., Hutchins, 1995a), in which cognition is understood as being on-line, that is, being tightly coupled with, embedded in, if not extended over, the body and the environment (Shapiro, 2010). This stands in stark contrast with more internalist notions of embodiment that are currently dominating the gesture literature and that focus on decoupled, or "off-line" cognition and the sensorimotor nature of mental representations (Wilson, 2002). We suggest that the embedded/extended account of the cognitive function of gestures could be successful in explaining how gestures fulfill a cognitive function if it makes clear how gestures as self-generated bodily acts *generate and support* rather than execute thought processes (Clark, 2013). Therefore, we focus on the idea that gestures may at times serve as external tools of the cognitive system that *replace and support* otherwise solely internal cognitive processes. By reviewing research on the beneficial role of gesture production in (visuo-spatial) cognition (e.g., Chu & Kita, 2008; Delgado, Gómez, & Sarriá , 2011) and connecting the resulting insights with research on embedded cognition (e.g., Gray & Fu, 2004; Hutchins, 1995a; Kirsh & Maglio, 1994) we aim to contribute to a more embedded/extended account of gestures.

Before we will elaborate on the main goals of this paper, we need to point out what this article is not about. First, we do not suggest that current perspectives in the gesture literature are incorrect. In fact, our embedded/extended perspective is largely complementary to, and in some instances builds on, contemporary accounts of the function of gestures we review here. Second, although we argue in favor of a more embodied account of gestures and their cognitive function, this does not require us to make any additional, more radical, claims about the supposed sensorimotor nature of conceptual representations that are currently under discussion in the literature (e.g., Arbib, Gasser, & Barrès, 2014; Dove, 2010; Zwaan, in press). Third, we will not provide philosophical claims about whether gestures should be considered as an extended as opposed to an embedded cognitive phenomenon (e.g., Adams & Aizawa, 2001; Clark, 2008, 2013; Wheeler, 2013). That is, we do not make explicit claims about whether gestures as extra-neural events are part of the cognitive process (Extended claim) or whether gestures merely support internal cognitive processes but strictly speaking should not be considered as part of the cognitive process (Embedded Claim). Rather, we aim to provide an empirical view through the embedded/extended perspective, on the basis of the shared anti-internalist goal of these perspectives, by focusing on extra-neural factors that support, shape, and replace internal cognitive processes. We suggest that our embedded/extended account of the cognitive function of gestures can fill an explanatory gap in the current literature concerning the possible intra-cognitive role of gestures and is supported by extant findings.

This article is structured into four main sections. The next section reviews findings that show that co-speech and -thought gestures have a (beneficial) cognitive function (primarily in visuospatial cognition). Section three provides an overview of some important theoretical perspectives on the role of gestures in cognition. We suggest that the current theoretical perspectives on the function and underpinnings of gestures leave an explanatory gap concerning how gestures as external bodily acts might be conducive to internal cognitive processes. Having exposed the explanatory gap, we introduce an embedded/extended account of gestures (Clark, 2008; 2013) and provide a new interpretation of the research reviewed in the previous section in light of recent research in the field of embedded cognition (Ballard, Hayhoe, & Pelz, 1995; Gray & Fu, 2004; Kirsh, 2009; Kirsh & Maglio, 1994; Risko, Medimorec, Chisholm, & Kingstone, 2013). Finally, we summarize and discuss our main points.

The Function of Gesture: Empirical Evidence

The Inter-cognitive Role of Gestures

Before we consider evidence for the beneficial or supportive role of gestures for cognitive processes, it is important to acknowledge the evidence for the common assertion that gestures fulfill a communicative function. When speakers produce gestures, this seems to be intended to increase listeners' understanding of their message. Indeed, when speaker and listener are face-to-face, more gestures with semantic content are produced than when there is no visual contact (Alibali, Heath, & Myers, 2001). Also, when speakers are aware of listeners' knowledge gaps, they tend to convey the information unknown to listeners in both speech and gesture, while they tend to only use verbal information when relevant knowledge is already shared between the interlocutors (Holler & Stevens, 2007). These results suggest that speakers adjust their gestures for their listeners' benefit. And indeed, listeners' comprehension has been shown to improve by speakers' use of gestures from an early age on. For example, three to five year olds understand indirect requests (Kelly, 2001) and new abstract concepts (Valenzeno, Alibali, & Klatzky, 2003) better when the request is accompanied by deictic (i.e., pointing) gestures. In addition, preschoolers understand complex spoken messages better when these are accompanied by representational gestures (McNeil, Alibali, & Evans, 2000). Moreover, co-speech gestures do not only contribute to *what* is understood, but also to *how* something is understood. When deictic gestures are used, listeners are more likely to correctly interpret utterances compared to when the utterance was not combined with a gesture, suggesting that co-speech gestures play a role in pragmatic understanding. For example, when hearing the utterance "it's getting hot in here", people were sooner inclined to interpret this as an indirect request (i.e., could you please open the window) when the speaker pointed to the window, than when the speaker did not point, in which case the listener might interpret the utterance as a mere statement (Kelly, Barr, Church, & Lynch, 1999). All in all, there is a great deal of evidence for the contention that gestures fulfill inter-cognitive (i.e., communicative) functions (Goldin-Meadow & Alibali, 2012).

The Intra-cognitive Role of Gestures

There is mounting evidence that gestures fulfill intra-cognitive functions in addition to inter-cognitive ones. This is relevant to our present purposes. For example, co-speech gestures affect speakers' own cognitive processes. Several studies have suggested that lexical access is disrupted or promoted when gesticulation is prohibited versus allowed to naturally emerge. When speakers are prohibited from gesturing during speech with spatial content, they are less fluent than when gesticulation is allowed, suggesting that lexical access is disrupted (Morsella & Krauss, 2004; Rauscher et al., 1996; see however, Hoetjes, Krahmer, & Swerts, 2014). Moreover, speech is more fluent when co-speech gestures are produced and gesture rates are higher when lexical access is difficult (e.g., during the tip of the tongue phenomenon; Chawla & Krauss, 1994). Furthermore, when gesticulation is prohibited, the content of speech is less likely to be spatial in nature, suggesting that gestures support speech that is spatial in content (Rimè, Schiaratura, Hupet, & Ghysselinckx, 1984). Not only can online speech be influenced by co-speech gestures, these gestures can also have an influence off-line. For example, making gestures during the recollection of a previous event, can improve retrieval of details of that event compared to when gesticulation is not allowed (Stevanoni & Salmon, 2005). In addition, gesticulation prior to recalling previously learned words aids recall performance (De Nooijer, Van Gog, Paas, & Zwaan, 2013).

Gestures primarily arise during the processing of visuospatial information (e.g., Alibali et al., 2001; Allen, 2003; Kita & Özyürek, 2003; Seyfeddinipur & Kita, 2001). For example, people are more likely to gesture when describing visual objects from memory as opposed to when the object is visually present (Morsella & Krauss, 2004; Wesp, et al., 2001; see also Ping & Goldin-Meadow, 2010), although gesticulation also occurs in the presence of the to-be described object (Morsella & Krauss, 2004). Moreover, gestures occur more often when objects are difficult to describe in speech, such as complex, not easily describable drawings (Morsella & Kraus, 2004). Indeed, the emergence of gesticulation appears to be related to the cognitive demands of the task (Cook, Yip, & Goldin-Meadow, 2012; Goldin-Meadow, Nusbaum, Kelly, & Wagner, 2001; Marstaller & Burianová, 2013; Ping & Goldin-Meadow, 2010; Smithson & Nicoladis, 2014; Wagner, Nusbaum, & Goldin-Meadow, 2004). For example, participants who were given the dual task of remembering letters while explaining a difficult math problem, remembered more letters when they were allowed to gesture while explaining the problem than when they were not allowed to gesture (Goldin-Meadow et al., 2001). This suggests that gesticulation reduced the working memory load imposed by explaining the math problem, leaving more capacity available for performing the secondary task of remembering letters. Gesticulation when describing a mental rotation problem emerges primarily when describing the task-relevant rotation itself as opposed to describing the task-relevant static end-point of the rotation (Hostetter, Alibali, & Bartholomew, 2011). This finding suggests that it is the high spatial cognitive demand, which is arguably higher during dynamic spatio-temporal rotation as opposed to describing static spatial information, that invokes the use of gestures (see also Smithson & Nicoladis, 2014). Furthermore, it has been found that encouraging participants to gesture during a mental rotation task enhances their performance (Chu and Kita, 2011).

The findings described here primarily involved iconic gestures. However, even deictic (pointing) gestures occur more often when cognitive demand is higher. Infants and young children (between 1 and 2 years of age) sometimes point for non-communicative reasons (Bates, Camaioi, & Volterra, 1975; Delgado, Gómez, & Sarría, 2009). Furthermore, pointing gestures can aid the regulation of the speaker's attention in non-communicative and challenging problem-solving situations (Delgado et al., 2011). In two studies, children ranging in age from 2 to 4 years old saw a toy being hidden in one of three containers on a rotation table. This was followed by a delay of 45 to 60 seconds during which the children either had to remember where the toy was hidden by the experimenter (cognitive demand group) or had to wait for the experimenter to retrieve the toy for them. During the delay the experimenter left the room. Additionally, the difficulty of the memory task was varied for half of the trials such that the table was turned for 540 degrees. Analysis of the video-taped sessions showed not only that solitary pointing gestures occurred, but also that they occurred significantly more often in the cognitive demand condition than in the waiting condition (although no effects were found for task difficulty). A second experiment with children ranging from 4 to 6 years old who performed a picture-matching task showed that constraining gestures resulted in poorer performance on the task than non-constraining gestures, but only for children who habitually pointed in the constrained condition, suggesting a cognitively beneficial role of solitary pointing gestures. This finding is surprising because deictic gestures have primarily been considered as serving communicative functions (Tomasello, Carpenter, & Liszkowski, 2007). Additional research on pointing gestures was conducted in the context of keeping track of counting. Children, adults, and even primates effectively use the hands in counting objects by pointing and touching gestures as to mark counted objects, and synchronize with counting expressed in speech (Alibali & Dirusso, 1999; Boyson, Berntson, Shreyer, & Hannan, 1995; Kirsh, 1995). For example, participants who were allowed to use their hands for pointing during the counting of coins were faster and made fewer mistakes than those who were not allowed to use their hands (Kirsh, 1995). Thus, pointing gestures sometimes regulate visuo-spatial attentional processes, being especially helpful under high cognitive task demands.

These results converge with a recent correlational study that examined whether individual differences in spatial working memory capacity, spatial transformation ability, and conceptualization ability (amongst others) were associated with frequency of use of several types of gestures (Chu, Meyer, Foulkes, & Kita, 2013). Lower scores on all of these variables predicted higher frequency of spontaneously produced representational and conduit[10] gestures in a natural setting. Other evidence is consistent with this pattern. Particularly people with low working memory capacity are negatively impacted on a working memory task

[10]Defined as "iconic depictions of abstract concepts of meaning and language" (McNeill, 1985, p.357).

when they are not allowed to gesture as opposed to people with high working memory capacity (Marstaller & Burianová, 2013). Thus, in addition to the findings that gestures emerge during spatial information processing, gestures are also more likely to be produced by, and more likely to affect cognitive processes of, people with low spatial working memory and information processing ability (see also Chu & Kita, 2011).

Further evidence for gesturing as a compensatory mechanism comes from a study by Chu and Kita (2008).The type of spontaneous gestures that participants used during a mental rotation task followed a trajectory from external to more internalized solution strategies. That is, participants first gestured concretely *as if* manipulating the object to be rotated and subsequently changed their strategy and used their flat hand as stand-in for the object that needed to be rotated. Moreover, frequency of gesture use in aiding a spatial rotation task diminished over time, suggesting that cognitive operations became gradually internalized. A related phenomenon is that intermediate advanced abacus users use gestures during mental calculation. In the absence of the abacus, trained participants apply finger gestures as if manipulating an abacus ready to hand; but as abacus users become more advanced, they exhibit a reduced reliance on gestures during mental calculation (Hatano, Miyake, & Binks, 1977; Hatano & Osawa, 1983). In line with the findings of Chu and Kita (2008) this shows that the use of gestures becomes more infrequent as familiarity with the task increases.Moreover, when describing the solution of a particular spatial problem, people's gesticulation aligns with the medium that the problem has been introduced in (Cook & Tanenhaus, 2009). For example, participants who described solutions of the Tower of Hanoi with physical discs as opposed to a computer simulation tended to spontaneously produce gestures that aligned with the physical actions performed with physical discs.

Thus, if we consider (a) that working memory capacity is limited, and (b) that new tasks often impose a higher working memory demand that diminishes as the learner becomes more experienced with a task (e.g., Chase & Ericsson, 1982; Kalyuga, Ayres, Chandler, & Sweller, 2003) then the findings we just reviewed suggest that gestures are likely to emerge in novel situations so as to provide the cognizer with some kind of external support. We will discuss the nature of this external support in our embedded/extended account of the cognitive function of gestures.

Finally, gestures can aid in acquiring a solution during problem solving (Alibali, Spencer, & Kita, 2004; Boncoddo, Dixon, & Kelly, 2010; Stephen, Dixon,

& Isenhower, 2009). For example, participants were presented with two glasses with differing widths and equal heights and were asked to imagine the glasses being filled with water to the same level. Participants judged whether the water would spill when glasses were rotated at equal angles (Schwartz & Black, 1999). Participants were able to predict the answer correctly much more often when rotating the empty glasses with their eyes closed, compared to when they were only allowed to think about the solution (i.e., mentally rotate). Although the previous study was in a sense a form of direct action (by allowing the objects to be manipulated), there is evidence that suggests that gestures, as non-direct manipulations, equally support the use of particular problem-solving strategies. For example, a study in which participants were presented with an interlocking gear problem (Alibali et al., 2004) found that they judged the direction of movement of a gear through different strategies, depending on whether or not gesticulation was allowed. When they were allowed to gesture, participants were more likely to simulate the rotations of each gear by finger gestures in order to provide the solution of the end-gear's rotational direction (depictive strategy), whereas participants who were prohibited from gesticulation were more likely to achieve the solution through the parity rule (direction gear *x* has the same direction as gear *x*+2). Note that the participants who used the depictive strategy were not better at the task than those using the parity rule (Alibali et al., 2004; also see Hegarty, Mayer, Kriz, & Keehner, 2005). Indeed, the parity rule strategy is generally considered to be the most effective strategy (Boncoddo et al., 2010). It is interesting in this regard to note that preschoolers are more likely to achieve understanding of the parity rule through gesticulation (Boncoddo et al., 2010). That is, preschoolers who used more gestures supporting a depictive strategy, more efficiently acquired a strategy based on the parity principle, in comparison to preschoolers who gestured less. Thus in this particular instance, the repeated use of gestures by participants is more likely to lead to discovery of new strategies during problem-solving although the use of gestures does not necessarily invite learners to adopt the most efficient strategy (see also Stephen et al., 2009).

The research reviewed here provides evidence that gestures have an intra-cognitive cognitive function for the gesturer. Furthermore, it produces two intriguing and related questions that we think need to be answered in a theoretical account of the cognitive function of gesticulation. First, why do gestures occur more often when cognitive demand is high? Second, why are spatial cognitive ability and working-memory capacity negatively related to the use of gestures?

Current Theory About the Origin and Function of Gesture

In this section we will discuss several prominent accounts that aim to elucidate the underlying mechanisms and function of gestures, most prominently the Gesture-as-Simulated-Action account (GSA; Hostetter & Alibali, 2008) and subsequently the Lexical Gesture Process Model (LGP; Krauss, Chen, & Gottesmann, 2000), the Information Packaging Hypothesis (IPH; Kita, 2000), and the Image Maintenance Theory (IMT; Wesp, Hesse, Keutmann, & Wheaton, 2001). We evaluate these models directly after summarizing their main points, by assessing their explanatory power regarding the question: How do gestures-as-bodily-acts support cognitive processes?

We have chosen to address this collection of accounts for several reasons. The GSA account is a prominent contemporary account that attempts to integrate the literature of embodied cognition and the literature on gesture into a single perspective. Yet, as mentioned in the introduction, it seems that this attempt has resulted in a "disembodied" perspective on gesticulation. The other accounts have been very influential in elucidating the cognitive function of gestures. Moreover, they differ significantly from the GSA account but also from each other. The result is a representative (but not exhaustive) overview of theories about the possible cognitive function of gestures.

Gesture-as-simulated-action (GSA) Account

The GSA account (Hostetter & Alibali, 2008) relies heavily on the insights from embodied cognition that representations are based on the sensorimotor system (Barsalou, 1999, 2008; Glenberg & Kaschak, 2002). This embodied view is supported by mounting evidence that perceptuo-motor faculties of the brain are activated during concrete but also supposedly symbolic and abstract conceptual processes (e.g., Barsalou, 2008; Pulvermüller, Moseley, Egorova, Shebani, & Boulenger, 2014). For example, merely reading words that have olfactory, gustatory, or motor connotations (e.g., garlic, jasmine, salt, sour, kick, pick) as opposed to reading neutral words, activates brain regions that are involved in smelling, tasting, and moving (Barrós-Loscertales et al., 2011; Gonzales et al., 2006; Hauk, Johnsrude, & Pulvermüller, 2004).

The GSA approach predicts that cognitive processes, such as conceptual processing, co-occur with sensorimotor reactivations. More importantly it is contended that meaningful cognitive processing is *dependent* on these reactivations or simulations of sensorimotor states (Barsalou, 2008; Hostetter & Alibali, 2008).

Indeed, conceptual processing is hampered when participants are primed with inconsistent perceptual or motor information (e.g., Glenberg, Havas, Becker, & Rinck, 2005; Kaschak, Zwaan, Aveyard, & Yaxley, 2006). For example, participants are quicker in verifying the sensibility of sentences (such as "Andy delivered the pizza to you vs. You delivered the pizza to Andy") when their response actions were consistent with the implied motion of the sentences (moving the hand forward or backward), whereas they were slower when the movement contrasted with the implied motion (Glenberg & Kaschak, 2002). As such, it is suggested that induced sensorimotor states impinge on conceptual representational states since both systems are tightly coupled (Barsalou, 2008).

Hostetter and Alibali (2008) have suggested that the phenomenon of co-speech and co-thought gestures fits nicely with the idea that cognitive processing depends on activations in the sensorimotor system. In fact, according to the GSA account gestures *are* the bodily realizations (or as they call it, "visible embodiments") of otherwise covert sensorimotor activations. The main question that the GSA account aims to address, therefore, is how sensorimotor activations come to be reflected in gestures. Hostetter and Alibali (2008, p. 503) first provide a simple answer: "Simulation involves premotor action states; this activation has the potential to spread to motor areas and to be realized as overt action. When this spreading activation occurs, a gesture is born". More specifically, the GSA account suggests that gestures emerge through sensorimotor re-activations underlying thought and speech processing that 'leak into' the motor-executive system:

> "As an analogy, we might imagine activation spreading from premotor areas to motor areas through a gate. Once the gate is opened to allow more activation for one task (speaking), it may be difficult to inhibit other premotor activation (that which supports gestures) from also spreading through the gate to motor areas, the activation for the simulations 'rides along' and may be manifested as a gesture" (Hostetter & Alibali, 2008, p. 505).

Hostetter and Alibali (2008) further propose three underlying factors that determine when gestures are likely to occur. First, the strength of the particular perceptuo-motor activation must surpass a certain *gesture threshold* for actual physical embodiment (i.e., gesticulation) to arise. This activation strength is dependent on the degree to which speakers evoke visuospatial imagery during conceptual processing. For instance, they argue that the same conceptual content can be processed verbal-propositionally or with visuo-spatial imagery (e.g., in the case of route-descriptions), the latter type of encoding being more likely to evoke gesticulation (e.g., Alibali et al., 2001; Allen, 2003; Kita & Özyürek, 2003; Seyfeddinipur, & Kita, 2001). Second, visuo-motor simulations are likely to evoke gesticulation when the conceptual content that is being processed involves an action. For example, talking about action is likely to evoke gestures because it is dependent on motor-information (Alibali & Hostetter, 2008). Third, it is speculated that the height of speakers' gesture-threshold can vary across individuals and situations. To illustrate, a higher degree of neural interconnectivity between pre-motor and motor areas may lower the gesture threshold of a particular individual. Furthermore, inhibiting gesticulation requires cognitive effort and as such the threshold might be lowered when cognitive load is high (e.g., Goldin-Meadow et al., 2001).

Explanatory power of the GSA account

So how does the GSA account answer our question of how gestures-as-bodily-acts support cognitive processes? First, it is held that speech production and thought processes are dependent on the conceptual system recruiting sensorimotor representations. Furthermore, according to Hostetter and Alibali (2008), gestures arise from and are dependent on the strength of sensorimotor activations. However, the model does not allow the conclusion that gestures-as-bodily-acts aid cognition, because gestures only *execute* sensorimotor information, they do not *produce* it. The sensorimotor information that is produced (e.g., proprioceptive and visual consequences of movement) does not fulfill a cognitive function in the GSA account. This is indicated by the motor-leakage metaphor, as gestures simply "ride along" with sensorimotor activations (Hostetter & Alibali, 2008, p. 505) and can be understood as a mere "outgrowth" (Risko et al., 2013) or "visible embodiments" (Hostetter & Alibali, 2008) of internal embodied simulations. Thus, the GSA account leaves us with the question why do cognitive processes sometimes recruit the body (gestures), as opposed to relying on purely internal mechanisms? Furthermore, what is the explanatory power of the GSA account in terms of the empirical literature on the cognitive function of gestures provided above? Most notably, why is high cognitive demand result in more use of gestures. This is explained by the GSA account in "that inhibiting activation from spreading to a gesture requires more cognitive resources than does producing the gesture" (Hostetter & Alibali, 2008, p. 505). From this point of view, gesticulation is the default and is simply hard-wired with cognitive processes. By accepting this, we would simply deflate the idea of there being any function of gestures as bodily acts, endow the cognitive system with functionally unnecessary expenditure of energy (hand-movements), and allow only a negative cognitive effect of not gesturing. Although this idea of costly active inhibition may very well be a correct explanation for some instances of gesticulation, we think its possible scope for explaining the function of gesture is somewhat reduced by the realization that possessing a superfluous and energy-demanding gesture system does not seem very adaptive or flexible. Moreover, we think that a non-deflationary account of the function of gesture is possible and in fact more promising for understanding the empirical findings on the cognitive function of gestures reviewed in this paper.

Lexical Gesture Process Model

The lexical gesture process model proposed by Krauss, Chen and Gottesman (2000) tries to explain why speech might be facilitated by gesticulation. According to this theory, gestures do not only fulfill a communicative role, but may serve to facilitate lexical retrieval on the part of the gesturer as well. Gestures that share features with the lexical semantic content of the word will facilitate lexical access. Krauss and colleagues (2000) hypothesize that this is the case because gesturing results in "cross-modal priming" in which features of the concept represented by the gesture can facilitate lexical retrieval. According to this Lexical Gesture Process (LGP) account, gesture production draws upon the activated representations in working memory that are expressed in speech. The assumption is that the content of conceptual memory is encoded in multiple ways, and that activation of one representational format can spread to activation in another representational format. In this account gestures derive from non-propositional representational formats (mostly visuo-spatial), as opposed to speech, which draws on propositional symbolic formats. LGP further suggests that non-propositional information becomes expressed in speech through a spatial/dynamic feature selector that transforms spatially and dynamically formatted information into a set of "abstract properties of movement". The abstract specifications are then translated into a motor program by a motor planner. Motor systems output the set of instructions from the motor planner and the gestural movement is monitored kinesthetically. The motoric features that are picked up by the kinesthetic monitor promote retrieval of the concept for speech through *cross-modal priming*. Krauss and Hadar (1999, pp. 21) specify:

> "the spatio-dynamic information the gesture encodes is fed via the kinesic monitor to the formulator, where it facilitates lexical retrieval. Facilitation is achieved through cross-modal priming, in which gesturally represented features of the concept in memory participate in lexical retrieval. Of course, it is possible to locate the site of gestural input more precisely (e.g., the grammatical encoder or the phonological encoder)".

Explanatory power Lexical Gesture Process Model

Does LGP allow for a cognitive role of gestures-as-bodily-acts? That is, does it answer the question why gestures are produced, and how they are cognitively relevant? An affirmative response is appropriate, although the mechanism seems underspecified and unparsimonious. Indeed, when a gesture is outputted by the motor-system, the "kinesthetic" feedback that is produced acts as input to the formulator (i.e., the grammatical or phonological encoder or both) and can then facilitate lexical selection by way of additional cues or "cross-modal priming". Thus, in this model, motor-information is externalized and is fed back into the system to promote lexical retrieval through supporting the processes of the "grammatical encoder" and the "phonological encoder". Yet the question remains why this motor-information needs to loop out of the brain and then be retrieved again by the kinesthetic monitor. According to LGP, gesture will only facilitate lexical access when the gesture features match the lexical semantic content of the concept. Therefore, gestures will only facilitate lexical access when the kinesthetic information that was already present in a verbal form is fed back into the formulator. Thus it seems that the brain is "primed" with information that is already present in the internal system, given that gestures are outputs of an already constructed motor program. Thus, it is unclear with what kind of information the cognitive system is primed. Of course, gestures might indeed fulfill this function, but the model currently presented is not very illuminating why and how gestures-as-bodily-acts fulfill a cognitive function. So, although LGP also suggests an intra-cognitive role for gestures, it is still difficult to appreciate the added value of the kinesthetic information that is fed back into the system with regard to cognitive processing.

Information Packaging Hypothesis

A third prominent theory in the gesture literature is the Information Packaging Hypothesis (IPH; Kita, 2000). This theory proposes that gestures aid speech production by breaking images into smaller bits to enhance the verbalize-ability of communicative content. A key idea is that there are two modes of thinking that tend to converge during the linguistic act. There is analytical thinking as opposed to spatio-motoric thinking from which gestures follow, which involves the organization of information through hierarchical structuring and involves decontextualized conceptual templates. According to Kita, these templates can be non-linguistic (in the case of scripts), or linguistic, such as in the case of a lexical item's semantic and pragmatic specifications. The templates are not multimodal as in the case of the GSA account, thus they do not involve "activation of 'peripheral' modules" (Kita, 2000, p. 164), yet can be translated into the other mode of thinking, which is spatio-motoric thinking. The spatio-motoric mode of thinking constitutes gestures and involves information organized in action schemas. Gestures should be considered as actions in a virtual environment, and are derived from practical actions.

A core idea behind IHP is that the two modes of thinking collaboratively organize information during speaking. Kita (2000, pp. 163) suggests that (a) "The production of the representational gesture helps speakers organize rich spatio-temporal information", (b) "Spatio-motoric thinking, which underlies representational gestures helps speaking by providing an alternative informational organization that is not readily accessible to analytic thinking" and (c) "Spatio-motoric thinking and analytic thinking have ready access to different sets of informational organizations. However, in the course of speech production, the representations in the two modes of thinking are coordinated and tend to converge."

Explanatory power Information Packaging Hypothesis

Does IPH have explanatory power of how gestures-as-bodily-acts support cognitive processes? The IPH does provide a clear account of how gestures aid the "packaging of information" given that gestures are considered as the result of spatio-motoric thinking that is already internally realized. That is, just like the GSA, the IPH seems to regard gestures as mere output of spatio-motoric thinking, with the latter having the actual cognitive function (information packaging). Even if we allow for a possible different reading of the information packaging hypothesis, in which gesticulation actually supports spatio-motoric thinking, the IPH account does not go into any detail about how gestures-as-bodily-acts feedback to or support internal cognitive processes to perform the function of spatio-motoric information packaging.

Image Maintenance Theory

The final theory under review here is the Image Maintenance Theory (IMT) by Wesp and colleagues (2001). Although this theory is only briefly presented in an empirical paper it has become an influential view on the cognitive role of gestures (Alibali, 2005). Arguably, the main thesis of the image maintenance theory, which is often contrasted with the LGP, is "that gestures are not directly involved in the search for words; rather, they keep the non-lexical concept in memory during the lexical search, a process of data maintenance not unlike that needed in other problem-solving activities" (Wesp et al., 2001, p. 592). This is further explained; "a prelinguisitic representation of spatial information is established through spatial imagery and maintenance of these spatial images is facilitated by gestures" (Wesp et al., 2001, p. 595). Wesp and colleagues (2001) base this idea on the idea that spatial information is held in the visuospatial scratchpad of working memory (Baddeley, 1986). The items (visuospatial information) in the scratchpad decay rapidly and must be rehearsed to be maintained in working memory. Just like articulatory loops, gestures serve the function of "refreshing" the visual scratchpad to sustain activation of the image in working memory. Importantly, gestures are therefore not necessary for lexical retrieval but may indirectly facilitate it through, "motoric refreshing" of the image (p. 597).

Explanatory power Image Maintenance Theory

Does the IMT have explanatory power of how gestures-as-bodily-acts, support cognitive processes? The answer is yes, although much is still needed to understand its function. "Yes" because the IMT suggests that the production of a *physical* gesture supports the maintenance of an internal spatial image (a cognitive process); without the physical gesture the internal spatial image becomes unstable and its activation is likely to decay. Yet, Wesp et al.'s (2001) account does not provide sufficient detail beyond this notion. How do gestures refresh motoric spatial images? What is the mechanism by which gestures-as-bodily-acts refresh motor spatial images? Furthermore, aren't gestures redundant given that they provide the gesturer with information that is already present in the system that outputs the gestures (e.g., visual information)? Although these questions remain unanswered, of all the accounts presented here, the IMT is most compatible with an embedded/extended account that assumes gestures are cognitively relevant because they are bodily.

Summary of Findings From the Theoretical Overview

In the previous subsections we have discussed four models that have been put forth to explain the underlying mechanisms of gestures. We sought an answer to our question: How do gestures-as-bodily-acts support cognitive processes? Our review of the literature suggests that the cognitive function of gestures-as-bodily-acts cannot be adequately explained, or remains underspecified, in several different theories about the underpinnings and functions of gestures. In the GSA account gestures are seen as by-products of sensorimotor activation but cease to be supporting cognition the moment they are outputted by the motor-system. The IPH suggests that gestures help package the spatio-motoric thinking during speech, yet this account also assumes that gestures are the result of these processes as they are the realizations of spatio-motoric internal processes; they are pre-packaged the moment they are externalized as gestures and do no packaging of their own. In the LGP account, the gestures that are produced are fed back into the cognitive system to provide it with cross-modal primes. As such, gestures, as physical acts, attain a function. Yet, the LGP account is unclear about what exactly is primed, or what novel information gestures provide to the system, that was not already activated or present. Interestingly, the IMT does seem to ascribe a definite cognitive function to gestures by positing that they support the maintenance of mental images.

It is important to stress that our review is aimed at answering a specific question that may be different from the questions that the theories we discussed were designed to address. We have only considered these theories' explanations (*explanantia*) of a particular aspect of gesticulation that we think needs to be explained (*explanandum)*, namely how gestures-as-bodily-actions have a cognitive function. This means that we do not suggest that the theories under discussion are wrong, nor do we suggest that they are incompatible with the upcoming perspective; rather the *explanantia* they offer are not (yet) suitable to cover the *explanandum* that is the focus of the current paper. In the next section we aim to fill this explanatory gap through a more embedded/extended perspective on the cognitive function on gestures.

Towards a more embedded/extended perspective to the cognitive function of gestures

In this section we attempt to answer the main question of how gestures can fulfill cognitive functions. In the following subsection we will briefly introduce the embedded/extended cognition perspective (inspired by Clark, 2013), which is followed by a representative overview of research in this domain. Subsequently we apply the relevant theoretical and empirical findings to the cognitive function of gestures, which yields challenges and hypotheses for future research.

An embedded/extended perspective: Theory and research

Embedded/extended cognition is considered part of the broader development of embodied cognitive science (Shapiro, 2010; Wilson, 2002) and has its roots (amongst others; Gallagher, 2009) in situated cognition (Bredo, 1994), robotics (Brooks, 1991) and the dynamical systems approach to cognition (Chemero, 2009). According to a loose description of "the" embedded/extended perspective on cognition (cf. Wilson, 2002), the main thesis is that the cognitive system is a coupled brain-body-world system (Clark, 2008; Wheeler, 2007). As such, cognition involves an ongoing transaction between current states of the brain, body, and the environment (Clark, 2008). Within this view, the classic internalist picture of cognition is disputed; thinking is something we do, rather than something that simply happens within us. Understanding cognition, therefore, requires a broader level of analysis that allows the study of how we use our body and the world during the unfolding of cognitive processes. For example, Hutchins (1995b) analyzed the goings-on of commercial airlines and suggested that a purely internalist perspective was ill-suited to understand its workings; flying a plane involves task-relevant information that is neither fully instantiated in the cockpit, the pilot, or co-pilots, it is rather distributed among them and all parts work together (see also Hutchins, 1990a). Everyday examples of embedded/extended cognitive phenomena would be, for instance, asking another person to remind you of something, using a tall building for navigating your way home, or reducing working memory load by taking notes during a conversation. Or in the case of drawing: "One draws, responds to what one has drawn, draws more, and so on. The goals for the drawing change as the drawing evolves and different effects become possible, making the whole development a mutual affair rather than a matter of one-way determinism" (Bredo, 1994, pp. 28).

In philosophy there is a debate on whether states of the body and the environment can be considered extra-neural contributors to cognition (Wilson, 2002), or in a more radical reading, external vehicles of cognition (Clark, 2008; Clark & Chalmers, 1998). According to the radical extended perspective, the internalist view is provoked by the classic thesis that "If, as we confront some task, a part of the world functions as a process which, were it to go on in the head, we would have no hesitation in accepting as part of the cognitive process, then that part of the world is (for that time) part of the cognitive process" (Clark & Chalmers, 1998, p.8). The less radical thesis, the notion of embeddedness, also stresses a tight coupling between the agent and the world and suggests that the body and environment can, often in unexpected ways, causally impact cognition, yet suggest that the body and the environment are not part of cognition (Adams & Aizawa, 2001; Rupert, 2009). Thus the difference between embedded and extended cognition is whether extra-neural conditions causally impact cognition (embedded thesis) or are constitutive of it (extended thesis). As mentioned in the introduction, we will side-step this technical debate; for our present purposes it suffices to say that we follow the joint anti-internalist approach of embedded *and* extended cognition, which suggests that the cognitive system works in concert with the body and the environment.

The embedded/extended perspective has given rise to a large amount of empirical research on the way the cognitive system uses the body and the environment (e.g., Ballard, et al., 1995; Fu, 2011; Haselen, Steen, & Frens, 2000; Kirsh & Maglio, 1994; Martin & Schwartz, 2005; Risko, et al., 2013; see also Pouw, Van Gog, & Paas, 2014). A seminal study by Kirsh and Maglio (1994; see also Stull, Hegarty, Dixon, & Stieff, 2012) found that expert Tetris players make more use of *epistemic actions*; actions that uncover (hidden) information that is cognitively demanding to compute. These types of actions are different from actions that bring one closer to one's goal (pragmatic actions). For example, advanced players, instead of rotating 'zoids' (i.e., falling block arrangements in Tetris) through mental simulation to judge whether it will fit the zoids in the bottom deck, they preferred rotating them physically as this allowed a direct matching of orientation and fit. The cognitive operation of rotation to determine a possible fit was thus off-loaded onto the environment.

Another classic study (Ballard et al., 1995; see also Haselen et al., 2000; Hayhoe, Pook, & Rao, 1997) showed that the cognitive system opts for retrieving information just-in-time, thereby minimizing constrains on working-memory. Participants were asked to recreate a configuration of colored blocks from a *model* by picking up colored blocks from a *resource space* and putting them in a *work-space*. The model, resource-, and work-space were all displayed in front of the participants. Eye-movement data were collected during this task. Participants made many switches of eye fixations between the model, work and -resource space. This indicated that participants adopt a 'minimal memory strategy' in which information is gathered incrementally as opposed to memorized in one fell swoop. Instead of memorizing the position and color all at once, participants first memorized the color to be searched from the model, then after finding a color match in the resource space, looked up the position of the block of the model. Thus, information is gathered just in time to minimize working memory constraints (see also Cary & Carlson, 1999, who obtained similar results in an income calculation task).

Yet, findings indicate that the cognitive system does not seem to have an a-priori preference for using the environment rather than internal cognitive resources in solving a cognitive problem; which strategy is adopted depends on the context. For example, when Ballard and colleagues (1995) increased the distance between the workplace and the model, participants were more likely to adopt a memory-intensive strategy. This finding resonates with the study by Gray and Fu (2004; see also Fu, 2011) in which participants were confronted with the task of programming a simulated VCR. In this task, retrieval costs of attaining task-relevant information were subtly manipulated. That is, the ease of retrieval was manipulated in such a way that participants could either acquire the information through a simple glimpse or through performing an additional mouse-click to make the information available. The cognitive strategy that the subjects chose changed as a function of the ease of retrievability. When external information was directly accessible, participants primarily relied on retrieving information externally. Attaining this "perfect-knowledge-in-the-world" was shown to be a reliable strategy, as it reduces the number of mistakes made during the task. Moreover, when the information was only indirectly available, participants were more likely to rely on internal memory, which produced a larger number of mistakes. The reason why participants in this condition relied on "imperfect-knowledge-in-the-head" was that the internally stored information was more quickly available compared to externally available information, as was predicted by a computational model that expressed the amount of time it takes to retrieve or recall information. Thus people seem to opt for the quickest problem-solving strategy in which the cognitive system "tends to recruit, on the spot, whatever mix of problem-solving resources will yield an acceptable result with a minimum of effort" (Clark, 2008, pp. 13).

Situational constraints bring about a trade-off decision whether the cognitive system relies on computation performed "on-line" (with the environment) or "off-line" (internally) (Wilson, 2002). Relevant in this regard is a recent set of experiments conducted by Risko and colleagues (2013) in which participants were presented with a varying number of letters that were either presented upright or tilted at 45 or 90 degrees. Participants spontaneously rotated their head, which indeed seemed to promote readability of tilted presentation of letters. Furthermore, participants were more likely to rotate their head when more letters were presented and tilt of the letters was more extreme, indicating that head-tilting (which they call external normalization) occurs when the cognitive demand of not tilting the head by means of "internal normalization" increases (more cognitive effort to read more letters in tilted position, and more extreme tilt of the letters). Thus, when internal computational demand increases, an externally mediated cognitive strategy becomes more attractive. This was also found in a study by Kirsh (2009), in which participants played a mental tic-tac-toe game with the experimenter. During the mental tic-tac-toe game participants have to keep their own "moves" and those of the opponent, in mind. In the critical conditions, participants were given a sheet of paper with a tic-tac-toe matrix depicted on it or a blank sheet. External support of a tic-tac-toe matrix aided participants' efficiency of playing the game in comparison to having no support or a white sheet. Apparently, participants are able to project the progression of the moves on the matrix through visual simulation. This is very similar to chess-players who think through moves on a chess-board without manipulating the board (Kirsh, 2009). Interestingly however, the external support was only beneficial when the tic-tac-toe game was complex (4x4 matrix as opposed to a 3x3 matrix), and especially for participants who scored low on spatial ability. Thus, this study suggests that projection on external support is especially helpful when cognitive demand is high, and relatedly, primarily for those who are low in spatial cognitive ability.

As a final example, the study conducted by Martin and Schwartz (2005) shows how active manipulation of the environment may foster learning through exploration of the solution space. In two studies, children (nine to ten years old) were learning how to solve fraction operator problems (e.g., one-fourth of eight candies), using physical tiles and pie-wedges that were movable *and* in another set of trials, using line drawings of pies or tiles which they could highlight and circle with a pen. The difficulty that children often experience in this task is that they focus on the numerator, leading them to understand "one-fourth of eight candies" to be "one candy." Martin and Schwartz predicted that physical interaction with manipulable objects would increase the chance that children come to interpret that one-fourth of eight means four groups of two because rearranging the tiles results in new groupings. Thus they reasoned that the agent and the environment mutually adapt each other (as in the case of drawing), where one acts without a preconceived goal on the environment which in turn feeds back information that might align with the correct solution. Indeed, children performed better with manipulable objects than without them (Experiment 1 and 2). Interestingly, presenting the children with the correct organization of tiles did not aid understanding; rather the physical open-ended interaction with the environment drove understanding and performance on the task (see also Manches, O'Malley, & Benford, 2010).

Let us summarize. First, the cognitive system makes use of the environment to distribute computational load but also to enable exploration of a problem-space that is difficult to enact off-line (i.e., to achieve through purely internal computations). Moreover, the cognitive system is not a-priori driven to reduce internal computational load by off-loading onto the environment, rather the environment is exploited if it offers a cheaper resource than internal means of computation to achieve an acceptable performance on a task (Gray & Fu, 2004). Although not conclusive, it further seems that when cognitive demand is high, either due to external constraints (higher cognitive load of the task) or internal constraints (e.g., low visuospatial cognitive ability) the cognitive system is more likely to opt for and benefit from external computational strategies. However, these findings do not allow us to draw definitive conclusions about when and how the cognitive system trades external with internal computational resources. Thus one of the major challenges for research in embedded/extended cognition is to determine which external (e.g., availability of external information) and internal (e.g., working memory ability) constraints affect whether and how problem-solving strategies become externally or internally mediated (Risko et al., 2013).

Furthermore, is it possible to identify a trajectory of problem-solving strategies as expertise develops? Specifically, does the cognitive system first rely on external support—given that it is still ill-equipped to perform stand-alone internal computations—and are computations increasingly performed off-line when the cognitive system becomes more equipped (e.g., because of acquired strategy knowledge or chunking mechanisms) to hold task-relevant information internally? Even though such questions cannot yet be answered by the embedded/extended cognition frameworks, it is not difficult to see the relevance of this framework for gesture research; there is a clear analogy between these findings and the findings from some of the gesture studies reviewed in the section on "the intra-cognitive role of gestures".

An embedded/extended perspective on the cognitive function of gestures

Recently, Clark (2008, 2013; see also Wheeler, 2013) provided a purely extended perspective on gesticulation. Clark (2013) provides a detailed discussion of why gestures should be seen as constitutive to—as opposed to merely causally impinging on—cognitive processes (cf. Wheeler, 2013). Here we only briefly address his account to further develop an embedded/extended perspective that is able to provide an explanation of the empirical data on the cognitive function of gestures as well as produce hypotheses and identify challenges for further research.

According to Clark (2013) we should *not* understand the cognitive role of gestures purely in terms of its neural pre- and post-cursors:

> "The wrong image here is that of a central reasoning engine that merely uses gesture to clothe or materialize performed ideas. Instead, gesture and overt or covert speech emerge as interacting parts of a distributed cognitive engine, participating in cognitively potent self-stimulating loops whose activity is as much an *aspect* of our thinking as its *result*" (p. 263).

Furthermore, he states that:

> "The physical act of gesturing is part and parcel of a coupled neural-bodily unfolding that is itself usefully seen as an extended process of thought." (p. 257)

Clark further argues that by producing a gesture, something concrete is brought into being (arm posture) that subsequently affects ongoing thinking and reasoning. Much like using a notepad, gestures provide a *stable* physical presence that embodies a particular aspect of a cognitive task. We can appreciate Clark's point if we consider that speech dissolves in midair and working memory allows only for a certain amount of thoughts to be consciously entertained. We can argue that gestures are not only a way to externalize speech and thought content, but also allow for temporal cognitive stability that might be more reliable than internal means of temporal cognitive extension (e.g., consciously attending to a thought to keep in mind).

Thus the key to an embedded/extended perspective on gestures is the view that gestures fulfill a cognitive function *because* they are bodily. That is, in contrast to what the GSA and the IPH propose, gesticulation produces an external physical presence that somehow supports internal cognitive processes. According to Clark's (2013) purely extended account, this physical presence instantiated in gesture is actually part of thinking itself. Indeed, he thinks that a more moderate account of gestures' function in which they merely *affect* inner neural cognitive processes is misconstrued. His argument for an extended cognitive understanding of gestures relies on the appreciation that some crucial forms of neural activity arise in coordination with gestures, wherein gesture and neural activity are interdependent in achieving a particular cognitive state. Thus although, in some instances "'neural goings-on' may be sufficient for the presence of some cognitive state or the other" in other instances gestures, at times, should be given a genuine cognitive status (p. 261) because "gesture and speech emerge as interacting parts of a cognitive system" (p. 263) whereby no meaningful categorization can be made of what should be considered cognitive or non-cognitive on the basis of the distinction between inner (neural activity) and outer (gestures).

How and when do these specific physical conditions fulfill a supporting role for a particular cognitive function? It is instructive to compare the research from the embedded/extended cognition tradition with research on the cognitive function of gesture. We need to reconsider the research by Kirsh and Maglio (1994), which showed that expert Tetris players operate on the environment to alleviate internal computational load (epistemic actions). Determining where a zoid fits is not dependent on internally computed rotations of the zoid, but is achieved by actual rotation of the zoid. In mental rotation tasks in which participants have to judge whether a 3-d zoid matches one out of several 3-d zoids depicted in different rotational angles (classic S-M cube task; Shepard & Metzler, 1971), participants use gestures to aid in their judgments (Chu & Kita, 2008; 2011). We would submit, that gestures in this case *are* epistemic actions that reveal information that is hidden (since the 3-d zoids do not rotate by themselves) and difficult or more costly to compute internally. Chu and Kita (2008) also found that when participants first approach the mental rotation task they are more likely to use hand-movements *as-if* actively rotating the block. We would speculate that in this case gestures fulfill the function of providing a physical platform that supports the internal representational stability (a term earlier used by Hutchins, 2005) of a rotating 3-d zoid (see also Pouw et al., in press). In this case the zoid is visually 'projected' into the hands (Kirsh, 2009) and is manipulated *as if* it were actually in the hand. In this case the hands offer a reliable external support for performing the cognitive function of rotating the projected 3d-zoid through gestures. Furthermore, using pointing gestures to keep track of something in the environment similarly produces a reliable physical attentional marker that alleviates internal attentional tracking processes (e.g., Delgado et al., 2011; Kirsh, 1995). This might also be the case with abacus users doing mental calculations that perform gestures on, what seems to be, a mentally projected abacus (Hatano et al., 1977; Hatano & Osawa, 1983). In this case physical gesticulation seems to be preferred by these users as opposed to internally simulating changes on the abacus. We would argue that because gestures allow a stable external physical presence, they support internal representational stability of the dynamically changing abacus during calculation. In line with Kirsh (2009), we argue that in these cases the cognitive system seems to be neither purely off-line nor online; rather, it uses partly environmental resources (e.g., gestures) and internal cognitive resources (e.g., visual simulation) to perform a task. Gestures are essentially a way to put on-line extra-neural resources into the mix of problem-solving resources.

Another possible embedded/extended function of gesture is exploration of a problem space. Schwartz and Martin (2005) found that manipulation of objects promoted the understanding of fraction-operating principles. Relevantly, gesturing might sometimes allow the gesturer to become aware of structural correlations that would be difficult to generate through internal computation. For instance, this seemed to be the case in the rotating-gear problem, in which the number gestures used that simulated each rotation of a gear predicted the discovery of a more efficient problem-solving strategy that involved pick-up of the regularity that each gear N+2 rotates in the same direction (Delgado et al., 2011).

With regard to when gestures emerge to fulfill an embedded/extended function, the research that we have discussed in the domain of embedded/extended cognition has another interesting alignment with the gesture literature. We can summarize both streams of findings in one converging main principle:

> *When the costs of internal computation are high, either induced by external constraints (higher cognitive demand of the task; more cost of retrieving information from the environment) or internal constraints (e.g., lower working memory ability) the cognitive system is more likely to adopt, if cheaply available, an externally supported problem-solving strategy; be it the environment or gestures* (Cook et al., 2011; Goldin-Meadow, et al., 2001; Gray & Fu, 2004; Kirsh, 2009; Marstaller & Burianová, 2013; Ping & Goldin-Meadow, 2010; Risko et al., 2013; Smithson & Nicoladis, 2014; Wagner et al., 2004).

In other words, "cognitive processes flow to wherever it is cheaper to perform them" (Kirsh, 2010, p. 442). Understood in this manner, it is not surprising that people who are describing a physical object tend to gesture less when the object is present as opposed to absent (Morsella & Krauss, 2001), since the task-relevant information is cheaply available in the environment. Or that gestures are more likely to be used to lighten the cognitive load when pressure is put on internal computational system (cognitive demand of the task) (e.g., Goldin-Meadow, 2001; Smithson & Nicoladis, 2014).

This embedded/extended perspective on the cognitive function of gestures, leads to several testable questions and further challenges for future research.

First, an interesting avenue for further research is to determine how changes in the external constraints—such as the cognitive demands of a task—and in the ease of availability of external resources, changes the likelihood of gesturing. For example, one could devise a mental rotation task in which participants can rotate a 3d-zoid either through a mouse, by using gestures, or solely by internal strategies. According to the present perspective, if we manipulate the speed in which the 3d-zoid can be manipulated by a mouse, we would predict that participants are more likely to use gestures when the manipulation takes more time (as relative cost decreases). Another, more unorthodox manipulation would be to put varying weights on the wrists of participants, which may induce costs in terms of energy expense, leading participants to an earlier adoption of an internal solution strategy. Many more constraints could be considered to assess the trade-off decision between internal and external resources that the cognitive system seems to make.

Second, gesture use evolves (Chu & Kita, 2008). When the task is more familiar, hand-gestures evolve from "*as-if* manipulations" to a stand-in-for relation of the 3d-zoid by means of a rotating flat hand, eventually eliminating the use of gestures altogether. In a similar vein, when abacus users become more advanced they tend to use less and less gestures during mental calculations. Indeed, it seems that gestures itself are costly to perform, and contrary to the GSA account, may under certain circumstances hinder performance (De Nooijer, Van Gog, Paas, Zwaan, in press), or learning (Post, Van Gog, Paas, & Zwaan, 2013) relative to other strategies. Interesting in this regard, is research that suggests that different types of body-movements have their own cognitive load (or come with particular cognitive costs) and may at times be traded for less costly bodily movements. That is dancers who rehearsed a dance-routine performed better when they rehearsed through "marking" (minimal movements and use of gestures to stand in for full-out movements) as opposed to rehearsing the routine full out (Warburton, Wilson, Lynch, & Cuyckendall, 2013). Thus, it seems that under certain conditions, gestures, once cheap resources to think with, become relatively costly in comparison to, and are therefore traded in for, purely internal strategies. This raises several questions. For example, do gestures help in the internalization process? Thus, are embedded/extended solution strategies shaping the way internal computations are performed?

Relatedly, when the cognitive system has a lower ability to produce internal object rotations (i.e., low spatial cognitive ability) it will rely more on external resources such as gestures (e.g., Chu, Meyer, Foulkes, & Kita, 2013; Marstaller & Burianová, 2013). An important research question that relates to this idea is whether people who score "low" on spatial cognitive ability test are actually only scoring low on *mental* spatial cognitive ability, and may not underperform when gestures are allowed. Indeed, when gesture is prohibited people who are low in working memory perform only more poorly on a mental rotation task with no performance deficits in the gesture condition, suggesting that they can fully compensate with external problem-solving strategies (Marstaller & Burianová, 2013). Furthermore, consider findings that prohibiting gesturing has a negative effect on performance. Seen in this light, this negative effect of not gesturing may not arise because it imposes cognitive load, and thereby imposes constraints on cognition (as proposed by the GSA account), but precisely because the prohibition to gesture withholds the cognitive system from the use of external resources in the performance of a task. Thus, whereas the GSA account suggests that not-gesturing imposes a cognitive load since the agent has to prevent automatic activations of gestures, we propose that the prohibition of gesturing takes external bodily resources away from the agent and drives the agent to rely exclusively on internal computational processes. This is an important empirical question that future research should address, as it is both related to how we should define and measure cognitive abilities, as well as to the particular cognitive function of gestures.

A more fundamental question that currently remains unanswered in the embedded/extended perspective on gesturing is what type of information is being made available through gesturing. Is it the proprioceptive, kinesthetic, haptic and/or visual consequences of movement that allow gestures to support cognitive processes? Or both, as these systems are tightly coupled (e.g., Radman, 2013)? For example, it is well-known that the visually impaired people use gestures (Iverson, 1998). Do they still benefit from gestures through proprioception or other consequences of movement? Clark (2013) raised a similar question in relation to patients with a rare disease that leads to loss of proprioception; yet these patients are still able to gesture quite naturally (see Gallagher, 2005).Would gestures still fulfill an embedded/extended cognitive function for such patients through visual feedback? This question is somewhat harder to address since the disease is, luckily, quite rare. An interesting avenue for research therefore would be to interfere with the information that gestures might provide as to identify factors that might underlie the embedded/extended cognitive function of gestures. For example, obstructing visibility of one's own gestures, by putting a screen at the level of the shoulders (Gallagher, 2005). Thus the current challenge for the present account is to provide an account of what information gestures produce that might be supportive for cognitive processes.

Conclusion

By means of our review of the empirical literature we have tried to assess explanatory power of current theories with regard to the question of how gestures might fulfill cognitive functions. Although all the accounts we have addressed here claim that gestures indeed fulfill a cognitive function, we have shown that in these accounts, this claim often does not refer to gestures, but rather to their neural precursors. Importantly, there are accounts that suggest that gestures fulfill the cognitive role of priming or activating internal action representations (e.g., Goldin-Meadow, & Beilock, 2010; Kraus, 1988), yet we think the reason why bodily movements fulfill this function is not clearly stated and seems to differ from the embedded/extended cognitive function we have identified here. We have tried to analyze the cognitive functions of gestures, by integrating the literature of embedded/extended cognition with the gesture literature. There is a considerable amount of overlap between the ways cognizers have been found to use their environment as well as how gestures support cognitive processes. Although further research into the exact mechanisms of embedded/extended functions of gestures is necessary, we put forth the notion that gestures provide the cognitive system with a stable external, physical, and visual presence that can provide a platform for thought.

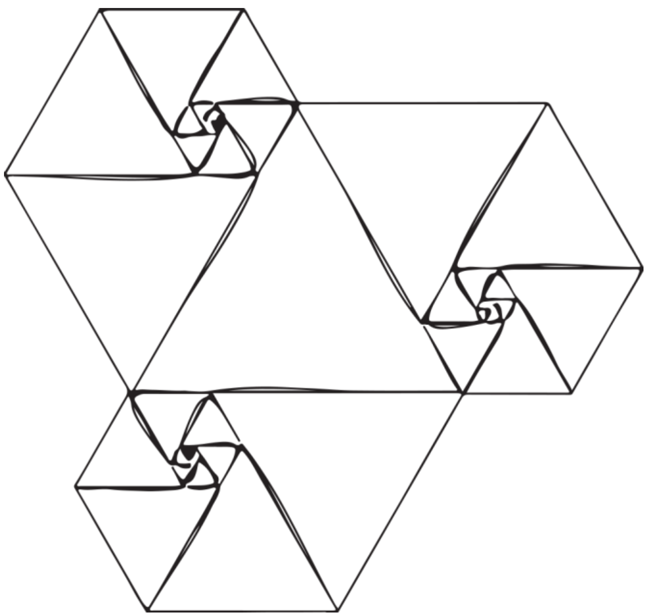
Importantly, we should stress two related concerns that apply to the current proposal. First, it is evident that the embedded/extended view on gestures, as presented here, does not address the full gamut of gesticulation. We have primarily focused on co-thought gestures in problem-solving contexts instead of for example beat gestures, or gestures that primarily emerge in communicative contexts. Therefore, at this point we remain agnostic to whether all gestures fulfill an embedded/extended cognitive function (for the gesturer). Indeed, extant "alternative" theories that we have addressed here may very well be complementary to our proposal. These theories are complementary to our proposal in that they might address cognitive functions and underpinnings of gestures that we have not addressed here. For example, it is possible that gestures emerge from action-related motor simulations that are activated during visuo-spatial cognition (Hostetter & Alibali, 2008) with the added proposal that the bodily externalizations of these motor simulations have a cognitive function themselves of the kind we have proposed here. Thus although we maintain that current theories in the gesture literature are not very suitable to address why gestures-as-bodily-acts might fulfill a cognitive function, our proposal does not deny any explanatory power of these theories regarding other aspects of the nature and cognitive function of gestures.

Secondly, it is clear that gestures have a developmental trajectory and primarily emerge in intersubjective contexts (e.g., Iverson & Thelen, 1999; Lizkowski, Brown, Callaghan, Takada, & de Vos, 2012; McNeill, 1992; Tomasello, 2008). As such, the current embedded/extended account of the cognitive function of gestures is still presented in an "ontogenetic vacuum" *and* is still rather individualistic. Although this is a concern that needs to be addressed in future work, there is much room for exploring how the embedded/extended function of gestures might be related to developmental and social dimensions. For example, Iverson and Thelen (1999) have provided a detailed account of how the hands, mouth and the brain should be regarded as one dynamical system; more specifically of how these components become entrained throughout development. Although they focus primarily on the way language and gesture become constitutively interdependent, the kind of gestures that have been the focus of this paper (gestures in problem-solving contexts) can be scaffolded onto their developmental account as another way of how "perception, action, and cognition can be mutually and flexibly coupled" (Iverson & Thelen, 1999, p. 37). On the other hand, how does our account relate to the intersubjective context in which gestures most often emerge? It would fare well with appeals coming from embodied cognitive science which suggest that an important way humans achieve interpersonal understanding is not from a spectatorial third-person stance, but rather from an interactive and second-person stance (e.g., Anderson, Richardson, Chemero, 2012; De Jaegher & Di Paolo 2007; De Jaegher, Di Paolo, Gallagher, 2010; Pouw & de Jong, submitted; Schilbach et al., 2013). In these approaches interpersonal understanding involves "know-how that allows us to sustain interactions, form relations, understand each other, and act together" (De Jaegher et al., 2010, p. 442), instead of two brains trying to predict each other's mental contents through observation alone. In such a portrayal of intersubjectivity, gestures are always already considered as having an embedded function for both the gesturer and the interlocutor since gestures are co-constitutive of the social coordination itself. To put it another way, in social interaction gestures are a non-neural component that is part of an organism-organism-environment coordinative structure (Anderson et al., 2012). The challenge for further work is to show how non-social embedded/extended gestures that we have focused on here might develop from these social contexts.

In closing, our aim with this article to point out the necessity of understanding the role of the body in thinking. We tried to accomplish this by developing an embedded/extended perspective on the cognitive role of gestures. In this perspective, the body is not a trivial output-appendage of the cognitive system but an important component thereof. The body is a resource with particular qualities that is recruited in the coordination of cognitive processes. This perspective intended to promote research that tries to further address when, why, and how gestures are recruited during cognitive processes.

Chapter 6

Problem solving is supported by co-thought gestures when task complexity is high and visual working memory capacity is low*



This chapter has been submitted for publication as Pouw, W. T. J. L, Eielts, C., Van Gog, T., Zwaan, R. A., & Paas, F. (2016). *Problem solving is supported by co-thought gestures when task complexity is high and visual working memory capacity is low*. Manuscript submitted for publication.

Problem Solving is Supported by Co-thought Gestures when Task Complexity is High and Visual Working Memory Capacity is Low

Research into co-thought gestures, which arise during problem solving in silence, without communicative intent, holds great promise for furthering our understanding of the nature of gesture. We investigated whether co-thought pointing gestures during mental problem solving support subsequent performance in a Tower of Hanoi and a chess task. Congenial to the idea that gestures offer new or improved embodied cognitive resources that may support internal cognitive resources, we found that gesturing benefited problem-solving performance under conditions of high cognitive load; when tasks were more complex, for participants with a lower visual working memory capacity. An interaction effect of gesticulation on problem solving with regards to spatial working memory capacity was not obtained. In conclusion, the current study provides evidence that co-thought pointing gestures, either performed under instruction or spontaneously produced, aid in problem solving (as compared to no gesture support) under conditions of high cognitive load.

Introduction

Hand-gestures, such as pointing to objects or acting on imagined objects, do not only arise in communicative contexts, but also in a wide variety of procedural learning or problem-solving tasks, such as counting coins (Kirsh, 1995), mental rotation tasks (Chu & Kita, 2008), mental abacus calculations (Brooks, 2014), tracking moving items in space (Delgado, Gómez, & Sarriá, 2009), solving fraction problems (Zurina & Williams, 2011), route-learning (e.g., Logan, Lowrie, & Diezmann, 2009; So, Ching, Lim, Cheng, & Ip, 2014), and rotating gear problems (e.g., Alibali, Spencer, Knox, & Kita, 2011; Stephen, Dixon, & Eisenhower, 2009). These non-communicative gestures produced without speech, are referred to as *co-thought gestures*. The fact that co-thought gestures spontaneously occur without communicative intent, strongly suggests that the cognitive function of gestures may go beyond support of (communicative) speech processes (Chu & Kita, 2016; Pouw, de Nooijer, van Gog, Zwaan, & Paas, 2014). Yet, despite their revealing nature of the possible function(s) of gesture, they have been largely neglected in the study of gesture.

Most research designed to investigate the cognitive function of gestures has focused on co-speech gestures; hand-gestures that are synchronized with and meaningfully related to speech (e.g., Cook, Yip, & Goldin-Meadow, 2012; Goldin-Meadow, Nusbaum, Kelly, & Wagner, 2001; Marstaller & Burianová, 2013; for review see Goldin-Meadow & Beilock, 2010). Such research has revealed that participants who gesture while explaining a solution of a math or physics problem perform better on a concurrent secondary task (e.g., remembering strings of letters or 2-d spatial coordinates) than when participants are inhibited to gesture. This research suggests that gesturing participants have more cognitive resources available than participants who are inhibited from gesturing.

If gestures do indeed provide resources to think with, then how do they do so and what do these resources consist of? With regards to co-speech gestures it is hypothesized that they support semantic processing (e.g., Cook, Yip, & Goldin-Meadow, 2012). For example, during verbal explanation of a solution of some problem, some of the information that needs to be communicated can be more efficiently conveyed by depicting it through gesture as opposed to conveying it in analog format through speech. As such, distributing communicative content over multiple modalities would reduce speech processing load for the speaker (Cook et al., 2013).

Yet co-thought gestures occur in the absence of speech or communicative intent, and may therefore shed light on an entirely different cognitive function of gesture. Research on co-thought gestures during mental rotation tasks (Chu & Kita, 2008, 2011, 2016) has shown that when confronted with the task of judging whether two objects presented under different orientations are the same or not, problem solvers produce gestures as-if actually rotating the object to match the target object. Not only are these gestures produced when problem solvers verbally explain how they solve the problem; they also arise when confronting the problem in silence and even occur when silent speech processes are inhibited by a secondary task (Chu & Kita, 2008, 2011, 2015).

Importantly, these co-thought gestures are not merely epiphenomenal, because they are consistently found to boost mental rotation performance compared to not gesturing (Chu & Kita, 2011). In similar vein, other studies have found that co-thought gestures such as pointing and tracing can improve performance on route learning or tracking moving items in space as compared to not gesturing during such tasks (e.g., Delgado et al., 2009; Hegarty, Mayer, Kriz, & Keehner, 2005; Logan et al., 2014; Macken & Ginns, 2014; So et al., 2014).

Thus, co-thought gestures too seem to provide the gesturer resources to think with, but these resources do not seem to relate to (communicative) speech processing, as co-thought gestures do not co-occur with speech or communicative intent. So what function do co-though gestures fulfill and how? The embedded/extended account of gesture puts forth a non-speech related cognitive function of gestures (Pouw et al., 2014). Namely, gestures produce stable visual and proprioceptive sensory consequences that can be used to think with. As such, gestures, as external bodily movements, embed and extend internal cognitive resources that allow the cognitive system to solve problems in new or improved ways. For example, producing action-like gestures during mental rotation produces kinematic regularities (normally co-occurring with actually rotating an object) that can be used by the gesturer to predict the consequences of such a rotation. The visual and proprioceptive information produced by the gesture is either not yet available, or more costly to simulate internally, and therefore aids in solving the mental rotation problem (see also Pouw & Hostetter, in press).

This is a central prediction of the embedded/extended account of gesture: gestures are more likely to be used, and more likely to positively affect the cognitive system when internal cognitive means alone (e.g., working memory processes; mental imagery processes) are insufficient or inefficient for the task at hand. This idea aligns with a number of findings on co-speech gestures (see Pouw et al., 2014 for a review). Namely, speakers who have lower visual or verbal working memory capacity are more likely to use gestures when they talk than speakers with higher capacity (Chu, Meyer, Foulkes, & Kita, 2013; Gillespie, James, Federmeier, & Watson, 2014). When speakers are confronted with complex visually distracting information when telling a story, they are more likely to gesture than with simple information (Smithson, & Nicoladis, 2014; see also Melinger & Kita, 2007). And gesturing during explanation of a math problem alleviates cognitive load, but only for those gesturers who have a relatively lower verbal working memory capacity (e.g., Marstaller & Burianová, 2013; see Galati, Weisberg, Newcome, & Avraamides, 2015 for similar findings in a spatial domain). In sum, when cognitive load is high, either because of the complexity of the task or because of limited availability of internal cognitive resources, gestures are more likely to be adopted and more likely to support cognitive processing and problem solving.

In contrast to co-speech gestures (e.g., Hostetter & Alibali, 2008) it is still unknown under which conditions co-thought gestures arise (Chu & Kita, 2016). Although there is some evidence that co-thought gestures are more likely to arise when task complexity is high (e.g., Chu & Kita, 2008; Logan et al., 2009) it is not clear how they relate to the availability of internal cognitive resources such as working memory capacity. Therefore, we conducted two experiments to investigate when co-thought gestures are used and how they affect performance, under varying conditions of cognitive load.

The Present study

We investigated when co-thought gestures were produced during mental problem solving and whether producing such gestures would affect problem-solving performance. Specifically, we compared effects of instructed and spontaneous gesturing to 'spontaneous' non-gesturing[11] on performance on problems of varying complexity. According to the embedded/extended account of gesture, the spontaneous use of co-thought gestures, as well as an effect of such gestures on performance, is most likely when cognitive load is high; that is, with tasks of higher complexity and for individuals with relatively lower visual and spatial working memory capacity. These particular working memory capacity measures have been shown to be predictive of co-speech gesture frequency (Chu et al., 2013).

[11] i.e., no explicit inhibition of gesture is taking place, to ensure that a positive effect of gesturing on performance would actually be due to the production of gesture, rather than constitute a negative effect of having to inhibit automatic gesture production; see e.g., Chu & Kita, 2011, Exp. 2 & 3; Cook, Yip, & Goldin-Meadow, 2010; Goldin-Meadow et al., 2001; Wagner, Nusbaum, & Goldin-Meadow, 2004).

We tested this hypothesis using two different problem-solving tasks: the Tower of Hanoi (TOH; often used in gesture research, e.g., Trofatter, Kontra, Beilock, & Goldin-Meadow, 2015)[12], and chess (not yet studied in gesture research), thereby providing an opportunity for conceptual replication within the study. Both TOH and chess problem solving have been found to positively and substantially correlate with visuospatial working-memory capacity (Robbins et al., 1996; Zook, Davalos, Delosh, & Davis, 2004), which is important for testing our assumptions regarding cognitive load in these tasks. The effects of gesturing on both tasks were investigated consecutively, with the same participant sample and a similar design.

We hypothesized that (1) participants should make more co-thought gestures under high cognitive load conditions as determined by task complexity (i.e., higher complexity = higher load), visual and spatial mental working-memory (lower capacity = higher load), and the combination of task complexity and working memory (gesture likelihood hypothesis); and (2) co-thought gesturing (spontaneous and instructed) should positively affect mental problem-solving performance (as evidenced by speed and efficiency of subsequent actual performance) on both problem-solving tasks (TOH and chess) in comparison to non-gesturing, under high cognitive load conditions as determined by task complexity (i.e., higher complexity = higher load), visual and spatial mental imagery ability (lower capacity = higher load), and the combination of both (gesture effect hypothesis).

General Method

Participants and Design

A total of 76 Dutch university students participated in this study in partial fulfillment of course requirements or for a financial compensation of 7.50 euros (about 8.30 USD). One participant had to be excluded from the entire study due to a technical malfunction during the working memory capacity tasks. For the TOH, two additional participants had to be excluded from the sample as they did

[12] It is important to note that although there are findings that gestures affect (or are affected by) solving the Tower of Hanoi (Beilock & Goldin-Meadow, 2010; Cook & Tanenhaus, 2009; Garber & Goldin-Meadow, 2002; Goldin-Meadow & Beilock, 2010; Trofatter et al., 2015) these studies are not diagnostic for present purposes as they have focused on iconic co-speech (as opposed to co-thought) gestures. Additionally, when performance was object of study, it was assessed between physically different TOH tasks, making the task performance dependent on (secondary) manipulative features of the problem-solving task (the weights of the discs), not on the primary problem-solving content (Goldin-Meadow & Beilock, 2010; Trofatter et al., 2015).

not comply with the gesture instruction. This resulted in a total sample of 73 for the TOH (41.1% men, M_{age} = 20.60, *SD* = 2.06, age range 18 to 31 years). Because of delays in programming the chess task, 16 participants had already participated in the TOH study before the chess task experiment was added to the procedure. Moreover, in addition to the one participant mentioned above, four other participants had to be excluded due to malfunctioning of the video or the task program. As a consequence, the total sample for the chess task consisted of 55 participants (31.9% men, M_{age} = 20.47, *SD* = 2.19, age range 18 to 31 years).

Participants were assigned randomly to one of the two conditions (instructed gesture vs. gesture allowed), under which they performed both TOH and chess tasks. Nine participants reported (when asked) that they had played the Tower of Hanoi in the past, but they were equally divided across conditions, $\chi^2(1) = .123$, $p = .725$. There were no differences between conditions for self-reported knowledge of chess, $t(53) = -.980$, $p = .332$.

Materials

Visual working memory capacity: Visual Patterns Test

The Visual Patterns Test (VPT; Della Sala, Gray, Baddeley, & Wilson, 1997) is a measure for visual working memory capacity. An adaptation of the VPT was used, which was developed (and kindly provided to us) by Chu and colleagues (2013). Participants were shown a matrix, in various patterns, wherein half of the cells (i.e., squares of 15 mm × 15 mm) were colored black. Each pattern was displayed for 3 seconds, after which all the squares turned white and each square was labeled with one unique letter. Participants indicated the pattern of black squares by naming the letters of the corresponding squares out loud in a non-specific order. The VPT consisted of 25 trials, with blocks of 5 trials per difficulty (from seven to 11 black squares). Before the start of the task participants were provided with two practice trials (3 and 4 black squares, respectively). If participants failed to recall all the letters, it was scored as an incorrect response. After five consecutive incorrect responses within one block of trials the experimenter stopped the task.

Spatial working memory capacity: Corsi Block Task

Spatial-sequential working memory capacity was measured with the Corsi Block Task (from here on CBT; Corsi, 1972), as used by Chu and colleagues (2013). In each trial (20 trials), nine empty irregularly placed squares (15 mm × 15 mm) were displayed on the computer screen. One square at a time turned black for one second, with an inter stimulus interval of 500 ms between each transition. The first block of five trials consisted of a sequence of five squares turning black, with each subsequent difficulty level adding one black square to the sequence (the fourth block sequence thus consisted of eight squares turning black). After the last square in the sequence turns black, a letter appeared in the center of each square. The participant reported aloud the letters in the correct order, following the order of the squares turning black. If the participants failed to recall all the letters in the correct order, the trial was scored as an incorrect response. After five incorrect responses in one block of trials the experimenter stopped the task.

Tower of Hanoi

The Tower of Hanoi (TOH) consisted of a wooden structure with a rectangular base (17 cm tall, 50 cm wide, 1cm deep) with three evenly spaced pegs (13.5 cm tall, 1 cm in diameter) mounted on top. The 'easy' version of the task consisted of three wooden discs (size disc 1: 7.5 cm in diameter; size disc 2: 7 cm in diameter; size disc 3: 6.5 cm in diameter) and the more complex version of the task of four discs (size disc 4: 6 cm); all discs were 1 cm in height. They were initially stacked on the left most peg, and could be placed on the other pegs during the problem-solving process. When participants took longer than 300 s to solve either of the two trials, the trial would be aborted.

Chess Problem-solving Tasks

Chess rules tutorial. Participants were provided a short animated tutorial on the computer about a set of five chess rules that helped prepare participants to solve the problem-solving task. The tutorial was designed specifically for participants with little if any knowledge of chess. The tutorial was self-paced and participants proceeded to the next stage of the tutorial by clicking a "Next" - button. The total duration of the tutorial is approximately 10 minutes. The instructions covered rules regarding the alternate turn taking in chess, the allowed movement of the rook, bishop and knight pieces, and the ability to capture a chess piece of the other color. Each of these rules was presented separately.

Chess rules knowledge test. Participants performed a chess knowledge task where they had to indicate all the squares to which a chess piece is allowed to move by selecting them on the chessboard using mouse clicks. This was performed for the rook, the knight, and the bishop separately. The purpose of this task was to ensure that participants were sufficiently familiar with the chess rules required to perform the problem-solving task. Once participants were able to exclusively select all the correct squares they automatically proceeded to the square selection task for the next chess piece. If participants repeatedly failed to select the correct squares, the experimenter verbally repeated the movement rule of the chess piece, until the participant selected the correct squares.

Chess problem-solving task. For each of the three tasks, participants were presented with two chessboards shown on the computer screen (see Figure 1). The chessboard on the left represented the start state and the chess board on the right indicated the end state. Three (low complexity), four (medium complexity) and five (high complexity) solving steps were required to transform the start state into the end state. For each task, there was only one correct solution. The time for mental preparation varied across the different levels of complexity (30 s for the 3-step problem, 60 s for the 4-step problem, and 150 s for the 5-step problem). Ending the mental preparation phase, participants were cued by an auditory tone to start solving the task. The chess pieces of the begin state could be moved across the chessboard by clicking on them with the left mouse button and dragging them to another square on the board. The chess pieces were placed on the selected square by releasing the left mouse button. Every placement of the chess piece was permitted, even when it was an invalid chess move. The board state would be automatically reset to the begin state after the maximum number of moves (3, 4 or 5) was exceeded and the solution was incorrect. Once participants correctly solved the problem, they automatically proceeded to the next task. When participants were unable to solve the task within 300 s, the task was aborted and they automatically proceeded to the next task.

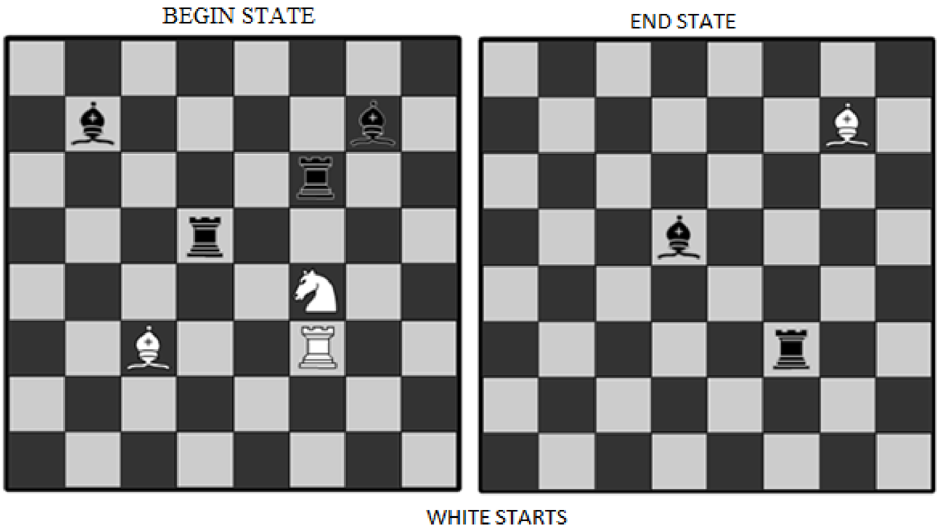


Figure 1. Board composition in the medium difficulty task, with the start state on the left and the end state on the right.

Self-report Questions

Mental effort, difficulty, interest. We obtained an indication of experienced mental effort, perceived difficulty, and experienced interest, after each problem-solving trial because experienced effort and difficulty should be affected by task difficulty as well as working memory resources (Paas & van Merriënboer, 1993), whereas interest is assessed to check that this would not differ (and therefore might explain hypothesized differences) among groups. These indications were obtained via self-reports after each problem-solving trial on a 5-point rating scale (see Paas, Tuovinen, Tabbers, & Van Gerven, 2003): "How much mental effort did you exert during the task? (mental effort; 1 = 'very low mental effort' to 5 = 'very high mental effort'), "How difficult did you find this task" (difficulty; 1 = 'not difficult' to 5 = 'very difficult'), and "How interesting did you find this task" (interest; 1 = 'not interesting' to 5 = 'very interesting').

Prior knowledge and experience. Before the start of the experiment participants rated their knowledge and experience with chess on a 5-point scale with the anchors 1 (very low) to 5 (very high). Participants also reported previous experience with the TOH (yes or no).

Procedure

Before the start of the experiment, participants consented to being video-recorded during the experiment. Participants were tested individually with the experimenter present in the room (but they could not see the experimenter while they were working on the tasks). The order of the working memory capacity tasks was counterbalanced such that in both conditions half of the participants started with the spatial working memory task (CBT), whereas the other half would start with the visual working memory task (VPT). During the working memory tasks, participants had to respond verbally to each trial after which the experimenter would start the next trial. Responses were immediately scored on a scoring-sheet by the experimenter, and were also recorded on video.

After participants completed both working memory capacity tasks, the experimenter started the instructions for the TOH. Firstly, they were told about the end goal: getting the arrangement of discs on the outer left peg on the outer right peg, in the same order. Then they were told about the two constraints: I) only one disc at a time may be moved from one peg to another and II) a larger disc cannot be placed on a peg that already contains a smaller disc. After each instruction the experimenter verified whether subjects understood the instructions (they were asked to verbally repeat the rules). Participants were informed that before solving the TOH as fast as possible, they would mentally plan the moves without manipulating the TOH for 150 s (the TOH was placed just outside arms reach: 90 cm from the participant). Participants were told that they should find and rehearse the correct moves repeatedly during this phase. Half of the participants were additionally instructed "to gesture, in other words, to think with your hands during this mental planning phase in a way that suits you" (gesture instructed condition). During this instruction the experimenter also performed a quick demonstration of a typical index pointing-gesture, which consisted of pointing movements directed at the TOH pegs. Thus participants were cued to use pointing gestures in the instructed gesture condition, but could gesture in a way that suited them. The other half of the participants did not receive any gesture instructions (gesture allowed condition). Participants first solved the easier 3-disc TOH, after which they reported mental effort invested in the task, perceived task difficulty and their interest in the task. Subsequently, the same routine was repeated (i.e., mental problem-solving phase followed by physical solving phase) for the more complex 4-disc TOH.

After completion of the TOH tasks, instructions of the chess tutorial were started. Participants performed these exercises at self-paced tempo and were allowed to ask questions about the content of the tutorial. Upon completing the chess knowledge task, the experimenter would start the instructions for the chess problem-solving task. The task was preceded by an example of the solution to a simple 2-step example of the task. After the experimenter verified that the subjects understood the instructions and had no more questions, the planning phase for the first problem-solving task was started. Note that participants continued under the same condition as they had been assigned in the TOH phase. The instructions for the gesture instructed condition were repeated. After each planning phase participants were prompted with a short tone to notify them that they could solve the chess trial using the mouse to click-and-drag pieces. Directly after each of the three chess trials, participants were prompted to respond to the self-report questions on experienced mental effort, difficulty of the task and interestingness of the task. After the experiment, participants filled out a questionnaire including demographic questions (age, sex, study program, prior experience with the task) and questions about what they thought was the purpose of the experiment. Finally, participants were debriefed and thanked for their participation.

Scoring and data analysis

Visual Patterns Test

The final score was the proportion of the correct responses out of all 25 trials (higher proportion approximates higher visual working memory capacity).

Corsi Block Task

The final score was the proportion of the correct responses out of all 20 trials (higher proportion approximates spatial visual working memory capacity).

Tower of Hanoi

For each of the two problem-solving trials we obtained solving speed and number of solving steps, with lower solving speeds and lower number of solving steps reflecting a higher performance. For the 3-disc TOH and the 4-disc TOH the minimal amount steps necessary to solve the task were seven and fourteen steps, respectively. A step was counted when participants placed a disc on another peg (so not on the same peg) once the disc was let go (i.e., if they changed their mind before releasing the disc, this was not counted as a step). Performance of participants who did not solve the task in 300 s was not scored.

Chess task

For each of the three chess trials we obtained solving speeds and number of attempts (i.e., lower solving speed and number of solving attempts means higher performance). Solving-steps were automatically counted when participants clicked-and-dragged a piece and let it go, and for each attempt participants had to perform 3, 4, 5 solving steps for chess task 1, 2, 3, respectively. When participants did not solve the task in 300s performance was not scored.

Gesture

Each participant's video data were coded for gesture prevalence for each TOH and chess trial. Three types of gesture prevalence were possible in each trial, namely non-gesturing vs. spontaneous gesturing vs. instructed gesturing; which we will collectively refer to as gesture prevalence type. Note, that the participants in the gesture allowed group, are thus re-divided into two categories of gesture prevalence type (non-gesturing vs. spontaneous gesturing), and since the participants in the instructed gesturing group all gestured, they were all assigned to the instructed gesturing prevalence type.

Additionally, we looked at form (pointing vs. iconic), and frequency of gesture. Almost all instructed and spontaneous gesturing consisted of pointing gestures (see Figure 2 for examples) with three exceptions. Two participants briefly used counting gestures and one participant briefly used iconic gestures with two hands picking and moving all the discs. Therefore, as a measure of gesture frequency we only counted the number of pointing gestures; in such a way that each rest point after a pointing gesture (either whole hand -or finger-pointing) was considered as one instance. For the TOH task, a research assistant counted all the gestures, and the first author recounted 10.4 % to check for reliability (interrater reliability was high, $r = .96$, $p < .001$). For the chess task, the second author counted all the gestures, and the first author recounted 10% to check for reliability (interrater reliability was high, $r = .90$, $p < .001$).

Note, for the chess task, although we could ascertain whether participants gestured we were unable to count the exact gesture frequencies for a number of participants for task 1 ($n = 8$), 2 ($n = 8$), and 3 ($n = 6$). This was because of an erroneous positioning of the video camera which was problematic since participants sometimes gestured more closely to the screen (the computer-monitor obstructed the view on the participants' hands). Since time differed across task complexity for the chess task we will also report a gesture ratio measure that provided number of gestures per 10 seconds; gesture frequency divided by mental preparation time x 10 (task 1, 2, 3 = 3 x 10s, 6 x 10, 9 x 10 s). Chu et al. (2014) used a similar procedure to control for time in gesture frequencies.

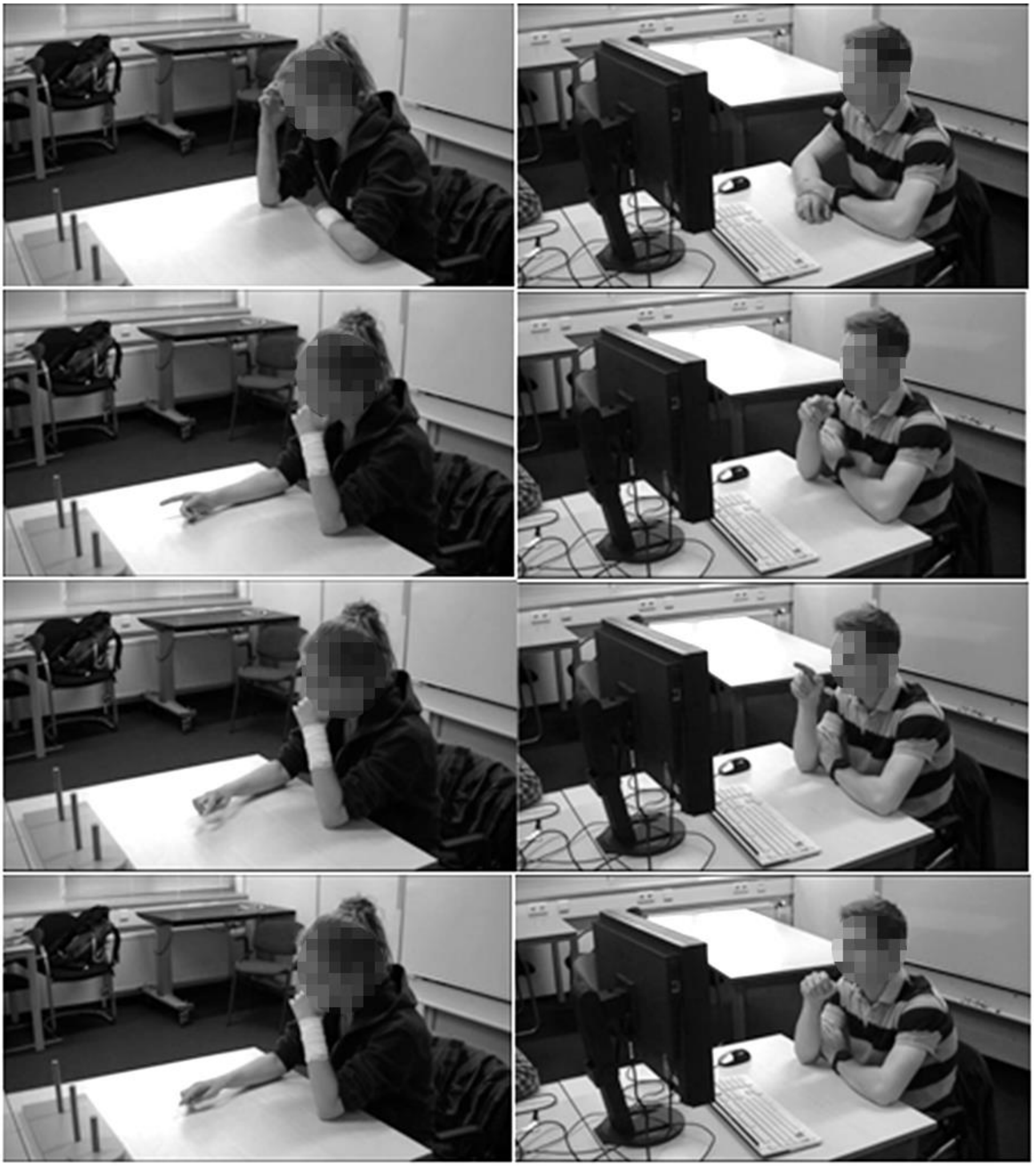


Figure 2. Examples of spontaneous gestures arising during mental problem solving for the TOH and chess task (1 frame per second).

Results Tower of Hanoi

Before discussing the analyses that address our hypotheses, we first address overall performance results on the 3-disc and 4-disc TOH (solving speed and solving steps), and visual working memory capacity (VPT), spatial working memory capacity (CBT), and gesture production data.

Overall Performance TOH

Three-disc TOH. Overall mean solving speed and solving steps 3-disc TOH and the correlation between solving speed and solving can be obtained from Table 1. In terms of solving steps most participants had a ceiling performance (82.2 % performed the task in 7 steps). Since both solving speed (skewness = 4.87, *SE* = 0.28, kurtosis = 27.23, *SE* = 0.56) and solving steps (skewness = 3.12, *SE* = 0.28, kurtosis = 9.63, *SE* = 0.56) were heavily skewed, we logarithmically transformed these variables. This did not completely resolve issues of skewness or kurtosis for solving speed (skewness = 1.99, *SE* = 0.28, kurtosis = 6.40, *SE* = 0.56), nor solving steps (skewness = 2.73, *SE* = 0.28, kurtosis = 7.09, *SE* = .56).

Four-disc TOH. Overall mean solving speed and solving steps for the 4-disc TOH, and the correlation between solving speed and solving can be obtained from Table 1. Note that 37.7 % had a ceiling performance of 15 solving steps. Three participants were unable to solve the task within 300 s. One participant forgot to follow the rules and made several mistakes during the 4-disc TOH (and was therefore excluded from analysis of the 4-disc TOH). Solving speed (skewness = 0.90, *SE* = 0.29, kurtosis = -0.361, *SE* = 0.57) showed no issues with normal distribution. However, solving steps was not normally distributed (skewness = 1.54, *SE* = 0.29, kurtosis = 2.05, *SE* = 0.57) and we therefore used a log transformation (skewness = 0.54, *SE* = 0.28, kurtosis = -0.04, *SE* = 0.57).

Visual and Spatial Working Memory Capacity

Next we checked for (unexpected and unwanted) differences in working memory measures between conditions. Table 1 displays the averages, standard deviations and correlations for VPT and CBT, and solving speed of both 3- (log-transformed) and the 4-disc TOH. There were no differences between the gesture allowed and instructed condition in the proportion correct trials on the VPT, $t(71) = -0.494$, $p = .623$ or CBT, $t(70) = 0.328$, $p = .744$. The VPT (but not CBT) showed a significant negative correlation with solving speed on the 3-disc, and both the VPT and CBT negatively correlated negatively with solving speed on the 4-disc TOH, such that higher VPT and CBT scores were associated with faster solving speeds and a lower amount of solving steps.

	Mean	SD	1	2	3
1. VPT score	0.50	0.15			
2. CBT score	0.41	0.15	.36**		
3. Solving speed 3-disc TOH	20.52	16.00	-.19*	-.16	
4. Solving speed 4-disc TOH	84.61	57.83	-.57**	-.41**	0.21

*Note. * p < .05, ** p < .01. Averages, standard deviations and correlations for the VPT, and the solving speed for both TOH tasks*

Table 1. Means, standard deviations, and correlations of VPT, CBT, and solving speed on the Tower of Hanoi 3-disc and 4-disc.

Gesture production

As indicated by Figures 3 and 4, on the 3-disc TOH, 28.90% (*N* = 11/38) of the participants in the gesture allowed condition spontaneously gestured during the mental solving phase (which had a fixed time of 150 s) with a mean absolute gesture frequency of 23.45 (*SD* = 13.87; frequency minimum = 8, maximum = 59). In the gesture instructed condition the mean absolute gesture frequency was 46.03 (*SD* = 24.20; frequency minimum = 7, maximum = 105).

On the 4-disc TOH, 44.44% (*N* = 16/38) of the participants in the gesture allowed condition spontaneously gestured during the mental solving phase, with a mean absolute gesture frequency of 36.25 (*SD* = 25.40; frequency minimum = 5, maximum = 79). In the gesture instructed condition the mean absolute gesture frequency was 65.73 (*SD* = 18.52; minimum = 19, maximum = 104).

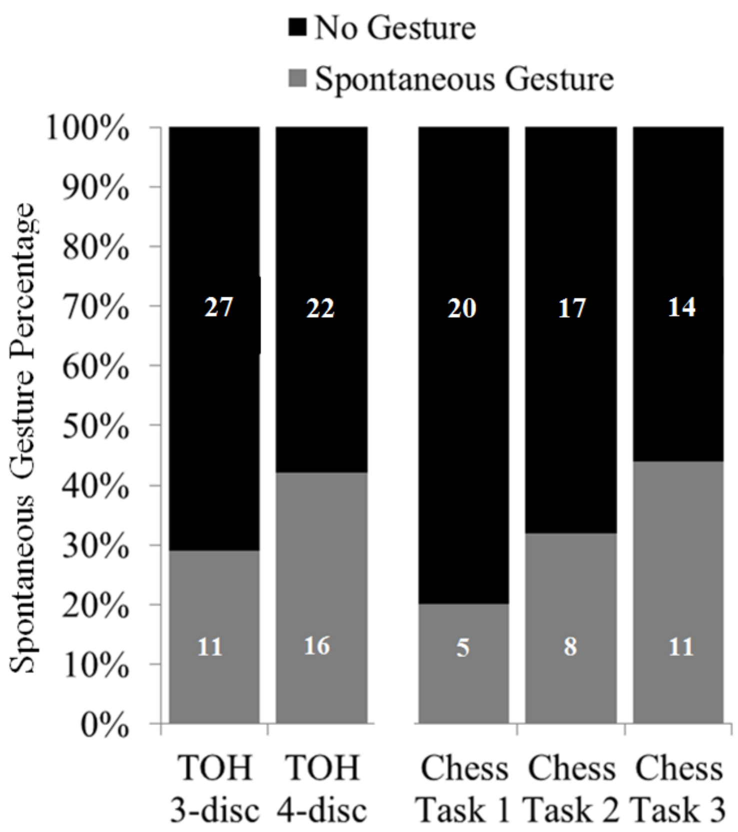


Figure 3. Gesture prevalence (no gesture vs. spontaneous gesture) in the gesture allowed group for both the TOH and Chess trials

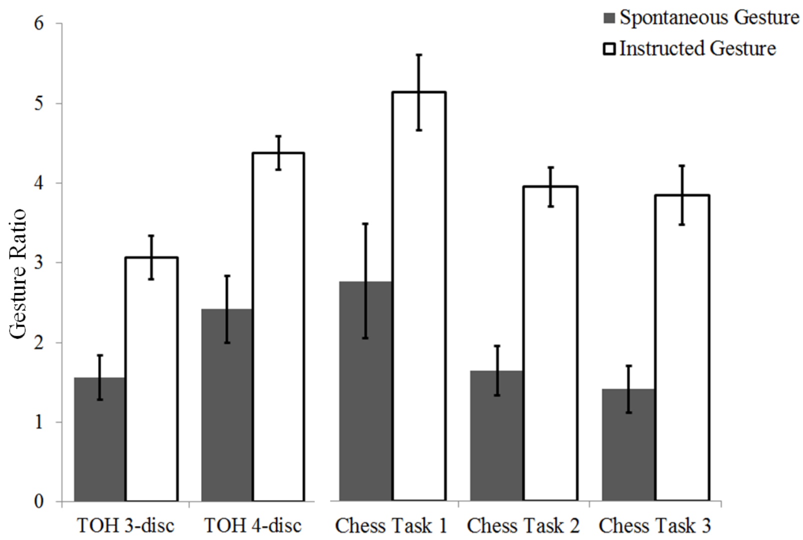


Figure 4. Gesture ratio indicating gesture Frequency per 10 seconds of mental preparation

Gesture Likelihood Hypothesis

The gesture likelihood hypothesis predicted that the percentage of spontaneous gesture prevalence (non-gesturing vs. spontaneous gesturing) and frequency (number of gestures) would be higher in the complex task and for those with a lower visual and/or spatial working memory capacity. The subsequent analyses test that hypothesis and also provide a manipulation check on whether participants who were instructed to gesture did indeed do so more often in terms of gesture frequency than those who spontaneously gestured (also see Figure 4).

Spontaneous gesture prevalence likelihood

To assess whether there was a significant difference in the percentage of prevalence of spontaneous gesturing (as opposed to non-gesturing) across complexity we performed a dependent Chi square test (McNemar Change Test), which revealed that spontaneous gesture prevalence was indeed more likely in the 4-disc TOH than in the 3-disc TOH, $\chi^2(1) = 4.000$, $p = .039$.

However, participants who gestured spontaneously during the mental solving-phase of the solution procedure for 3-disc TOH, did not differ from non-gesturing participants with regard to VPT score, $t(36) = -0.351$, $p = .728$. Similarly, participants who gestured spontaneously during mental solving phase of 4-disc TOH, did not differ from non-gesturing participants with regard to VPT score, $t(36) = -0.950$, $p = .348$.

Similarly, spontaneously gesturing participants during the 3-disc TOH, did not differ from non-gesturing participants with regard to CBT score, $t(35) = -.094$, $p = .962$. This was also the case for the 4-disc TOH, spontaneously gesturing participants did not differ from non-gesturing participants with regard to CBT score, $t(35) = .083$, $p = .926$.

Thus in contrast to our predictions, visual (VPT) and spatial (CBT) working memory capacity did not affect whether participants gestured or not (i.e., gesture prevalence).

Spontaneous and instructed gesture frequency likelihood

A repeated measures ANOVA was conducted to assess whether task complexity (within-subjects: 3-disc TOH vs. 4-disc TOH) and gesture prevalence type (between-subjects: spontaneous vs. instructed gesture) predicted gesture frequency. This allows us to test whether gesture frequencies were indeed higher in the instructed gesture condition (manipulation check) and whether task complexity affected gesture frequency (gesture likelihood hypothesis). We found a significant main effect of task complexity, $F(1, 52) = 26.62$, $p < .001$, $\eta_p^2 = .339$, with gesture frequency being lower during the mental solving phase for the 3-disc TOH (*M* = 35.68, *SD* = 27.945) than for the 4-disc TOH (*M* = 53.83, *SD* = 27.26). Additionally there was a main effect of gesture prevalence type, $F(1, 52) = 26.62$, $p < .001$, $\eta_p^2 = .453$, showing that participants who were instructed to gesture did so more often (*M* = 56.94, *SD* = 22.18) than participants who spontaneously gestured (*M* = 22.32, *SD* = 21.02). There was no interaction between complexity and gesture group, $F(1,52) = .02$, $p = .882$.

We additionally assessed whether visual working memory (VPT) covaried with gesture frequency by running a repeated measures ANCOVA, however this did not account for variance in frequency of gestures, $F(1, 51) = 0.34$, $p = .562$, and the effect of complexity on gesture frequency dissipated, $F(1, 51) = 0.22$, $p = .641$. Also no interactions were obtained for VPT with complexity, $F(1, 51) = 3.27$, $p = .077$, or gesture group $F(1, 51) = 0.20$, $p = .887$. When entering CBT as a covariate this similarly yielded non-significant results, $F(1, 50) = 0.32$, p = .574. Also no interactions were obtained for VPT across complexity or gesture type, $F(1, 50) < .009$, $p > .924$. These findings suggest that visual and spatial working memory capacities did not predict the frequency at which participants gestured (either when instructed to do so, or when spontaneously produced).

6

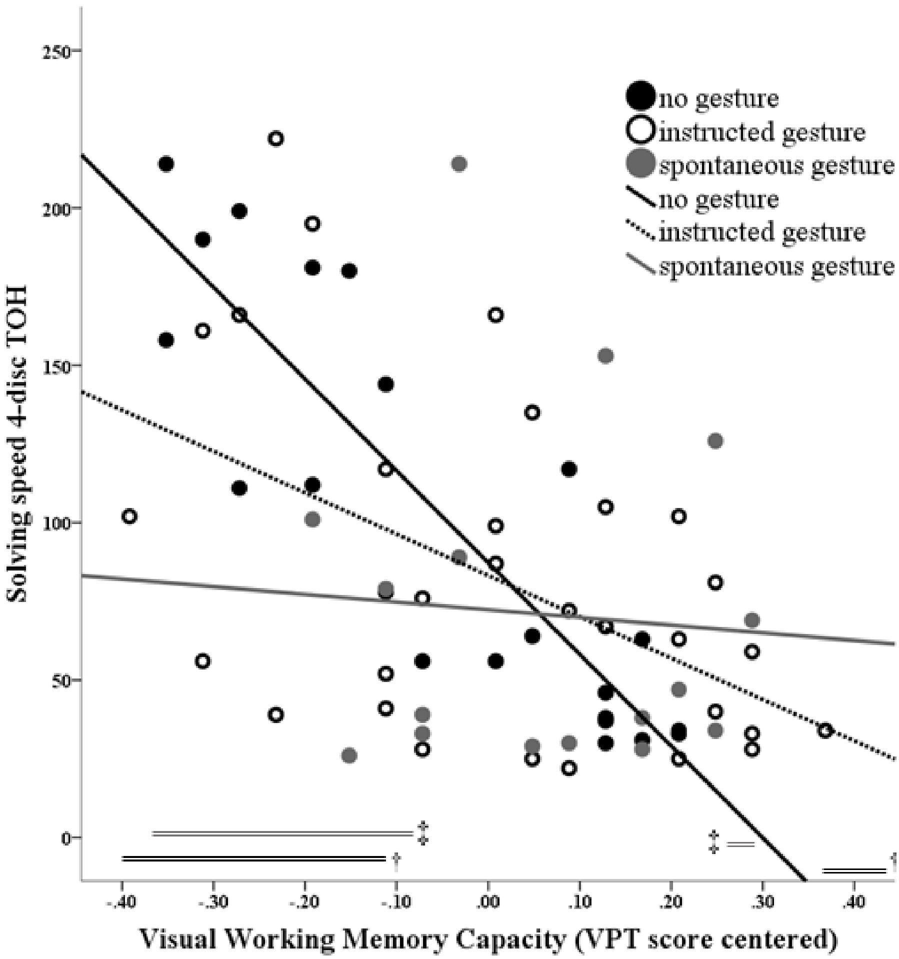
Gesture Effect Hypothesis

The gesture effect hypothesis predicted that co-thought gesturing (spontaneous and instructed) would positively affect mental problem-solving performance (as evidenced by solving speed of subsequent actual performance) under high cognitive load conditions in comparison to participants who did not gesture but were allowed to (i.e., 'spontaneous' non-gesturing). Because number of steps, which could serve as a measure of performance, was non-normally distributed (even after log-transformation), we will only analyze solving speed. We first analyzed whether VPT, gesture prevalence type (non-gesturing vs. spontaneous gesturing vs. instructed gesturing) and their interaction affected solving speed on the 3-disc and 4-disc TOH in two multiple stepwise hierarchical regression analyses (i.e., separate analysis for each task complexity level). Note that the gesture effect hypothesis predicts interaction effects - especially on the more complex task (i.e., solving speed 4-disc TOH) - of VPT and gesture prevalence type, such that those with lower VPT scores improve in performance when using gestures (i.e., instructed or spontaneous) as compared to those who did not gesture and that have similar VPT scores.

We entered VPT (mean centered) in the first step, and dummy variables spontaneous gesturing (0 = non-gesturing, 1 = spontaneous gesturing) and instructed gesture (0 = non-gesturing, 1 = instructed gesturing) in the second step. In the third step we entered two interaction terms of the centered VPT with instructed and spontaneous gesture prevalence. In Table 2 the results of the multiple stepwise hierarchical regression on solving speed of the 3-disc and 4-disc TOH are reported.

As can be seen in Table 2, no significant predictors were obtained for solving speed on the 3-disc TOH. However, on the 4-disc TOH, there was a main effect of VPT on solving speed, which remained so in the final model, $t(68) = -5.44$, $p < .001$, *Partial r* = -.570. We found no main effects of gesture prevalence type for spontaneous gesture, $t(68) = -.92$, $p = 0.361$, or instructed gesture, $t(68) = -.31$, $p = .758$. However in the final model (explaining 37% of the variance), the interaction terms of VPT and spontaneous gesture prevalence, $t(68) = 2.88$, $p = .005$, *Partial r* = .341, as well as instructed gesture prevalence, $t(68) = 2.42$, $p = .018$, *Partial r* = .292, were significant in the predicted direction. Namely, spontaneous and instructed gestures were positively affecting performance (i.e., lower solving speeds) on the task as compared to non-gesturing, especially when participants had a lower VPT score (see Figure 5).

Figure 5. Regression slopes for gesture type prevalence (spontaneous vs. instructed vs. no gesture) on solving speed on the TOH 4-disc, with centered VPT score on the horizontal axis.



Note. The "†" denotes the region where the effect of spontaneous gesture prevalence (vs. no gesture) was significant. The "‡" denotes the region of significance for instructed gesture prevalence.

6

3-disc TOH	B (SE)	β	Model	R^2_{adj}
Intercept	2.888(0.050)		1	.036
VPT	-.451 (0.236)*	-.221		
Intercept	2.910 (0.083)		2	.012
VPT	-0.444 (0.239)*	-.218		
Spontaneous gesture prevalence	-0.014 (0.153)	.012		
Instructed gesture prevalence	-0.03 (0.110)	-.058		
Intercept	2.896 (0.083)		3	.034
VPT	-0.905 (0.083)*	-.444		
Spontaneous gesture prevalence	0.025 (0.152)	.021		
Instructed gesture prevalence	-0.035 (0.109)	-.041		
VPT × Spontaneous gesture prevalence	-0.037 (0.782)	-.013		
VPT × Instructed gesture prevalence	0.837 (0.520)	.313		
4-disc TOH	B (SE)	β	Model	R^2_{adj}
Intercept	86.191 (5.769)		1**	.315
VPT	-168.68 (29.695)	-.570		
Intercept	94.254 (10.925)		2**	.303
VPT	-162.614 (30.709)**	-.550		
Spontaneous gesture prevalence	-13.850 (16.560)	-.102		
Instructed gesture prevalence	-10.265 (13.893)	-.089		
Intercept	87.321 (10.607)		3**	.377
VPT	-291.533 (53.623)**	-.985		
Spontaneous gesture prevalence	-14.962 (16.192)	-.110		
Instructed gesture prevalence	-4.080 (13.286)	-.035		
VPT × Spontaneous gesture prevalence	267.111 (92.805)*	.356		
VPT × Instructed gesture prevalence	160.130 (66.178)*	.397		

Note. * $p < .05$, ** $p < .01$

Table 2. Summary of stepwise hierarchical regression analysis for variables predicting log transformed solving speed on easy 3-disc TOH and solving speed on 4-disc TOH.

To further assess these interactions we performed Johnson-Neyman Regions of Significance Tests using PROCESS (Hayes, 2013) to assess the nature of the two interaction effects of instructed and spontaneous gestures with the VPT (see also Figure 5). We performed the first analysis by contrasting 'spontaneous' non-gesturing vs. instructed gesturing (IV; 0 = non-gesturing, 1 = instructed gesture prevalence), centered VPT (Moderator), and solving speed on the 4-disc TOH (DV). These analyses revealed that the region of significance ($p < .05$) was indeed obtained for a conditional positive effect of instructed gesturing on performance on 4-disc TOH (range b = -.66.815 to -29.183, range Se = 25.404 to 14.522, range $t(52)$ = -2.63 to -2.01, range p = .011 to .050) for those with a lower working memory capacity, range of significance for centered score VPT = -.392 to -.130, 26.41 % of the lowest VPT scorers. However, we further found that those with a higher working memory capacity (range of significance for centered score VPT = .361 to .448, 1.89% of the highest VPT scorers) seemed to be slowed down by instructed gesturing (range *b* = 51.269 to 65.238, range *Se* = 25.549 to 65.238, range $t(49)$ = 2.01 to -2.16, range p = .05 to .035).

Similarly, in the second analysis, this time only contrasting spontaneous gesture prevalence with 'spontaneous' non-gesturing (coded 0 = non-gesturing, 1 = spontaneous gesturing), we found a conditional positive effect of spontaneous gesturing, range *b* = -108.927 to -37.302, range *Se* = 37.198 to .18.313, range t (32) = -2.93 to -2.04, range p = .05 to .006) for those with a lower working memory capacity, region of significance for centered score VPT = -.351 to -.083, 33.33% lowest VPT scorers. Additionally, we obtained a (small) region of significance for participants with a higher working memory capacity (centered score VPT = .260 to .288, 2.78 % of the highest VPT scorers) that seemed to be slowed down by spontaneous gesturing (range *b* = 54.416 to 62.025, range *Se* = 26.714 to 28.8416, range $t(49)$ = 2.01 to -2.16, range p = .05 to .039).

3-disc TOH	B (SE)	β	Model	$R^2_{adjusted}$
Intercept	2.891 (0.050)		1	.033
CBT	-0.621 (0.334)	-.217		
Intercept	2.921 (0.085)		2	.012
CBT	-0.631 (0.338)	-.220		
Spontaneous gesture prevalence	0.012 (0.155)	.010		
Instructed gesture prevalence	-0.064 (0.112)	-.074		
Intercept	2.920 (0.086)		3	-.017
CBT	-0.514(0.519)	-.197		
Spontaneous gesture prevalence	0.013 (0.158)	.011		
Instructed gesture prevalence	-0.64 (0.113)	-.074		
CBT × Spontaneous gesture prevalence	-0.159 (1.125)	-.019		
CBT × Instructed gesture prevalence	-0.220 (0.731)	-.051		
4-disc TOH	B (SE)	β	Model	$R^2_{adjusted}$
Intercept	86.346 (6.470)		1**	.154
CBT	-155.747 (42.805)*	-.409		
Intercept	107.694 (12.009)		2**	.185
CBT	-155.252 (42.127)**	-.407		
Spontaneous gesture prevalence	-32.983(17.769)	-.243		
Instructed gesture prevalence	-28.005 (15.076)	-.243		

Intercept	107.708 (12.198)		3*	.160
CBT	-157.104 (69.062)**	-.412		
Spontaneous gesture prevalence	-33.768 (18.221)	-.249		
Instructed gesture prevalence	-28.043 (15.308)	-.244		
CBT × Spontaneous gesture prevalence	33.547 (128.081)	.035		
CBT × Instructed gesture prevalence	-7.447 (93.543)	-.13		

Note. * $p < .05$, ** $p < .01$ In sum, when instructed to gesture or spontaneously gesturing, participants who had a lower visual (but not spatial) working memory capacity performed better on the most difficult TOH trial (in terms of solving speed).

Table 3. Summary of stepwise hierarchical regression analysis for variables predicting log transformed solving speed on easy 3-disc TOH and solving speed on 4-disc TOH.

We also reran the previous analyses with the CBT (reported in Table 3), but no significant interactions of CBT and instructed or spontaneous gesture prevalence were obtained in explaining variance in solving speed on the 3-disc and 4-disc TOH.

Mental effort, Difficulty, and Interest

We assessed whether task complexity indeed affected self-report measures of invested mental effort, experienced difficulty, and interest, serving as a subjective manipulation check of experienced cognitive load. See Table 4 for means and SDs for self-report ratings. Task complexity indeed affected self-report ratings, $F(3, 70) = 123.38$, $p < .001$, $\eta_p^2 = .968$. Pairwise comparisons with Bonferroni corrected p-values showed higher scores on invested mental effort, $M_{diff} = -1.48$, $p < .001$, $\eta_p^2 = .554$, difficulty, $M_{diff} = -1.65$, $p < .001$, $\eta_p^2 = .793$, and self-reported interest, $M_{diff} = -0.690$, $p < .001$, $\eta_p^2 = .968$, on the 4-disc TOH as compared to the 3-disc TOH.

	Mental Effort *M (SD)*	Difficulty *M (SD)*	Interest *M (SD)*
TOH 3	2.05 (0.91)	1.84 (0.71)	2.99 (3.09)
TOH 4	3.53 (0.92)	3.43 (0.90)	3.70 (0.76)
Chess 1	2.78 (1.21)	2.31 (1.29)	2.93 (0.84)
Chess 2	3.27 (1.11)	3.15 (1.15)	3.29 (0.85)
Chess 3	3.29 (1.07)	3.02 (1.08)	3.32 (0.83)

Table 4. Means and standard deviations for self-report ratings of experienced mental effort, difficulty and interest across task complexity on TOH and chess tasks

6

Results Chess

All the analyses performed on the TOH data are repeated here for the chess task.

Overall Performance Chess Tasks

The results and performance distributions of the overall performance on chess task 1, 2, and 3, for solving speed and number of attempts are reported here (see also Table 5).

- insert Table 5 here -

Chess task 1. Overall mean solving speed for the 3-step chess task was 33.16 seconds (SD = 42.44), with overall mean of 1.83 attempts (SD = 1.61), and an intercorrelation of r = .817 (p < .001). Three participants were unable to solve the task, and were ceased in their attempt after 300 s and were therefore not included in the analysis. Solving speed was non-normally distributed with skewness of 2.30 (SE = 0.33) and kurtosis of 5.26 (SE = 0.64). A log transformation was performed on solving speed, which reduced skewness of this measure to 0.81 (SE = 0.33) and kurtosis to -.22 (SE = 0.64). For subsequent analysis we use log transformed solving speed as performance measure (see Table 5).

Chess task 2. Overall mean solving speed for the 4-step chess task 68.71 s (SD = 78.95), with overall mean of 2.96 attempts (SD = 2.98), and an intercorrelation of r = .91 (p < .001). Three participants were unable to solve the task. Solving speed was non-normally distributed with skewness of 1.53 (SE = 0.33) and kurtosis of 1.29 (SE = 0.64). For subsequent analysis we used log transformed solving speed as performance measure (skewness .31, SE = 0.33 and kurtosis -1.16, SE = 0.64).

Chess task 3. Overall mean solving speed for the 5-step chess task was 26.92 s (SD = 38.43), with overall mean of 1.63 attempts (SD = 1.7), and an intercorrelation of r = .95 (p < .001). Eight participants were unable to solve the task. Solving speed for this task was also non-normally distributed with skewness of 4.18 (SE = 0.34) and kurtosis of 18.19 (SE = 0.67). Again, a log transformation was performed on solving speed which reduced skewness to 1.83 (SE = 0.34) and kurtosis to 4.05 (SE = 0.676).

Visual and Spatial Working Memory Capacity

Table 5 displays the averages, standard deviations and correlations between VPT, CBT and the log transformed scores on each of the three chess tasks. There were no differences between the instructed gesture and gesture allowed condition on the proportion correct on the VPT, $t(53) = 0.33$, $p = 0.743$, nor the CBT, $t(52) = 0.842$, $p = .404$. For each task, there was a negative relationship between VPT score and solving speed (i.e., higher VPT associated with faster solving speeds). CBT also showed negative correlations with solving speed on chess task 1 and 3, but not task 2. Furthermore, prior knowledge of chess was associated with faster solving speed on each of the three tasks.

	Mean	SD	1	2	3	4
1.VPT Score	0.52	0.23				
2. CBT Score	0.43	0.16	.45**			
2. Prior knowledge	2.47	1.32	.46**	.38**		
3. Solving speed chess Task1 (log)	2.95	0.97	-.45**	-.32*		
4. Solving speed chess Task2 (log)	3.61	1.14	-.31*	-.20	.43**	
5. Solving speed chess Task3 (log)	2.93	0.69	-.41*	-.32*	.47**	.44**

Note. * $p < .05$, ** $p < .01$

Table 5. Averages, standard deviations and correlations for the VPT, and log transformed solving speed for chess task 1, 2 and 3.

Gesture Production

Here we provide an overview of the percentages and frequency of gestures across task complexity and condition (also see Figure 3 and 4).

On chess task 1, 20% (*N* = 5/25) of the participants in the gesture allowed condition spontaneously gestured during the mental solving phase, with a mean gesture frequency of 8.00 (*SD* = 4.76, minimum frequency = 3, maximum = 13), and a gesture ratio of 2.6 (*SD* = 1.6). In the gesture instructed condition the mean gesture frequency was 14.86 (*SD* = 7.5, minimum frequency = 3, maximum = 30), with a gesture ratio of 4.9 (*SD* = 2.5).

On chess task 2, 32.0% (*N* = 8/25) of the participants in the gesture allowed condition spontaneously gestured during the mental solving phase, with a mean gesture frequency of 9.50 (SD = 5.12, minimum frequency = 4, maximum = 18) and a gesture ratio of 3.0 (*SD* = 1.6). In the gesture instructed condition the mean gesture frequency was 22.84 (*SD* = 7.68, min = 5, max = 37.00) with a gesture ratio of 3.8 (*SD* = 1.3).

On chess task 3, 44% (N = 11/25) of the participants in the gesture allowed condition spontaneously gestured during the mental solving phase, with a mean gesture frequency of 12.20 (*SD* = 8.57, minimum frequency = 4, maximum = 28) and a gesture ratio of 1.4 (*SD* = 1.0). In the gesture instructed condition the mean gesture frequency was 33.36 (*SD* = 17.38, min = 6, max = 77) with a gesture ratio of 3.7 (*SD* = 1.9).

Gesture Likelihood Hypothesis

The gesture likelihood hypothesis predicts that higher percentage of gesture prevalence and frequencies (controlling for time, i.e., gesture ratio) will be observed when participants have a lower visual or spatial working memory capacity (VPT) and in more complex tasks (spontaneous gesture frequency task 3 > 2 > 1).

Spontaneous gesture prevalence likelihood. We assessed whether there was a significant difference in the percentage of prevalence of spontaneous gesturing (as opposed to non-gesturing) across complexity. We performed a dependent Chi square test (Cochran's Q test), which revealed that there were differences across spontaneous gesture prevalence across the three chess trials, $\chi^2(2) = 7.714$, $p = .021$. Pairwise comparisons only showed a significant difference in the likelihood of spontaneous gesture prevalence between chess task 1 and 3, $\chi^2(1) = 2.777$, $p = .016$, and not between task 1 and 2 or 2 and 3, $\chi^2(1) > 1.389$, $p > .495$.

We further assessed whether those with a lower visual or spatial working memory capacity were more likely to spontaneous gesture. This was not the case. Participants in the gesture allowed condition who spontaneously gestured during task 1 did not differ from non-gesturing participants with regard to VPT score, $t(23) = -0.02$, $p = 0.984$. Thus VPT did not relate to gesture prevalence in chess task 1. Participants in the gesture allowed condition who spontaneously gestured during task 2 did not differ from non-gesturing participants with regard to VPT score, $t(23) = 0.52$, $p = .608$. Similarly, participants in the gesture allowed condition who spontaneously gestured during task 3 did not differ from non-gesturing participants with regard to VPT score, $t(23) = 0.48$, $p = 0.636$. Thus VPT did not relate to spontaneous gesture prevalence in any of the three chess tasks. In similar vein, spontaneously gesturing participants during chess task 1, did not differ from non-gesturing participants with regard to CBT score, $t(22) = .226$, $p = 0.823$. Spontaneously gesturing participants during chess task 2, did not differ from non-gesturing participants with regard to CBT score, $t(22) = 1.18$, $p = 0.251$. Spontaneosly gesturing participants during chess task 3, did not differ from non-gesturing participants with regard to CBT score, $t(22) = 2.04$, $p = .054$.

Spontaneous and instructed gesture frequency likelihood. A repeated measures ANOVA was conducted to assess whether task complexity (within-subjects: task 1 vs. task 2 vs. task 3) and gesture type (between-subjects: spontaneous vs. instructed gesture) predicted gesture frequency after controlling for time on task (gesture ratio), which allows us to test whether gesture frequencies were indeed higher in the instructed gesture condition (manipulation check) and whether task complexity affected gesture frequency (gesture likelihood hypothesis). Note that participants are included in this analysis when they have gestured on task 1 or 2 or 3. Thus participants in the gesture allowed condition who did not gesture in any of the tasks are not included. In contrast to the hypothesis, we did not find an effect of task complexity on gesture ratio, $F(2, 58)= 0.33$, $p = 0.679$. However, there was a main between subjects effect of gesture type, $F(1, 29) = 29.94$, $p < .001$, $\eta_p^2 = .508$, showing that gesture ratio was higher for participants who were instructed to gesture ($M_{estimated} = 0.43$, $SE = 0.05$) than for participants who spontaneously gestured ($M_{estimated} = 0.12$, $SD = 0.04$). No interaction effect was obtained between complexity and gesture type, $F(2, 58) = 0.99$, $p = .378$.

We additionally added VPT as a covariate into the repeated measures ANCOVA, however this did not account for variance in frequency of gestures across complexity or gesture prevalence group, $F(2, 56) > .269$, $p > .765$. This was also the case when taking CBT into account as a covariate, $F(2, 54) < .291$, $p > .749$. This suggests that visual working memory capacity was not related to spontaneous (nor instructed-) gesture frequency.

Gesture effect hypothesis

We assessed whether visual working memory capacity (VPT) and gesture prevalence type ('spontaneous' non-gesturing vs. spontaneous gesturing vs. instructed gesturing) affected performance (log solving speed) on each of the chess tasks through three separate multiple stepwise regression analyses. For each regression analysis we entered VPT (mean centered for sample on chess task) in the first step, and dummy variables spontaneous gesture prevalence (0 = non-gesturing, 1 = spontaneous gesturing) and instructed gesture prevalence (0 = non-gesturing, 1 = instructed gesturing) in the second step. In the third step we entered two interaction terms of the centered VPT with instructed and spontaneous gesturing. Again we predicted that gestures would benefit performance in more complex tasks (complexity task 1 < 2 < 3) qualified by interaction effects of VPT with instructed and spontaneous gesture prevalence, such that especially those with a lower VPT scores improve in performance when using gestures (as compared to those who do not gesture).

As can be seen in Table 6, no significant predictors were obtained for solving speed on task 1 in the final model.

a. Chess task 1	B (SE)	β	Model	$R^2_{adjusted}$
Intercept	2.972 (0.120)		1**	.188
VPT	-1.890 (0.525)**	-.451		
Intercept	2.993 (0.204)		2**	.166
VPT	-1.866 (0.533)**	-.444		
Spontaneous gesture prevalence	-0.367 (0.489)	-.101		
Instructed gesture prevalence	0.011 (0.260)	.006		
Intercept	2.993 (0.207)		3*	.140
VPT	-1.934 (1.053)	-.461		
Spontaneous gesture prevalence	-0.427 (0.503)	.117		
Instructed gesture prevalence	0.011 (0.264)	.006		
VPT × Spontaneous gesture prevalence	1.140 (1.984)	.095		
VPT × Instructed gesture prevalence	-0.107 (1.258)	-.020		
b. Chess task 2	B (SE)	β	Model	$R^2_{adjusted}$
Intercept	3.626 (0.150)		1*	.081
VPT	-1.560 (0.662)*	-.313*		
Intercept	4.058 (0.257)		2*	.131
VPT	-1.509 (0.649)*	-.303		
Spontaneous gesture prevalence	-0.313 (0.480)	-.094		
Instructed gesture prevalence	-0.715 (0.324) *	-.316		
Intercept	4.100 (0.248)		3**	.195
VPT	1.127 (1.261)*	.226		
Spontaneous gesture prevalence	-0.322 (0.477)	-.097		
Instructed gesture prevalence	-0.747 (0.312)*	-.331		
VPT × Spontaneous gesture prevalence	-3.062 (2.093)	-.240		
VPT × Instructed gesture prevalence	-3.592 (1.492)*	-.561		
c. Chess task 3	B (SE)	β	Model	$R^2_{adjusted}$
Intercept	2.967 (0.093)		1**	.265
VPT	-1.285 (0.415)**	-.415		
Intercept	3.174 (0.178)		2**	.220
VPT	-1.17 (0.404)*	-.380		
Spontaneous gesture prevalence	.029 (0.153)	-.018		
Instructed gesture prevalence	-0.417 (0.215)	-.303		
Intercept	3.248 (0.171)		3**	.304
VPT	-3.084 (0.934)*	-.996		
Spontaneous gesture prevalence	-0.064 (0.246)	-.039		
Instructed gesture prevalence	-0.530 (0.208)*	-.386		
VPT × Spontaneous gesture prevalence	1.392 (1.180)	.237		
VPT × Instructed gesture prevalence	2.742 (1.066)*	.657		

Table 6. *Summary of hierarchical regression analysis for variables predicting solving speed on task one to three. p < .05, ** p < .01*

For task 2 (see also Figure 6a), in the final model (explaining 19.5% of the variance), instructed gesturing resulted in faster solving speed overall, $t(47) = -2.39$, $p = .021$, *Partial* $r = -.331$. This was not the case for spontaneous gesturing, $t(47) = -0.67$, $p = .506$, *Partial* $r = -.098$. However, we also obtained an interaction of instructed gesturing and VPT $t(47) = -2.41$, $p = .020$, *Partial* $r = -.331$. We performed a Johnson-Neyman Regions of Significance Tests to assess the nature of this interaction. This analysis showed, surprisingly, that especially those with a higher score on the VPT (centered score VPT = -.0104 to .4451, 60.34 % highest VPT scorers), benefited from gesturing when instructed to do so (range b = -0.574 to -1.611, range Se = 0.285 to 0.624, range $t(49)$ = -2.01 to -2.58, range p = .05 to .013).

For task 3 (see also Figure 6b), in the final model (explaining 30.04% of the variance), we again found a main effect of instructed gesturing on log solving speed, $t(47) = -2.55$, $p = .014$, *Partial* $r = -.366$, next to a main overall effect of VPT, $t(47) = -3.30$, $p = .002$, *Partial* $r = -.454$). No effects on solving speed of spontaneous gesturing, $t(47) = -0.26$, $p = 0.796$, *Partial* $r = -.040$, nor an interaction of spontaneous gesturing with VPT was found, $t(47) = 1.18$, $p = 0.244$, *Partial* $r = .179$. Additionally, the interaction term of VPT and instructed gesturing was a significant predictor for solving speed, $t(47) = 2.57$, $p = .014$, *Partial* $r = .369$. Further regions of significance analysis showed a conditional effect of instructed gesture resulting in faster solving speed (range b = -1.350 to -.344, range Se = 0.412 to 0.171, range $t(52)$ = -3.28 to -2.02, range p = .002 to .05) for those with a lower visual working memory capacity, range of significance for centered score VPT = -.475 to -.031 (47.91 % lowest VPT scorers).

It should be noted that it is possible that the effects found in task three are the result of two statistical outliers and should be interpreted more carefully. These participants performed relatively slowly on task 3 (as can be seen in Figure 6b). Yet we find no reason to exclude them based on information other than their performance. Note further that the task ceased when participants took longer than 300 seconds, ensuring the exclusion of extreme outliers (these two participants solving speed was 179 s and 219 s). However, we reran the analysis with these two data points excluded, which resulted in a non-significant fit of the overall regression model, but including a conditional significant effect of instructed gesture (range b = -0.368 to -0.282 range Se = 0.183 to 0.140, range $t(52)$ = -2.018 to -2.018, range p = .05 to .05, lowest p = .040) but for a smaller region, range centered score VPT = -0.133 to 0.033 (21.73% of VPT scorers).

Chess task 1	B (SE)	β	Model	$R^2_{adjusted}$
Intercept	2.990 (0.130)		1*	.082
CBT	-1.987 (0.837)*	-.371		
Intercept	3.093 (0.225)		2	.064
CBT	-2.202 (0.853)*	-.321		
Spontaneous gesture prevalence	-0.528 (0.522)	-.146		
Instructed gesture prevalence	0.109 (0.284)	-.056		
Intercept	3.124 (0.230)		3	.039
CBT	-2.861 (1.335)	-.459		
Spontaneous gesture prevalence	-0.580 (0.535)	-.160		
Instructed gesture prevalence	-0.143 (0.289)	-.068		
CBT × Spontaneous gesture prevalence	1.877 (3.381)	.084		
CBT × Instructed gesture prevalence	1.417 (1.807)	-.160		
Chess task 2	B (SE)	β	Model	$R^2_{adjusted}$
Intercept	3.601 (0.157)		1	.081
CBT	-1.409 (0.991)	-.197		
Intercept	4.147 (0.276)		2	.131
CBT	-1.778 (0.971)	-.249		
Spontaneous gesture prevalence	-0.586 (0.500)	-.177		
Instructed gesture prevalence	-0.835 (0.343)	-.366		
Intercept	4.063 (0.278)		3	.195
CBT	0.063 (1.571)*	.005		
Spontaneous gesture prevalence	-0.638 (0.499)	-.193		
Instructed gesture prevalence	-0.753 (0.342)*	-.330		
CBT × Spontaneous gesture prevalence	-5.627 (2.991)	-.297		
CBT × Instructed gesture prevalence	-2.087 (2.080)	-.201		

(Continued on next page)

Table 7. Summary of hierarchical regression analysis for variables predicting solving speed on chess task 1, 2, 3.

Chess task 3	B (SE)	β	Model	$R^2_{adjusted}$
Intercept	2.953 (0.098)		1*	.084
CBT	-1.418 (0.621)*	-.322		
Intercept	3.327 (0.197)		2**	.201
CBT	-1.602 (0.621)*	-.364		
Spontaneous gesture prevalence	-0.191 (0.285)*	-.117		
Instructed gesture prevalence	-0.618 (0.233)	-.446		
Intercept	3.277 (0.218)		3*	.207
CBT	-1.071 (1.173)	-.243		
Spontaneous gesture prevalence	-0.271 (0.303)	-.166		
Instructed gesture prevalence	-0.570 (0.251)*	-.411		
CBT × Spontaneous gesture prevalence	2.402 (1.810)	-.251		
CBT × Instructed gesture prevalence	-0.93 (1.066)	-0.14		

Note. * $p < .05$, ** $p < .01$.

Table 7 (continued). Summary of hierarchical regression analysis for variables predicting solving speed on chess task 1, 2, 3.

We repeated the hierarchical regression analyses reported above with CBT. As can be seen in Table 4 across all tasks, none of the predicted interactions of CBT with spontaneous or instructed gesture were obtained.

In sum, participants who were instructed to gesture or spontaneously gestured performed better on the chess task 2 and 3 overall (in terms of solving speed) than participants who did not gesture. The moderating effects of visual working memory capacity were mixed, such that stronger effects of instructed (but not spontaneous) gesture were obtained (as compared to no gesture) for those with a higher visual working memory capacity in task 2 (contrary to our predictions), but in task 3 stronger effects were obtained for those with a lower visual working capacity (in line with out predictions).

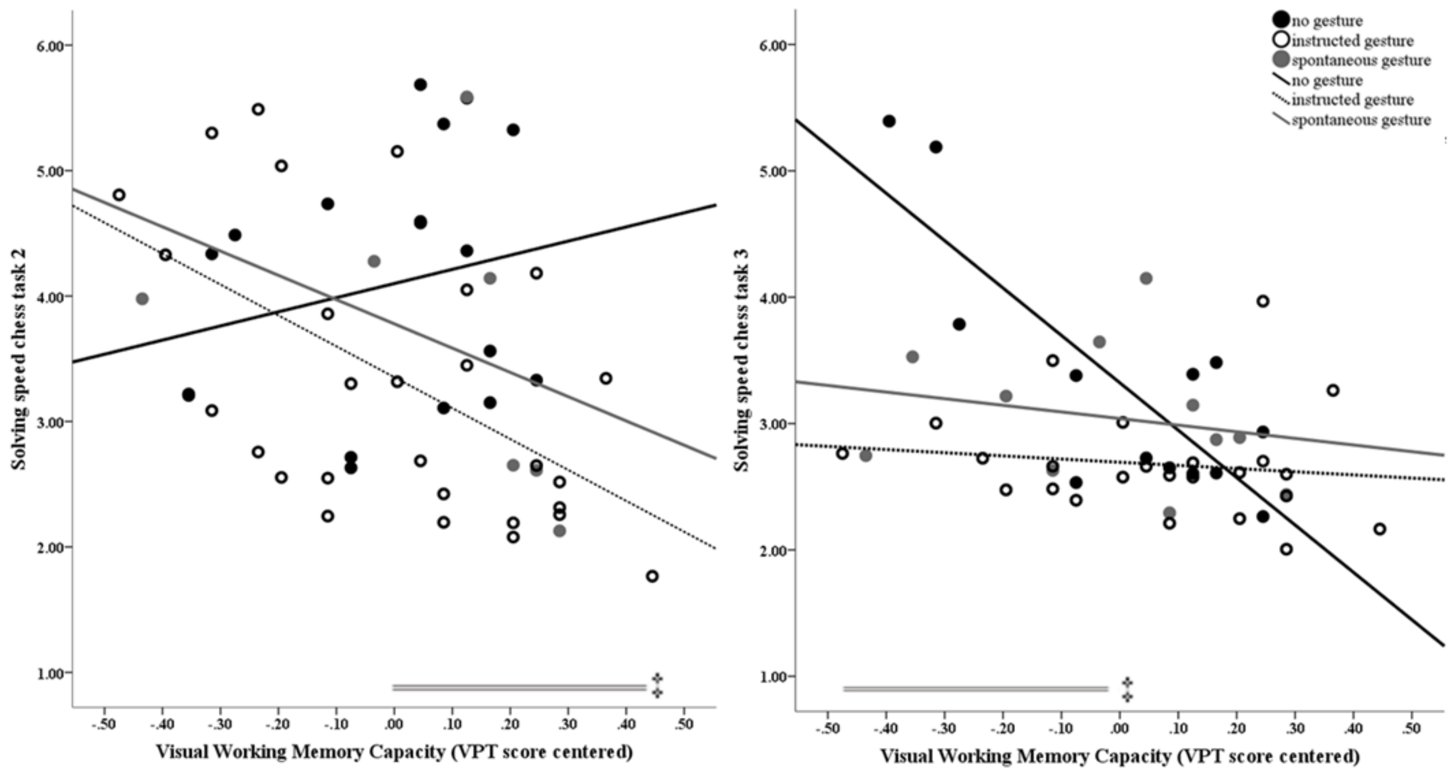


Figure 6a, 6b. Regression slopes for gesture type prevalence (spontaneous vs. instructed vs. no gesture) on log transformed solving speed on the chess task 2 (4-steps) and 3 (5-steps), with centered VPT score on the horizontal axis. Note that regressions lines differ from estimated regression coefficients estimated by the interaction model.

Mental effort, difficulty, and interest

As a subjective manipulation check of experienced cognitive load we assessed whether task complexity indeed affected self-report measures of experienced mental effort, difficulty and interest.

See Table 3 for means and SDs for self-report ratings across chess task complexity and gesture prevalence. Overall, mental effort significantly varied across the three chess tasks, $F(2, 102) = 7.66$, $p < .001$, $\eta_p^2 = .131$. Pair-wise comparisons with Bonferonni corrected *p*-values yielded a significant difference across complexity for the mental effort ratings, where mental effort was rated lower for task 1 compared to task 2, $M_{diff} = -0.48$, $p = .003$, and task 1 compared to task 3, $M_{diff} = -0.48$, $p = .015$. No differences in mental effort were observed between task 2 and 3, $M_{diff} = -0.00$, $p > .999$. Difficulty also varied significantly across complexity, $F(2, 104) = 14.73$, $p < .001$, $\eta_p^2 = .221$, wherein significant differences were found between task 2 and task 1, $M_{diff} = -0.79$ $p < .001$, and between task 1 and task 3, $M_{diff} = -0.70$, $p < .001$. There was no difference in perceived difficulty between task 2 and task 3, $M_{diff} = 0.09$ $p > .999$. The same pattern was observed for the interest ratings, $F(2, 104) = 7.91$, $p < .001$, $\eta_p^2 = .132$. Pairwise comparisons showed a significant difference between the reported interest for task 2 and task 1, $M_{diff} = -0.38$, $p < .001$, and between task 1 and task 3, $M_{diff} = -0.34$, $p = .008$. Again, there was no difference in interest between task 2 and task 3, $M_{diff} = .04$, $p > .999$.

Discussion

We hypothesized that gestures were more likely to arise in (gesture likelihood hypothesis) and positively affect (gesture effect hypothesis) problem solving under conditions of higher cognitive load (i.e., higher task complexity, lower visual or spatial working memory capacity).

Gesture effect hypothesis

In line with the gesture effect hypothesis, our results indicate that participants who gestured on the TOH, either spontaneously or instructed, subsequently performed the task faster than participants who did not gesture of their own accord, but only for those with a lower visual working memory capacity. Interestingly, this interaction effect was not found for spatial working memory capacity. On the more complex chess tasks 2 and 3, instructed gesturing indeed resulted in faster solving speeds overall as opposed to not gesturing. However, in contrast to the TOH task, spontaneous gesturing did not significantly speed up subsequent performance compared to not gesturing, possibly because of the smaller sample of spontaneous gesturers. Moreover, especially subjects with a lower visual (but again not spatial) working memory capacity who gestured when instructed to do so showed improved performance on chess task 3 (which required the most amount of moves to solve the problem) as compared to participants with similar working memory capacity who did not gesture. Interestingly, however, for task 2 those with a higher visual (but not spatial) working memory capacity, being instructed to gesture seemed to be less beneficial for subsequent performance as compared to those who did not gesture with similar visual working memory capacities. Again, no such interactions of gesture were found in relation to spatial working memory capacity.

These findings on co-thought gesture's effect on problem solving lend additional support to the general idea that gesturing may be especially effective when internal cognitive resources are limited (Pouw et al., 2014). Furthermore, gestures seem to provide resources to support, or counteract, limited *visual* rather than *spatial* working memory processes in the current task. Before addressing this interesting unexpected difference between visual and spatial working memory capacity we should address how gestures might have benefited problem solving in the current task. We speculate that when confronted with mentally exploring the solution space of the Tower of Hanoi and Chess task the participants simulate the transformations of the discs/pieces from one place to another through mental imagery. Such mental imagery processes are likely to work in concert with the visual information that is provided by the static presentation of the task set-up, allowing simulated moves to be projected on the pegs or chess board (Kirsh, 2009). Simulating the moves also entails continuously memorizing the positions of the discs/pieces that are moved during the mental simulation. Keeping track of simulated moves therefore requires visual working memory processes, and those who have lower capacities are more likely to lose track of the changing positions of the discs/pieces during their more unstable visual imaginations. We speculate that producing gestures offer stable visual and proprioceptive information regarding the hands in space, that allow a way to spatially index or "temporarily" locking simulated moves represented by the hand in space, thereby alleviating the relatively unstable visual imagery processes that would otherwise fulfill this tracking function. Indeed, in a recent companion study we have found preliminary evidence that when participants with a lower visual working memory capacity are instructed to gesture (vs. not to gesture) during mental preparation of the TOH they produce less eye movements, which suggests that moving the hands allow for a way to stabilize simulations in a less visually demanding way (Pouw, Mavilidi, Van Gog, & Paas, 2016).

The finding that spatial working memory capacity did not interact with an effect of gesture on performance was surprising, and can only be met with substantial speculation. A superficial explanation would be that the pointing gestures in the current case are more potent for keeping active (i.e., locking) the visuo-spatial static positions of the multiple pieces throughout the mental simulation rather than to keep track of the spatial sequential trajectories of the simulated discs/pieces. Yet, this explanation begs the question why pointing gestures would be potent in only this way. After all, these pointing gestures do contain information about spatial sequential trajectories. Additionally, it might be the case that mental imagery involved in the TOH and Chess is especially taxing for visual working memory capacity, as indicated by higher correlations of the VPT with performance as opposed to the CBT, and as such gestures become supportive of this component rather than another. Changing the task to a more spatially taxing task might thus reverse the current result for the differential effect of gesture with regards to visual vs. spatial working memory. Speculations aside, the current findings cannot provide insight into why we find this interesting difference. However, an interesting alignment with the current results is that visual working memory capacity has been found to be more predictive for co-speech gesture frequency than spatial working memory capacity (Chu et al., 2014), corroborating the idea that gestures may be especially potent to alleviate limitations in visual working memory capacity.

This study had some limitations that could lead us to over-interpret the evidence in favor of the gesture effect hypothesis, especially in the case of the chess tasks. Firstly, when designing the chess tasks we attempted to manipulate task complexity by varying the number of moves necessary to transform the begin state of the chessboard into the end state. We assumed that increasing the number of steps would linearly increase task complexity. This was not the case for the last two chess tasks, as participants performed better on chess task 2 than on 3 and no differences in perceived difficulty or invested mental effort (both indices of experienced cognitive load) were obtained. Possibly, there were some task characteristics that we did not control for which also might have influenced task complexity. For instance, in chess task 2, participants could transform the begin state into the end state in two ways, both using legitimate moves of the chess pieces. However, one of these solutions was incorrect because it required the invalid chess action of moving a chess piece of the same color twice in a row. This was something that participants sometimes seemed to forget, resulting in overall slower solving speeds. Perhaps this resulted in different task demands in chess task two as compared to task three. This difference in chess task demands may have also resulted in the gesture and visual working memory capacity interaction we found that is difficult to interpret from our perspective; namely, that especially those with a higher visual working memory capacity benefited from instructed gesturing on chess task 2.

Another objection to our present interpretation is that the current gesture effects can be attributed to the idea that gestures reflect efficient mental simulations rather than contributing to them. We think the present findings are not very supportive for this possibility. Firstly, a salient aspect of the present design is that we measure performance after a gesture manipulation, as such ensuring that gesture effects are in some way causally related to performance as opposed to being an epiphenomenon to some ongoing problem solving process. Additionally, had we only obtained a relation between spontaneous (and not instructed) gestures and performance it could still be that gestures are a consequence of effective mental planning rather than supportive of it. However, we found that manipulating gesture use through instruction showed similar positive gesture effects as spontaneous gestures in the TOH and chess task (although in this case we only obtained effects for instructed gesturing), thus providing evidence for a causal relation of gesture production and performance, rather than mental pre-cursors of gesture and performance (for a discussion see Pouw et al., 2014).

Gesture Likelihood Hypothesis

Our results provide only partial support for the gesture likelihood hypothesis dictating that gesticulation (in terms of prevalence or frequency) is more likely when cognitive load is high. With increasing task complexity, the percentage of participants who spontaneously gestured, increased, as expected, and gesturing was more frequent during the more complex TOH trial. In the chess task, however, no significant differences of gesture frequency were found when controlling for time on task (which was longer on the more complex tasks). Moreover, in contrast to research that suggests that limited working memory capacities relate to higher co-speech gesture frequency, neither visual nor spatial working memory capacity contributed to gesture frequency or prevalence in either the TOH or chess task.

This partial support should also be interpreted with some caution, because the higher the task complexity, the larger the number of steps needed to solve the task. Even though participants indeed experienced higher cognitive load (i.e., higher mental effort and difficulty ratings) in the more complex TOH task, the finding that gesture frequency was higher on the more complex TOH task might be due to the fact that this task required more steps to solve as opposed to the cognitive load the task imposed. Of course, in the current tasks, it is difficult to manipulate complexity without (a) increasing the number of steps to solve the task or (b) keeping the task the same but adding a secondary task to manipulate cognitive load. Future research could resolve this issue by adopting option b in the design (see e.g., Marstaller & Burianová, 2013).

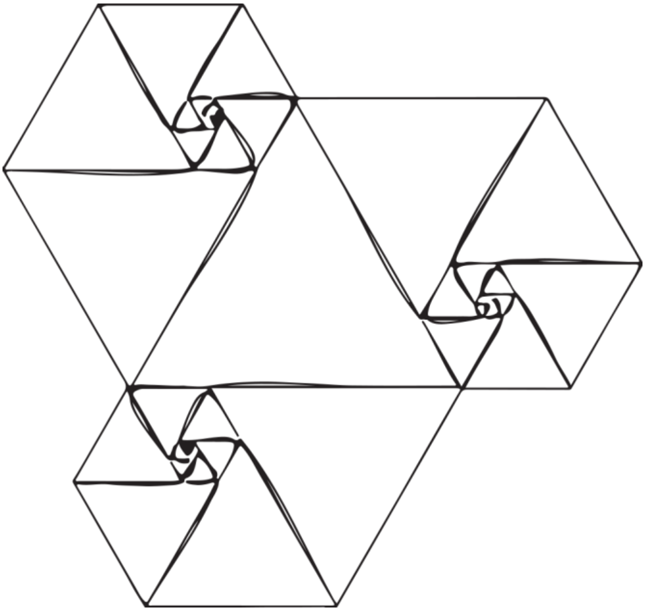
Yet, shortcomings aside, it should be emphasized that the current study shows for the first time that the present physical and digital setup of the TOH and Chess tasks, respectively, solicit spontaneous co-thought gestures. That spontaneous co-thought gesticulation was similar in nature (pointing gestures) across tasks is interesting as Chu & Kita (2016) have shown that when digital mental rotation stimuli are perceived as more manipulable co-thought gestures increase in rate when problem solving with such stimuli. Clearly, more research is needed to ascertain the factors which determine the likelihood and the form of co-thought gesticulation, by differentiating between factors that are inherent to the logical structure of solving the task, and the physical affordances related to solving the task (cf. Cook & Tanenhaus, 2009). Yet, that co-thought gestures spontaneously emerged during the current tasks is an important finding in and of itself, as it provides a paradigm for further investigation into natural occurrences of co-thought gesticulation, next to the few paradigms that are currently used (e.g., mental rotation, route learning). Moreover, in the case of the TOH task the current findings may inspire future research contrasting the cognitive function of co-thought and co-speech gestures. For example, the spontaneous gestures observed in the current TOH task were virtually all deictic (i.e., pointing) gestures, whereas previous research has established that co-speech gestures in explaining solving the TOH are often iconic in nature (e.g., grasping movements; e.g., Cook & Tanenhaus, 2009; Trofatter et al., 2015). This begs the question whether the form (pointing vs. iconic), is connected with different functions relating to problem solving versus speech processes (see e.g., Capuccio, Chu, & Kita, 2013; Chu & Kita, 2016).

Conclusion

The current findings add to the sparse literature on co-thought gestures. Firstly, we offer novel evidence that co-thought pointing- (rather than iconic-) gestures can be spontaneously solicited when participants are confronted with having to mentally solve the current TOH or chess task. Furthermore, gestures seem to be especially productive for problem solving performance when tasks are more complex (as corroborated by previous findings, Chu & Kita, 2008), and interact visual (but not spatial) working memory capacity. Moreover, given the current results that while some individuals already naturally use the body for cognitive benefit, encouraging others to do so may aid problem-solving performance. The present study therefore provides additional support that gestures' cognitive function may go beyond speech processing (a fundamental topic in gesture theory), and may (if developed further) improve educational environments by soliciting embodied problem solving. For instance, from the results we can imagine that current chess learning regimes can be more accommodating for beginning learners (as they experience a higher complexity of the task) when these learners are prompted to think out their moves with their hands.

Chapter 7

Co-thought gestures in children's mental problem solving: Prevalence and effects on subsequent performance*



7

* This chapter has been submitted for publication as Pouw, W., Van Gog, T., Agostinho, S., Zwaan, R. A., & Paas, F. (2016). *Co-thought gestures in children's mental problem solving: Prevalence and effects on subsequent performance*. Manuscript submitted for publication
Open Data: This study has been pre-registered on the Open Science Framework. Pre-registration form, materials, anonymized data-set, and syntax of statistical analyses can be retrieved from https://osf.io/dreks/.

Co-thought gestures in children's mental problem solving: Prevalence and effects on subsequent performance

Co-thought gestures are understudied compared to co-speech gestures, yet may provide insight into possible cognitive functions of gestures that are independent of speech processes. A recent study with adults showed that co-thought gesticulation occurred spontaneously during mental preparation of problem solving. Moreover, co-thought gesturing (either spontaneous or instructed) during mental preparation was effective for subsequent solving of the Tower of Hanoi under conditions of high cognitive load (i.e. when visual working memory capacity was limited and when the task was more difficult). We investigated whether co-thought gestures would also spontaneously occur and would aid problem-solving processes in children (N = 74; 8-12 years old) under high load conditions. Although children also spontaneously used co-thought gestures during mental problem solving, this did not aid their subsequent performance when physically solving the problem. If these null-results are on track, effects of co-thought gestures may be different in adults and children.

Introduction

The majority of our hand-gestures emerge in synchrony with speech, usually in service of some communicative goal. Yet, we also gesture when we think in silence, without any intention of communication. These so-called *co-thought gestures,* which may take the form of pointing to locations/objects, or simulating task-relevant actions (e.g., grasping and replacing), are observed in a variety of non-communicative tasks, such as mental rotation, or remembering a route (e.g., Chu & Kita, 2008; Logan, Lowrie, & Diezmann, 2014). Evidence suggests that such co-thought gestures are not merely epiphenomenal to thinking, because problem solvers' performance has been shown to improve from gesturing (as opposed to not gesturing or being prohibited from gesturing; e.g., Chu & Kita, 2011; Pouw, Eielts, Van Gog, Zwaan, & Paas, under review; So, Shum, & Wong, 2015). Yet, the cognitive function of co-thought gestures is understudied relative to co-speech gestures. As a consequence, it is still unclear when and why co-thought gestures are produced, and whether and how they support cognitive processes.

According to recent evidence, co-thought gestures and co-speech gestures may have a common cognitive origin (Chu & Kita, 2015). That is, the rate with which co-thought gestures are spontaneously produced in a silent mental rotation task corresponds (within participants) with the rate with which co-speech gestures are elicited in the same task. Moreover, when objects are seen as more difficult to physically rotate, both co-speech and co-thought gestures are less likely to spontaneously emerge (as opposed to objects with a more manipulable surface). Such findings hint at a possible common cognitive origin of co-thought and co-speech gestures that are not directly tied to speech processes or communicative intent (cf. McNeill, 2008). Rather, an action-generation system may underlie such gestures observed in mental rotation tasks, which is sensitive to affordances solicited by the objects that is thought or spoken about (Hostetter & Alibali, 2008).

In addition to the issue of which cognitive processes (e.g., action readiness) play a causal role in co-thought gesture production, there is the issue of whether, and if so how, co-thought gestures play a causal role in cognitive processes (Chu & Kita, 2011; Pouw, de Nooijer, Van Gog, Zwaan, & Paas, 2014). That co-thought gestures affect cognitive processes is supported by a handful of studies (e.g., Chu & Kita, 2011; Hegarty, Mayer, Kriz, Keehner, 2005; Logan, Lowrie, & Diezmann, 2014; Pouw, Eielts, Van Gog, Zwaan, & Paas, under review; Schwartz & Black, 1999; So, Shum, & Wong, 2015). For xample, mental rotation co-thought gestures are found to improve mental rotation performance (Chu & Kita, 2011). For example, judging whether a cup of a particular size will spill water can be improved when physically enacting a pouring movement through silent gesture (as opposed to mental inference alone; Schwartz & Black, 1999). Arguably, enacting a grasping movement allows bringing forth procedural knowledge from previous experience, which improved performance in this task. Route learning can be improved when participants rehearse the route on a road map with silent tracing gestures (as opposed to rehearsing it verbally; So et al., 2015). In sum, co-thought gestures are not merely epiphenomenal to cognitive processing; they seem to directly support those very processes.

With regard to when and why co-thought gestures are produced, evidence seems to suggest that - like co-speech gestures - they are used to support problem solving especially when internal cognitive resources are taxed (Pouw et al. de Nooijer, van Gog, Zwaan, & Paas, 2014). That is, compared to not gesturing, co-speech and co-thought gestures are more likely to arise when the problem at hand is more difficult (e.g., Chu & Kita, 2008; Hostetter, Alibali, & Kita, 2007). Such gestures seem to provide additional cognitive resources when engaging in cognitively demanding dual-tasks or when internal cognitive resources such as working memory capacity is low (e.g., Goldin-Meadow, Nusbaum, Kelly, & Wagner, 2001; Marstaller & Burianová, 2013).

In a recent study with adults, the idea was tested that co-thought gestures occur when internal cognitive resources are taxed, either induced by the difficulty of the task, or by limited cognitive capacities (Pouw et al., under review). Participants (aged 18-31) had to mentally prepare for physically solving the Tower of Hanoi at different levels of complexity (3-disc and 4-disc), and were allowed to gesture (i.e. they were not instructed to, but also not prohibited from gesturing) or instructed to gesture. Participants' subsequent problem solving performance (i.e., solving speed) improved when they gestured during mental preparation (either spontaneously or instructed), but only when the TOH task was more difficult, and when they had lower visual working memory capacity (determined via a Visual Patterns Test, explained below). In a subsequent study (Pouw, Mavilidi, van Gog, & Paas, 2016), it was found that adult participants (age 24-50 yrs) who were instructed to gesture during mental problem solving of the TOH, and especially those with a lower visual working memory capacity, were likely to reduce their eye movements during mental problem solving (as indicated by a drop in fixations directed at the Tower of Hanoi display when gesturing vs. not gesturing). This suggests that gesturing provides an alternate (i.e., non-visually guided) motoric strategy to support mentally simulation of the problem solving procedure, which may be especially fruitful for those problem solvers who lack internal cognitive resources for solving the task in a more disembodied way, using purely internal mental inference and imagery processes.

The aim of the current study is to investigate whether our findings with adults, showing that co-thought gestures are effective for fostering subsequent problem solving when cognitive load is high, would be replicated in children (we pre-registered this study). It is important to attempt to replicate these findings in a younger age sample (8-12 years), because children's visual working memory capacity is likely still in development (Gathercole, Pickering, Ambridge, & Wearing, 2004), and as such co-thought gestures could prove to be especially effective for this age sample (see also Paas & Sweller, 2012). Moreover, research on children's spontaneous adoption of co-thought gestures, and the effects on problem solving performance, is particularly scarce (for an exception, see Delgado et al., 2011).

In the current study, we assessed children's visual working memory capacity and then had them solve two consecutive trials of the Tower of Hanoi problem that differed in complexity, and hence, working memory load (3-disc and 4-disc). Before each problem-solving phase children were required to mentally prepare the task, by thinking through the solution in silence - just prior to physically solving the particular TOH trial. One third of the children were instructed to gesture during mental problem solving, whilst the rest were allowed (but not instructed) to gesture, and we assessed whether they spontaneously gestured (spontaneous gesture group), or not (no gesture group). We hypothesized that gesture prevalence during mental problem solving in children (either when children were instructed to gesture, or when spontaneously gesturing) would positively affect actual problem-solving performance (solving speed and solving steps) in all trials (TOH 3-disc and TOH 4-disc) as compared to natural non-gesturing. Furthermore, this effect would be more pronounced on the more complex task (TOH4) and for children with lower visual working memory capacity. We will further refer to this hypothesis as the gesture effect hypothesis.

Method

Participants & Design

This study was approved by the Human Research Committee of the University of Wollongong. Children between 8 and 12 years old were recruited for participation during their visit to a local Science Centre & Planetarium. This is a facility where children can learn about science, engineering and technology through interactive exhibits. The Tower of Hanoi is one of these interactive exhibits and thus was considered an ideal context for this study. The initial plan was to recruit 100 participants, but within the available time at the science centre we recruited 74 participants. We had to exclude 3 participants due to video camera failure, and an additional 3 participants, because they did not fit the age-requirement (younger than 8 or older than 12) years[13]. Additionally, 2 participants were excluded because they failed to comply with the task procedures (e.g., did not engage in the mental solving task), as observed by both the first and second coder of the video data (e.g., being continuously distracted, talking continuously during the mental solving phase). The final sample consisted of 66 participants (38 boys (57.6%), 28 girls (42.2%), M_{age} =9.83, *SD* = 1.20).

As stated in the pre-registration report, we assigned one third of participants to the gesture instruction condition (after exclusion: *N* = 23, 13 boys (56.5%), 10 girls (43.5%), M_{age} = 9.96, *SD* = 1.21), and two thirds to the no gesture instruction condition (after exclusion: *N* = 43, 15 boys (34.9%), 28 girls (64.1%), M_{age} = 9.75, *SD* = 1.21), This was done because participants in the no gesture instruction condition would later be subdivided in the analyses, depending on whether they gestured spontaneously during the mental problem-solving phase (spontaneous gesture group) or not (no gesture group). For the 3-disc Tower of Hanoi (TOH 3) the spontaneous gesture group included 13 participants, vs. 30 in the no gesture group and 23 in the instructed gesture group; in the 4-disc Tower of Hanoi (TOH 4) the spontaneous gesture group included 19 participants, vs. 24 in the no gesture group and 23 in the instructed gesture group.

[13] Although parents were briefed about the age-requirement, due to some miscommunication, some parents enrolled younger/older children for the study.

Materials & Measures

Visual working memory capacity: Visual Patterns Test

The Visual Patterns Test (VPT; Della Sala, Gray, Baddeley, & Wilson, 1997) was used as a measure of visual working memory capacity. We used an adaptation of the VPT developed by Chu and colleagues (2013). Participants were shown a matrix, in various patterns, in which half of the squares were colored black. Each pattern was displayed for 3 seconds, after which all the squares turned white. Subsequently, participants indicated by mouse-clicking on an empty grid the previous pattern of black squares. This was the adaptation made to simplify the task for children as when administered with adults they are required to verbally recall the pattern by naming letters that are assigned to the squares. Participants could select and deselect squares and continue to the next trial when they clicked on 'next' button. The VPT consists of 25 trials, with blocks of five trials per difficulty level (which increased from seven to 11 black squares). Before the start of the task participants were provided with two practice trials (of three and four black squares, respectively). If participants failed to recall one or more of the black squares, the trial was automatically scored as an incorrect response. After five consecutive incorrect responses within one block of trials the task automatically stopped. Performance scores were calculated as the proportion of correct responses out of all trials (i.e., number of correct trials/total number of trials [i.e., 25]).

Tower of Hanoi

The TOH consisted of a structure with a rectangular base with three evenly spaced pegs mounted on top. The task unfolded with a number of differently sized discs (practice trial: two discs, first trial: 3-discs, second trial: 4-discs) placed on the left-most peg; discs were placed decreasing in size from bottom to top. The discs were to be replaced during the problem-solving process onto the right-most peg in the same order (decreasing in size from bottom to top) while following the rules (see procedure). Identical to the procedure used in Pouw et al. (under review), Participants engaged in mental preparation for 2.5 minutes and then physically solved the problem when they took longer than five minutes for physical solving, the trial was aborted.

Gesture. Each participant's mental preparation phases for the 3-disc and 4-disc trials were coded for prevalence (no gesture vs. gesture) and type (pointing vs. iconic) of gesture. However, since there were virtually no iconic gestures observed (there were two exceptions where children momentarily gestured as-if grasping the discs next to performing pointing gestures) we only examined pointing gestures (see Figure 1 for examples of observed gestures).

We counted the number of pointing movements per participant as a measure of gesture frequency in such a way that each rest point after a pointing gesture (either whole hand -or finger-pointing) is considered as one instance and as a trial wherein gesture is prevalent (gesture prevalence). It should be noted that participants were able to, and sometimes did touch the rectangular base of the TOH (or just a place on the table) thereby "marking" a place through pointing gesture instead of pointing only in the air (these pointing-touch gestures were counted as pointing gestures). Additionally, there were four participants who did not point during mental preparation but when they asked a question to the experimenter during this session they did use pointing gestures. These gestures were not considered as gestures during mental preparation. An independent coder counted all the gesture instances in the sample and the first author recounted gestures of 15.15% of the participants who gestured to check for interrater reliability (interrater reliability was high, $r = .992$, $p < .001$).

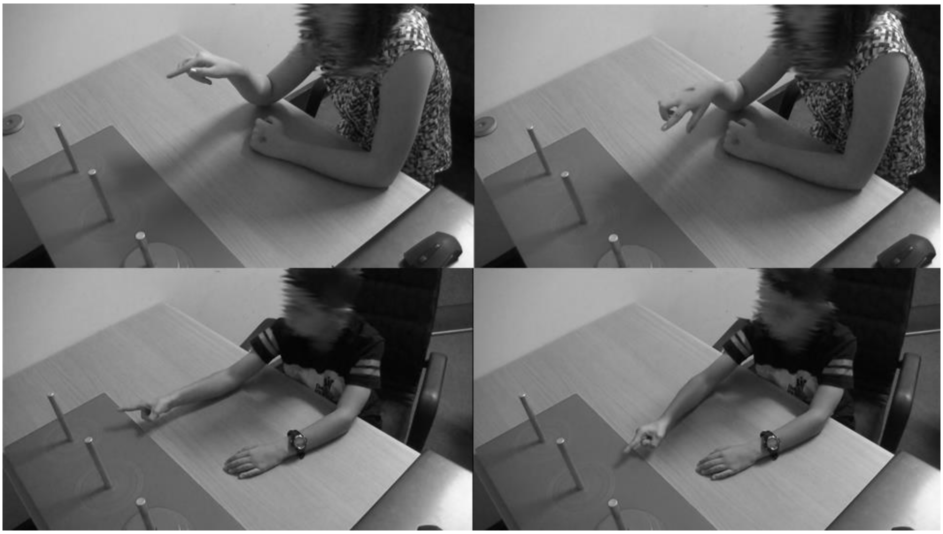


Figure 1. Two examples of spontaneous gestures arising during mental problem solving for the TOH (1 frame per second). Faces blurred for anonymization.

Self-report Measures: Mental effort, difficulty, interest

We obtained an indication of experienced mental effort, perceived difficulty, and experienced interest after the 3-disc as well as the 4-disc TOH problem-solving trial. These self-reports (for a discussion on self-report and cognitive load see Paas, Tuovinen, Tabbers, & Gerven, 2003) were verbally reported and administered by the experimenter on a 5-point rating scale: "How hard did you need to think to solve the task? (mental effort; 1 = 'not hard', to 5 = 'very hard'), "How difficult did you find this task" (difficulty; 1 = 'not difficult', to 5 = 'very difficult'), and "How interesting did you find this task" (interest; 1 = 'not interesting', to 5 = 'very interesting').

Prior experience. Before the start of the TOH task participants were asked whether they had played the Tower of Hanoi before ('no' responses: 83.8 %, 'yes' responses: 16.2%; yes responses were equally distributed among the no gesture instruction condition 16.6 % 'yes' and the gesture instruction condition 15.9% 'yes').

Procedure

The current study was conducted at the local science centre. Visiting parents/caregivers (from hereon: caregivers) and their children were invited to participate in a 30 minute study about problem solving. Caregivers were asked whether (one of) their children was between 8-12 years old, and were further informed about the nature of the study and given an information sheet and consent form. Additionally, caregivers were recruited and informed about the study via email through the member's list of the centre, upon which an appointment could be made to conduct the study at the science centre. As a reward for their participation, children were given a discount at the science centre's shop of 6 Australian Dollars.

When caregivers and children agreed on participation they were directed to a more quiet space with minimal background noise, and no other visitors present. However, some background noise of other visitors and exhibitions was unavoidable. The caregivers, were allowed to be (and often were) present during the experiment, and were seated behind the child and were asked to not help the child or otherwise intervene during the study in any way. Children were seated at a desk, with a computer, a video camera and the TOH present and were informed about the tasks. Additionally, children were told that there were no right or wrong answers, that results would not be shared with their caregivers, and that they were able at all times to abort the study.

In the first phase of the study children performed the VPT. Children would first read the instruction, upon which the experimenter verbally repeated it, and they proceeded with the VPT-practice trials. If children made a mistake during the practice trials, the experimenter would explain what went wrong and in such cases children re-did the practice trials until successfully completed. Children were then told that a longer series of trials would commence and that they would do that on their own for about 5 minutes, after which they would proceed to the next task.

Subsequently, children engaged in the TOH problem-solving trials. The experimenter explained the nature of the task with a practice example of 2-discs. The children were told that each trial would involve the experimenter putting a tower of discs on the left-most peg from large to small (i.e., smaller discs on top). They would need to solve the task by replacing the discs to the outer most right peg while taking two rules into consideration. Firstly, only one disc can be moved at a time. Secondly, only smaller discs can be put upon larger discs, not the other way around, and discs of any size could always be put on empty pegs. Additionally, the experimenter demonstrated that you could always move discs back to the original peg if needed, as long as the rules were not violated. Children then performed the practice TOH 2-disc trial. When children were unable to solve the task (which did not happen often), the experimenter repeated the instruction and children redid the practice trial until successfully completed. Subsequently, children were presented with a TOH 3-disc problem, and the rules of the task were repeated. They were told they would have to solve the puzzle as fast and accurately as possible, but before doing so they could prepare their solution for 2.5 minutes (150 seconds) without physically interacting with the TOH. This was called the mental preparation phase, and children were informed that thinking out the moves before the actual task could help them understand the problem. Participants in the gesture instructed condition were explicitly instructed to think with their hands using pointing gestures (as demonstrated by the experimenter who performed several pointing movements in the air directed at the TOH apparatus) during this mental preparation phase. Participants in the no gesture instruction condition were not given this additional instruction. During the mental preparation phase(s), the participant's hand gestures that (spontaneously) emerged during thinking out the solution for 2.5 minutes were video-recorded. Note, that sometimes children asked a question during the mental preparation phase. If the question concerned the task rules, the experimenter would give the answer, and instruct the child to (mentally) work again on the problem.

Directly after each preparation phase participants solved the task by physically manipulating the TOH (which was also video-recorded). This procedure, that is, a 2.5 minute mental preparation phase followed by actual problem solving, was performed first with the TOH 3-disc, and then with the more difficult TOH 4-disc.

During the actual solving of the TOH tasks, participants had to solve the task within 5 minutes (they were not informed about this time constraint to avoid them experiencing pressure). If they were not able to solve the task in time (which was sometimes the case on the TOH 4-disc), the trial was aborted, but the experimenter would give pointers to the child on how to finish the task. Although the rules of the TOH were mastered quite easily by the children, they did sometimes make a mistake with regard to one of the rules, and the experimenter would instruct the child to look again whether they did not violate any rules. This always led to self-correction upon which the child could proceed further.

After the TOH 3-disc and 4-disc procedure children were given a final task for exploratory purposes; it involved another 4-disc task with the rules inversed preceded by a self-explanation phase. Data on this task will not be reported here (as registered in the pre-registration report). Finally, children were informed about the nature of the study, thanked and awarded with a voucher for their participation, and contact details of caregivers were gathered for future communication of the results of the study (reported at group level, never on individual children).

Outliers

As mentioned in the pre-registered report, it is likely that the current sample will have considerable variability in visual working memory capacity and problem-solving competence. We used a similar procedure to control for extreme outliers as used by Chu, Meyer, Foulkes, and Kita (2013; see also Miyake et al., 2000). For each variable included in the regression analysis (VPT, solving speed, number of solving steps) any value laying 3 standard deviations under (or above) the mean will be set to exactly 3SD under (or above) the mean. This trimming procedure allows us to prevent loss of data with extreme values, without dramatically biasing the results.

Results

The results are divided in three parts. First, relevant descriptive statistics and correlations are reported. Subsequently we report the confirmatory analysis according to plan (see pre-registration report).

Descriptive Statistics and Correlations

TOH Problem-Solving Performance

All participants were able to solve the TOH 3-disc within 5 minutes. Mean solving time was 46.33 seconds (s) (*SD* = 0.72 s, observed range 9-272 s; one outlier > 3SD above the mean was observed and replaced as reported above; new range: 9-173). The mean number of solving steps for TOH 3-disc was 10.02 (*SD* = 5.16, observed range: 7-26; one outlier > 3SD above the mean was observed and replaced; new range: 7-25). Furthermore, 62.1% (41/66) of the children solved the TOH 3-disc in the fastest way possible (i.e., 7 steps).

Seven participants (10.6%) were not able to solve the TOH 4-disc within 5 minutes, so for these participants no score was obtained. Mean solving time was 111.02 s (*SD* = 65.84, observed range: 26-278 s; no outliers). The mean number of solving steps for the TOH 4-disc was 31.83 (*SD* = 13.22, observed range: 15-74; one outlier > 3SD above the mean was observed and replaced; new range: 15-63). Only 4.5% (3/58) of the participants were able to solve the task in the minimal number of moves (i.e., 15 steps).

VPT. The mean score on the VPT was .42 (*SD* = .20, observed range: .04-.92; no outliers were observed).

Gesture Production. In the no gesture instruction condition during the TOH 3-disc, 30.95% (13/42)[14] of the participants spontaneously gestured during the mental preparation phase with a mean gesture frequency of 48.85 (*SD* = 13.57; observed range: 2-166). In the gesture instruction condition, every participant gestured, and the mean gesture frequency was 77.13 (*SD* = 29.37, observed range: 27-134).

On the TOH 4-disc, 44.19% (19/43) of the participants spontaneously gestured during the mental preparation phase in the no gesture instruction condition, with a mean gesture frequency of 43.211(*SD* = 34.82, observed range 7-123). In the instructed gesture condition, all participants gestured, and the mean gesture frequency was 72.09 (*SD* = 34.79, observed range 35-158).

---

[14] Note that analysis including gesture and performance on the TOH3, one additional participant was excluded because of a camera failure during this trial (but not the subsequent TOH4 trial).

Gesture Production Relative to Adults. Figure 2 provides the graphical presentation of the gesture production data presented above, with an additional comparison of the gesture production rate (and percentage of spontaneous gesturers) of our previous study with the adult sample (Pouw et al., under review). Note that exactly the same procedure was used, thus a direct comparison is informative. Informal inspection of the figure shows that the likelihood that spontaneous gestures are adopted in the no instruction group seems to be comparable across age sample (i.e. there is only a 1.64% [TOH3] and 2.09% [TOH4] difference in the likelihood that participants spontaneously gesture). Additionally, it seems that children had a similar gesture frequency as adults on the 4-disc Tower of Hanoi, but a higher gesture rate on the 3-disc Tower of Hanoi. We will return to this interesting finding from the comparison in the discussion.

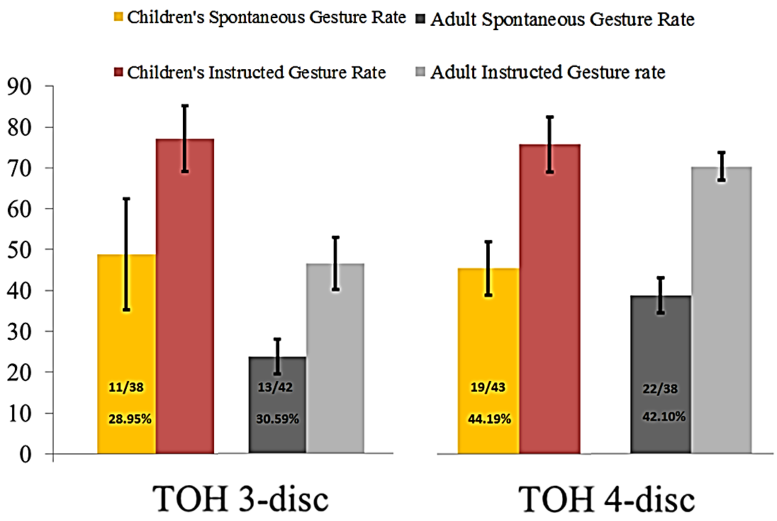


Figure 2. Gesture frequency, and spontaneous gesture likelihood on the TOH 3-disc and 4-disc Gesture production data for the current sample with children versus the adult sample as reported in (Pouw et al., under review). Next to the gesture rate during each mental preparation trial (150 seconds of mental preparation) per age group, the number of spontaneous gesturers (i.e., the number of participants from the total no instruction group that spontaneously adopted gesture) is presented.

Correlations between Problem-Solving Performance, VPT, and Gesture. Table 1 shows the overall correlations and correlations-per-gesture-group (no gesture-, spontaneous gesture-, and instructed gesture group) between problem-solving performance on the TOH 3-disc and 4-disc, VPT, and gesture frequency. As can be seen, overall, VPT is not a strong predictor for problem solving performance, showing only a significant correlation with solving steps on the TOH 4-disc (p = .033), such that participants with higher VPT scores solved the problem in a lower number of steps. Not surprisingly, solving speed and number of solving steps needed to solve the TOH tasks were highly correlated. A more interesting finding of the correlation analyses is that for participants in the instructed gesture group, the number of gestures performed (i.e., gesture frequency) was highly correlated with problem solving performance on TOH 3-disc, such that higher gesture frequencies resulted in faster solving times (p = .001) and a lower number of steps needed to solve the task (p = .019). However, surprisingly, this was not the case for those participants who spontaneously gestured. Furthermore, this effect of instructed gesture frequency on problem-solving performance was not replicated for the more difficult TOH 4-disc task.

Overall	Mean(*SD*)	1.	2.	3.	4.	5.
1. VPT	.41(.20)		-.074	-.014	-.328	**-.467***
2. Solving Time	47.75(47.42)			**.734****	.376	.244
3. Solving Steps	9.95(4.92)				.238	.127
4. Solving Time	103.78(63.59)					.127
5. Solving Steps	30.46(13.09)					
No Gesture Group TOH 3-disc (*N* = 29)	Mean(*SD*)	1.	2.	3.		
1.VPT	.41(.20)		-.174	-.004		
2. Solving Time	48.63(46.91)			**.715****		
3. Solving Steps	9.97(5.28)					
Spontaneous Gesture Group TOH 3-disc (*N* = 13)	Mean(*SD*)	1.	2.	3.	4.	
1. VPT	.41(.20)		-.361	-.360	.106	
2. Solving Time	48.15(51.52)			**.880***	-.042	
3. Solving Steps	10.15(4.36)				-.154	
4. Gesture Frequency	48.85(48.92)					
Instructed Gesture Group TOH 3-disc (*N* = 23)	Mean(*SD*)	1.	2.	3.	4.	
1. VPT	.45(.21)		-0.74	-.014	.152	
2. Solving Time	33.87(28.77)			**.734****	**-.634****	
3. Solving Steps	9.44(4.51)				**-.483***	
4. Gesture Frequency	77.13(29.37)					
No Gesture Group TOH 4-disc (*N* = 21)	Mean(*SD*)	1.	2.	3.		
1.VPT	.42(.21)		**-.469***	-.404		
2. Solving Time	111.24(63.78)			**.95****		
3. Solving Steps	31.57(13.59)					
Spontaneous Gesture Group TOH 4-disc (*N* = 16)	Mean(*SD*)	1.	2.	3.	4.	
1. VPT	.45(.15)		-.070	.347	.105	
2. Solving Time	94.00(64.05)			**.878****	-.333	
3. Solving Steps	29.00(12.70)				-.258	
4. Gesture Frequency	44.19(37.42)					
Instructed Gesture Group TOH 4-disc (*N* = 21)	Mean(*SD*)	1.	2.	3.	4.	
1. VPT	.45(.21)		-.328	**-.467***	.231	
2. Solving Time	123.76(69.34)			.738	-.243	
3. Solving Steps	32.23(9.84)				-.148	
4. Gesture Frequency	73.381(31.73)					

Table 1. *Means (and SD) and Correlations of Problem-Solving Performance, VPT, and Gesture. Note.* ** $p < .01$, * $p < .05$. When TOH 4-disc is concerned we only report means and correlations for participants who were able to solve the task.

7

Self-Report Data. Overall means for the TOH 3-disc and 4-disc are reported in Table 2. As can be expected the TOH 4-disc was generally reported to be more difficult, t (65) = -9.236, $p < .001$ (paired t-test), and to require more mental effort, t (65) = -8.443, $p < .001$ than the TOH 3-disc. The children found the TOH 4-disc more interesting than the TOH-3 disc, t (65) = -2.345, $p = .022$.

	TOH 3-disc	TOH 4 Hanoi 4-disc
Overall	Mean(*SD*)	Mean(*SD*)
1. Difficulty	2.56(0.93)	3.86(0.90)
2. Mental Effort	3.02(1.09)	4.08(0.89)
3. Interest	4.21(0.81)	4.42(0.70)
No Gesture Group	Mean(*SD*)	Mean(*SD*)
1. Difficulty	2.41(0.63)	3.75(0.99)
2. Mental Effort	2.76(1.09)	3.92(1.02)
3. Interest	4.241(0.83)	4.38(0.71)
Spontaneous Gesture Group	Mean(*SD*)	Mean(*SD*)
1. Difficulty	2.77(1.09)	3.89(0.88)
2. Mental Effort	3.46(0.78)	4.11(0.81)
3. Interest	4.08(0.95)	4.37(0.68)
Instructed Gesture Group	Mean(*SD*)	Mean(*SD*)
1. Difficulty	2.57(1.12)	3.95(0.84)
2. Mental Effort	3.13(1.18)	4.23(0.81)
3. Interest	4.21(0.74)	4.50(0.74)

Table 2. Means (and SD) of Self-reported Difficulty, Mental Effort, and Interest

Confirmatory Hypothesis Testing

The following analyses were registered in the pre-registration report.

Hypothesis 1

Our first hypothesis was that after controlling for age, visual working memory capacity would be positively related to problem-solving performance (i.e., that higher VWM would result in faster TOH solving speed and less steps needed to solve the task). We assessed whether this was the case with step-wise multiple regressions analyses, entering age and VWM as predictors for performance. As can be seen in Table 3, no significant effects of age or VPT score on performance (i.e., solving speed and solving steps) were obtained on either the TOH 3-disc or TOH 4-disc. This was unexpected and suggests that visual working memory capacity and age were not strong predictors for performance on the problem-solving task.

Hypothesis 2

Our second hypothesis was that after controlling for age, gesture prevalence (instructed and spontaneous) during mental problem solving would positively affect actual problem-solving performance (i.e., faster solving speed and less solving steps) in both trials (TOH 3-disc and TOH 4-disc) as compared to natural non-gesturing, but that this effect would be more pronounced on the most complex task (TOH4) and for participants with a relatively lower visual working memory capacity. We analysed whether this was the case with two multiple stepwise regression analyses per DV. For each DV (3-disc and 4-disc solving speed and number of steps) we first looked at the combined gesture effect (as opposed to parsing out the effects of instructed vs. spontaneously gesturing) as this is the most powerful analysis to assess the gesture effect hypothesis. In the first step of the stepwise regression analysis we entered age, and VPT (centered) as a predictor, adding gesture prevalence (coding: 0 'no gesture', 1 'gesture') in the second step and the third step the interaction term of the centered VPT and gesture prevalence, as predictors for solving speed. The results of these regression analyses are reported in Table 3. Age, VPT score, gesture prevalence, and the interaction of VPT and gesture prevalence, were unreliable predictors for performance (i.e., solving speed and solving steps) on both the TOH 3-disc and 4-disc.

Solving Speed TOH 3-disc	Step	B	(*SE*)	*β*	*t*	*p*	R^2_{adj}
1. Constant	1	109.376	51.580		2.121	**.038**	.034
2. Age		-6.696	5.215	-.188	-1.284	.204	
3. VPT (Centralised)		-20.511	31.019	-.097	-.661	.511	
1. Constant	2	107.996	51.867		2.082	**.042**	.025
2. Age		-6.170	5.302	-.173	-1.164	.249	
3. VPT (Centralised		-21.294	31.188	-.100	-.683	.497	
4. Gesture (0 = no gesture, 1 = gesture)		-6.854	10.576	-.081	-.648	.519	
1. Constant	3	108.005	52.299		2.065	**.043**	.009
2. Age		-6.171	5.346	-.173	-1.154	.253	
3. VPT (Centralised)		-20.690	43.135	-.098	-.480	.633	
4. Gesture (0 = no gesture, 1 = gesture)		-6.852	10.664	-.081	-.642	.523	
5. Interaction VPT & Gesture		-1.089	53.172	-.004	-.020	.984	
Solving Steps TOH 3-disc	Step	B	(*SE*)	*β*	*t*	*p*	R^2_{adj}
1. Constant	1	6.312	5.979		1.056	.295	-.021
2. Age		.357	.604	.089	.590	.557	
3. VPT (Centralised)		-2.892	3.596	-.121	-.804	.424	
1. Constant	2	6.241	6.029		1.035	.305	-.036
2. Age		.384	.616	.095	.623	.536	
3. VPT (Centralised		-2.932	3.625	-.123	-.809	.422	
4. Gesture (0 = no gesture, 1 = gesture)		-.352	1.229	-.037	-.286	.776	
1. Constant	3	6.265	6.068		1.032	.306	-.050
2. Age		.382	.620	.095	.617	.540	
3. VPT (Centralised)		-1.324	5.005	-.055	-.264	.792	
4. Gesture (0 = no gesture, 1 = gesture)		-.346	1.237	-.036	-.280	.781	
5. Interaction VPT & Gesture		-2.897	6.169	-.090	-.470	.640	
Solving Speed TOH 4-disc	Step	B	(*SE*)	*β*	*t*	*p*	R^2_{adj}
1. Constant	1	187.002	80.267		2.330	**.024**	.083
2. Age		-7.475	8.106	-.140	-.922	.360	
3. VPT (Centralised)		-81.942	51.162	-.243	-1.602	.115	
1. Constant	2	185.963	83.492		2.227	.030	.066
2. Age		-7.428	8.232	-.139	-.902	.371	
3. VPT (Centralised		-82.262	52.006	-.243	-1.582	.120	
4. Gesture (0 = no gesture, 1 = gesture)		.901	17.536	.007	.051	.959	
1. Constant	3	182.121	84.271		2.161	**.035**	.055
2. Age		-7.031	8.311	-.131	-.846	.401	
3. VPT (Centralised)		-114.036	75.769	-.338	-1.505	.138	
4. Gesture (0 = no gesture, 1 = gesture)		.194	17.687	.001	.011	.991	
5. Interaction VPT & Gesture		51.580	88.955	.118	.580	.564	

Table 3. Analyses of Hypotheses 1 & 2 (continues on next page)

Solving Steps TOH 4-disc	Step	B	(*SE*)	*β*	*t*	*p*	$R^2_{adjusted}$
1. Constant	1	39.973	14.973		2.670	**.010**	.033
2. Age		-.867	1.512	-.089	-.573	.569	
3. VPT (Centralised)		-12.164	9.544	-.198	-1.275	.208	
1. Constant	2	40.585	15.571		2.606	**.012**	.015
2. Age		-.895	1.535	-.092	-.583	.562	
3. VPT (Centralised		-11.976	9.699	-.195	-1.235	.222	
4. Gesture (0 = no gesture, 1 = gesture)		-.531	3.270	-.022	-.162	.872	
1. Constant	3	39.247	15.592		2.517	**.015**	.019
2. Age		-.757	1.538	-.078	-.492	.625	
3. VPT (Centralised)		-23.036	14.019	-.375	-1.643	.106	
4. Gesture (0 = no gesture, 1 = gesture)		-.777	3.272	-.031	-.237	.813	
5. Interaction VPT & Gesture		17.955	16.459	.225	1.091	.280	

Table 3. Analyses of Hypotheses 1 & 2 (Continued)

Hypothesis 3

Our third hypothesis was that after controlling for age, instructed and spontaneous gesture prevalence during mental problem-solving would positively affect actual problem-solving performance (solving speed) for all trials (TOH disc 3 and TOH disc 4) as compared to spontaneous non-gesturing, but that this effect would be more pronounced on the most complex task (TOH4) and for participants with a relatively lower visual working memory capacity. We analysed effects of spontaneous vs. no gesture and instructed vs. no gesture prevalence on performance (solving speed and solving steps) on the TOH 3-disc and 4-disc multiple stepwise regression analyses. For each DV (3-disc and 4-disc solving speed and number of steps) we entered age, and VPT (centered) in the first step, and dummy variables spontaneous gesture prevalence (0 = no gesture, 1 = spontaneous gesture) and instructed gesture prevalence (0 = no instructed gesture, 1 = instructed gesture) in the second step. In the third step we entered two interaction terms of the centered VPT with instructed and spontaneous gesture prevalence.

The regression analyses results for hypothesis 3 are shown in Table 4, and each analysis is visualized (without controlling for age) in Figure 3. Again, we did not find any significant results for the overall model fit for this set of predictors on any of the performance measures. Indeed, age, VPT score, spontaneous and instructed gesture prevalence, and the interaction of VPT with spontaneous or instructed gesture prevalence were unreliable predictors for performance (solving speed and solving) on the TOH 3-disc and 4-disc.

We did however find one significant interaction effect as predicted; of children scoring lower on the VPT, those who spontaneously gestured needed less steps to solve the TOH 4-disc as compared to children scoring lower on the VPT who did not gesture. However, this finding should be interpreted with caution given the low reliability of the overall model fit. In sum, our confirmatory analyses did not replicate previous findings with this younger age sample.

Solving Speed TOH 3-disc (See Figure 3a)	Step	B	(*SE*)	*β*	*t*	*p*	$R^2_{adjusted}$
1. Constant	1	109.376	51.580		2.121	**.038**	.064
2. Age		-6.696	5.215	-.188	-1.284	.204	
3. VPT (Centralised)		-20.511	31.019	-.097	-.661	.511	
1. Constant	2	113.119	52.113		2.171	**.034**	.086
2. Age		-6.699	5.327	-.188	-1.257	.213	
3. VPT (Centralised)		-17.888	31.370	-.084	-.570	.571	
4. Spontaneous Gesture (0 = no gesture, 1 = spontaneous gesture)		2.649	14.190	.025	.187	.853	
5. Instructed Gesture (0 = no gesture, 1 = instructed gesture)		-12.030	11.764	-.137	-1.023	.311	
1. Constant	3	117.751	52.494		2.243	**.029**	.108
2. Age		-7.178	5.367	-.201	-1.338	.186	
3. VPT (Centralised)		-17.506	43.023	-.083	-.407	.686	
4. Spontaneous Gesture		2.317	14.266	.022	.162	.872	
5. Instructed Gesture		-12.757	11.840	-.145	-1.077	.286	
6. Interaction VPT & Spontaneous Gesture		-57.844	72.731	-.118	-.795	.430	
7. Interaction VPT & Instructed Gesture		32.810	58.869	.094	.557	.579	

Solving Steps TOH 3-disc (see Figure 3 b)	Step	B	(*SE*)	*β*	*t*	*p*	$R^2_{adjusted}$
1. Constant	1	6.312	5.979		1.056	.295	-.021
2. Age		.357	.604	.089	.590	.557	
3. VPT (Centralised)		-2.892	3.596	-.121	-.804	.424	
1. Constant	2	6.437	6.103		1.055	.296	-.052
2. Age		.363	.624	.090	.583	.562	
3. VPT (Centralised)		-2.802	3.674	-.117	-.763	.449	
4. Spontaneous Gesture		.012	1.662	.001	.007	.994	
5. Instructed Gesture		-.549	1.378	-.055	-.399	.691	
1. Constant	3	6.813	6.178		1.103	.275	-.072
2. Age		.326	.632	.081	.516	.608	
3. VPT (Centralised)		-1.145	5.063	-.048	-.226	.822	
4. Spontaneous Gesture		-.040	1.679	-.003	-.024	.981	
5. Instructed Gesture		-.585	1.393	-.059	-.420	.676	
6. Interaction VPT & Spontaneous Gesture		-7.558	8.560	-.136	-.883	.381	
7. Interaction VPT & Instructed Gesture		-.335	6.928	-.008	-.048	.962	

Table 4. Analyses of Hypothesis 3 (continues on next page) (see Figure 3 a)

Solving Speed TOH 4-disc (See Figure 3b)	Step	B	(*SE*)	*β*	*t*	*p*	$R^2_{adjusted}$
1. Constant	1	187.002	80.267		2.330	**.024**	.083
2. Age		-7.475	8.106	-.140	-.922	.360	
3. VPT (Centralised)		-81.942	51.162	-.243	-1.602	.115	
1. Constant	2	204.768	83.195		2.461	**.017**	.092
2. Age		-9.308	8.205	-.174	-1.134	.262	
3. VPT (Centralised)		-75.900	51.448	-.225	-1.475	.146	
4. Spontaneous Gesture (0 = no gesture, 1 = spontaneous gesture)		-18.481	21.179	-.127	-.873	.387	
5. Instructed Gesture (0 = no gesture, 1 = instructed gesture)		14.885	19.414	.110	.767	.447	
1. Constant	3	197.225	84.905		2.323	**.024**	.067
2. Age		-8.542	8.377	-.160	-1.020	.313	
3. VPT (Centralised)		-107.979	75.492	-.320	-1.430	.159	
4. Spontaneous Gesture		-20.377	21.640	-.140	-.942	.351	
5. Instructed Gesture		14.755	19.760	.109	.747	.459	
6. Interaction VPT & Spontaneous Gesture		97.373	127.453	.118	.764	.448	
7. Interaction VPT & Instructed Gesture		31.584	94.911	.061	.333	.741	

Solving Steps TOH 4-disc (see Figure 3d)	Step	B	(*SE*)	*β*	*t*	*p*	$R^2_{adjusted}$
1. Constant	1	39.973	14.973		2.670	**.010**	.033
2. Age		-.867	1.512	-.089	-.573	.569	
3. VPT (Centralised)		-12.164	9.544	-.198	-1.275	.208	
1. Constant	2	42.647	15.755		2.707	**.009**	.012
2. Age		-1.101	1.554	-.113	-.709	.482	
3. VPT (Centralised)		-11.278	9.743	-.184	-1.158	.252	
4. Spontaneous Gesture		-2.656	4.011	-.100	-.662	.511	
5. Instructed Gesture		1.003	3.676	.041	.273	.786	
1. Constant	3	38.478	15.306		2.514	**.015**	.08
2. Age		-.680	1.510	-.070	-.450	.654	
3. VPT (Centralised)		-23.345	13.609	-.380	-1.715	.092	
4. Spontaneous Gesture		-3.850	3.901	-.145	-.987	.328	
5. Instructed Gesture		1.201	3.562	.049	.337	.737	
6. Interaction VPT & Spontaneous Gesture		53.566	22.976	.358	2.331	**.024**	
7. Interaction VPT & Instructed Gesture		4.247	17.110	.045	.248	.805	

Table 4 (continued). Analyses of Hypothesis 3 (see Figure 3 a)

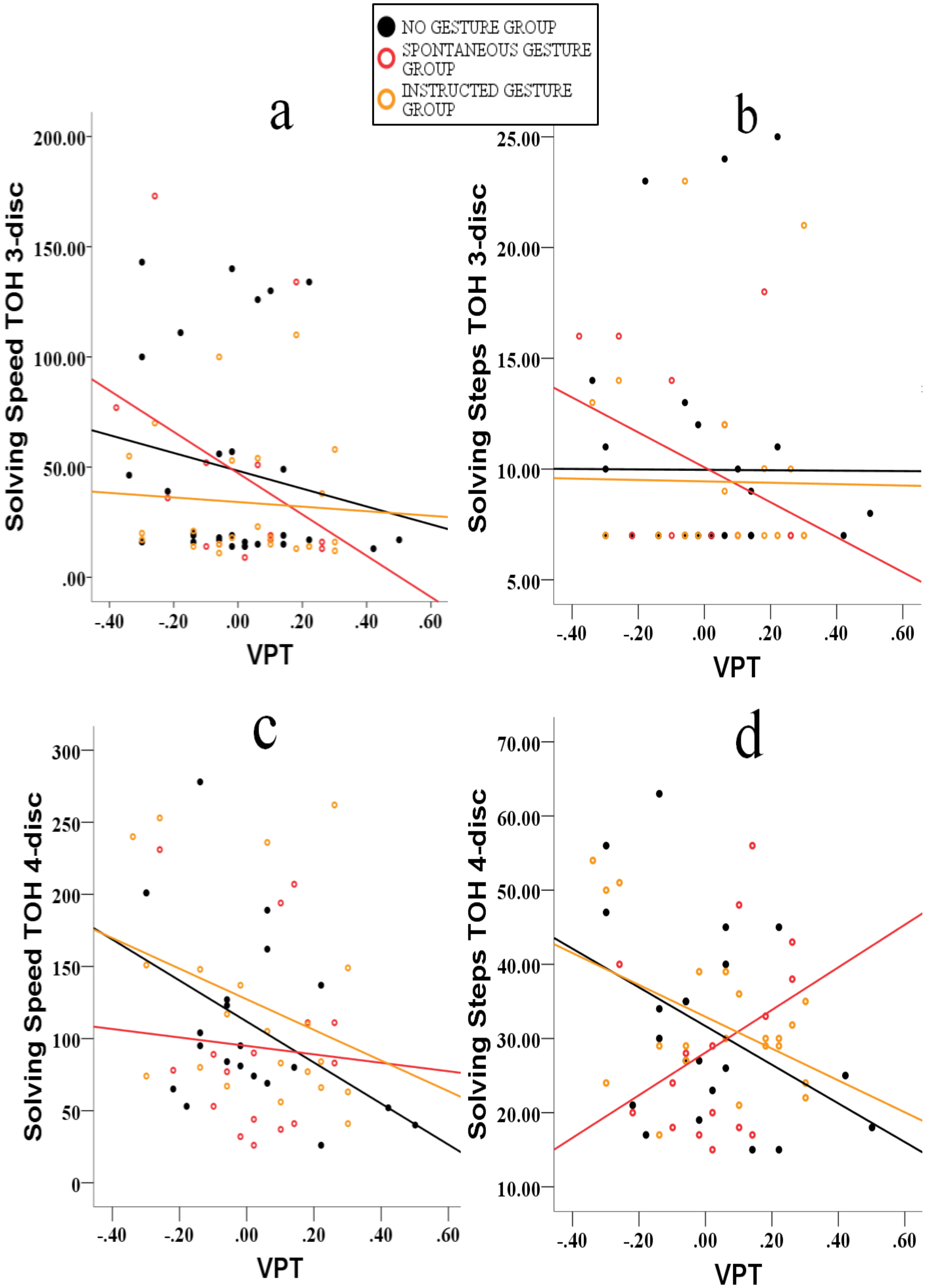


Figure 3a-d. Confirmatory analyses graphs. Note that the slopes are modelled on the observed data and differ from the estimated regression coefficients of the complete regression model (which includes age as a covariate).

Discussion

In the current study we aimed to replicate previous findings with adults (Pouw et al., under review) in young children. In adults, participants with lower visual working memory capacity who gestured either spontaneously or when instructed to do so during mental problem solving, performed better when subsequently solving the more difficult Tower of Hanoi problem than participants who did not gesture. In the current age sample with 8-12 year old children, we did not replicate these findings. That is, there were no effects of gesture (either spontaneous or instructed) on problem solving performance, either overall or as a function of visual working memory capacity or task difficulty.

There are several possible reasons why our hypothesized gesture effect was not replicated in the current sample. Firstly, in contrast to the study with adults, children's problem-solving performance was not correlated with performance on the visual working memory task. One possible explanation is that we used a more simplified version of the task, wherein children recreated the visual pattern by selecting locations through mouse-clicking, whereas the study with adults required verbally recalling the pattern by naming the letters that were assigned to the previous locations during the response phase. This difference in the task potentially recruits different cognitive processes. Given that we assume that gestures become effective when task-relevant resources are taxed, if we failed to gauge such task-relevant resources with the current task, then the current study was not able to test the cognitive load hypothesis. However, we doubt whether the difference of the simplified task with the task used with adults recruits different resources in such a dramatic way. Yet, problematically, this does beg the question why children's visual working memory resources were not correlated with task performance. One possible explanation is that children might not use a similar strategy as adults, and use more various strategies that do not recruit such visual imagery processes. Indeed it has been argued that children of the current age group are still developing the planning skills that are required to solve the Tower of Hanoi (Schiff & Vakil, 2015),

Relatedly, a general worry of the present results is that a younger age sample inevitably produces more noisy data, which might lower the detection of a possible gesture effect. One way that this manifested itself, is that some participants seemed more engaged with the task as than others (i.e., some children were not very enthused about performing the mental preparation task for [the full] 150 seconds). Even though we did not obtain any differences in self-reported motivation scores across conditions, given that we are unable to assess with certainty how engaged children are during the mental preparation phase, it might be that this natural variability prevents detection of a potential gesture effect. Yet, in any case the current sample does provide information about the ecological robustness of a potential co-thought gesture effect in children; which would not be very robust in a sample comparable to the present one (8-12 yrs old).

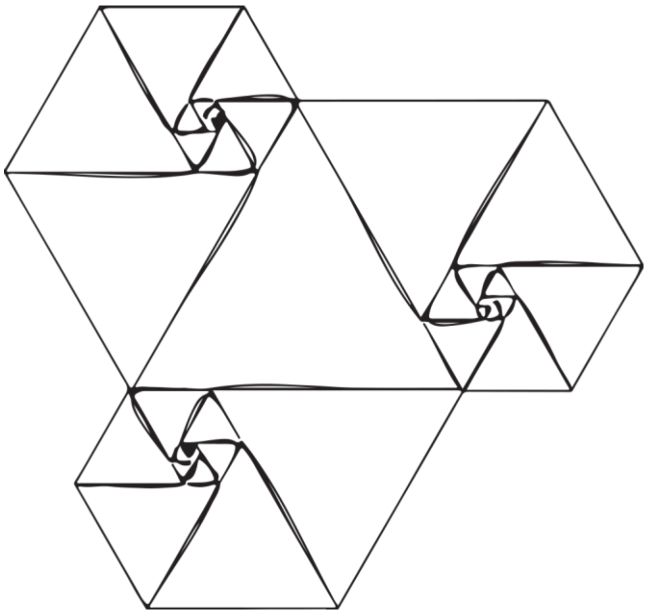
Although the results did not confirm our hypotheses, the current study provides novel evidence that children (in comparable ways as adults) spontaneously adopt co-thought gestures when being confronted with mentally simulating the problem space of the Tower of Hanoi. This is interesting, as it provides evidence that co-thought gesturing is already part of the cognitive toolkit in earlier development, and such gesturing persists throughout adulthood (Pouw et al., under review). Thus, at a minimum this study provides a productive paradigm that naturally solicits co-thought gestures from children, which will be useful for the further investigation of the cognitive role of co-thought gesticulation in younger age samples.

There is another aspect that needs to be emphasised to put the current results in an appropriate context. Firstly, in the current study we did not compare the effect of gesture with the inhibition of gesture, but rather with spontaneous non-gesturing. We reasoned, if we did obtain an effect of gesture it would allow us to conclude that the production rather than the inhibition of gestures affects cognitive processing (for a discussion of the theoretical importance of this difference, see Pouw et al., 2014; Cook, Yip, & Goldin-Meadow, 2010; Goldin-Meadow et al., 2001; Wagner, Nusbaum, & Goldin-Meadow, 2004). However, future studies could compare the role of inhibiting gestures as well, to see whether gesture effects on performance does arise. That is, given the fact that children did spontaneously gesture in the current study, it could be the case that actively inhibiting children to move their hands would result in a gesture effect on performance (albeit in a negative way).

The current study does provide more insight on the role of co-thought gestures in thinking and the potential boundary conditions concerning the beneficial effect on problem solving. Namely, children (8-12 yrs) spontaneously use gestures in similar ways (pointing gestures), and in similar amounts (Figure 2) as compared to adults, although do not necessarily positively affect problem solving as compared to non-gesturers. As such, it keeps the question that motivated this study in the first place very alive: what is the cognitive function of co-thought gesture? Why does the current task invoke spontaneous pointing gestures, while others evoke more iconic pantomimic gestures (e.g., Chu & Kita, 2008)? Moreover, the current study can inspire a more systematic study into the development of co-thought gestures in children, as it provides a paradigm in which these gestures are naturally adopted. This is needed, as current studies on the developmental emergence/ontogenetics of gestures largely ignore the phenomenon of co-thought gestures (e.g., Goldin-Meadow, 1998; Kendon, 2004; McNeill, 2008).

Chapter 8

Gesturing during mental problem solving reduces eye movements, especially for individuals with lower visual working memory capacity *



8

*This chapter has been published as Pouw, W. T. J. L., Mavilidi, M., Van Gog, T., & Paas, F. (2016). Gesturing during mental problem solving reduces eye movements, especially for individuals with lower visual working memory capacity. *Cognitive Processing, 17,* 269-277. doi: 10.1007/s10339-016-0757-6

Gesturing during mental problem solving reduces eye movements, especially for individuals with lower visual working memory capacity

Non-communicative hand gestures have been found to benefit problem-solving performance. These gestures seem to compensate for limited internal cognitive capacities, such as visual working memory capacity. Yet, it is not clear how gestures might perform this cognitive function. One hypothesis is that gesturing is a means to spatially index mental simulations, thereby reducing the need for visually projecting the mental simulation onto the visual presentation of the task. If that hypothesis is correct, less eye movements should be made when participants gesture during problem solving than when they do not gesture. We therefore used mobile eye tracking to investigate the effect of co-thought gesturing and visual working memory capacity on eye movements during mental solving of the Tower of Hanoi problem. Results revealed that gesturing indeed reduced the number of eye movements (lower saccade counts), especially for participants with a relatively lower visual working memory capacity. Subsequent problem solving performance was not affected by having (not) gestured during the mental solving phase. The current findings suggest that our understanding of gestures in problem solving could be improved by taking into account eye movements during gesturing.

Introduction

Gesturing can benefit problem solving, especially under conditions of high cognitive load (e.g., Chu & Kita, 2011; Marstaller & Burianová, 2013; for a review see Pouw, De Nooijer, Van Gog, Zwaan, & Paas, 2014). Yet, the exact mechanisms through which the cognitive system exploits manual activity are still not clear. Based on the literature discussed below, we hypothesized that gestures (pointing) allow for spatially indexing mental simulations in space, which come to stand in for eye movements that visually project mental simulations onto the presentation of the task in the external environment (Cappucio, Chu, & Kita, 2013; Cooperrider, Wakefield, & Goldin-Meadow, 2015; Pouw et al., 2014). To test that hypothesis, we investigated whether gesturing (pointing) vs. not gesturing during mental problem solving (Tower of Hanoi; hereon TOH) affected eye movements.

Gesturing during problem solving (or mentally problem solving) has been shown to benefit (subsequent) problem-solving performance, especially when cognitive load is high (for a review Pouw et al., 2014). That is, when the task is more complex (e.g., Chu & Kita, 2011; Delgado, Gómez, & Sarriá, 2011; Logan, Lowrie, & Diezmann, 2014) and/or when cognitive resources (such as working memory) are limited (e.g., Marstaller & Burianová, 2013). For example, participants who spontaneously used pointing gestures, or had been instructed to gesture while mentally solving the TOH for 150 s (in silence), subsequently performed better on solving the problem as compared to participants who did not gesture (Pouw, Eielts, Van Gog, Paas, & Zwaan, under review). However, gesturing was only beneficial for performance compared to not gesturing under conditions of higher cognitive load: for participants with lower visual working memory capacities, and only on more complex trials. In line with these results, there is evidence that gestures are indeed spontaneously employed to compensate for visual processing load: spontaneous gestures have been found to increase in rate when subjects are wearing glasses that project visually complex information compared to when simple information is projected (Smithson & Nicoladis, 2014). 8

The effect of cognitive load on gestures' effectiveness can be interpreted from an embodied and embedded cognition perspective (Cappucio et al., 2013; Clark, 2013; Pouw et al., 2014). According to this interpretation, gestures offer the cognitive system stable extra-neural tools for visuo-spatial thinking from which new or improved cognitive resources can emerge. That is, gestures embed, support and extend ongoing internal cognitive processing (e.g., working memory).

Yet, a major challenge for current research on the role of gesture in problem solving is to specify *how* gestures support cognitive processes (Cappuccio, Chu, & Kita, 2013; Cooperrider, et al., 2015; Pouw et al., 2014). One potential mechanism was proposed by Cappuccio and colleagues (2013). Focusing on the role of pointing gestures, they suggest that gesturing during problem solving provides a compensatory mechanism for visual processing: "pointing hence represents a stand-in for the corresponding series of acts of ocular redirection; the benefits received from monitoring these acts affect capabilities such keeping track of what has been counted, individuating objects, focusing on a particular object, anchoring number words to objects… … double-check, re-organize, concentrate, and parse in time/space the task..." (p. 141).

Indeed, there is evidence that eye movements ("ocular redirection") reflect and even support mental simulations during on-line problem solving. For example, Kirsh (2009) confronted participants with a tic-tac-toe game, in which they had to keep track of their own and the opponent's moves in working memory. It was found that participants (especially those with a low spatial ability) performed better on the most difficult task when they could play the game while looking at a tic-tac-toe matrix as opposed to an empty sheet of paper. The tic-tac-toe matrix allowed a way to "project" mentally simulated information on a presentation of the task in the environment. In similar vein, findings from eye tracking research on solving the TOH suggest that problem solvers actively explore possible moves visually when presented with a 2D presentation of the task, anticipating (or simulating) the placement of the disc from one peg to another with an eye-movement (e.g., Patsenko & Altmann, 2010). As argued by Spivey and Dale (2011), eye-tracking research in problem solving (e.g., Thomas & Lleras, 2007) suggests that eye movements not only reflect but also support ongoing problem solving by anchoring cognitive processes in the environment. This visual projection strategy, though, produces substantial cognitive load, because of the need to not only visually plan, but also visually monitor the "correctness" of each step of the mental simulation, mapped onto an external visual presentation that has not (yet) changed. Thus, this strategy might be especially difficult for those with lower visual working memory capacity.

Although not explicitly stated by Cappuccio and colleagues (2013), it can be argued that gestures are likely to be "monitored" through proprioception (i.e., the sense of the relative positions of body and limbs in space; see Pouw et al., 2014 for a discussion). Gesturing, we would suggest, provides an additional non-visual based spatial presentation that can anchor mental simulations. That gesture's function is (at least in part) proprioceptive, is in line with recent research that shows that gestures affect problem solving of the TOH even when gesturers cannot see their own hands (Cooperrider, et al., 2015). When gestures are proprioceptively monitored, it can be hypothesized that gestures can come to "stand-in" for eye movements as an anchor for mental simulations in the external environment, thereby reducing the number of eye movements being made. Furthermore, this effect should be stronger under conditions of higher visual working memory load, that is, when tasks are more complex or (when task complexity is equal) for those individuals who have lower visual working memory capacity.

We investigated this hypothesis in the present study. Participants performed two trials of the TOH of similar complexity: Each trial consisted of a 4-disc problem but with normal or inversed rules, wherein each solution path is exactly the same (see method for details). In one of the two trials participants were instructed to gesture (pointing in silence) during a 60 s mental solving phase that preceded actual problem solving, in the other trial participants did not gesture. If pointing gestures indeed allow for spatially indexing a mentally simulated move of a disc in space surrounding the body (peri-personal space), then the need to project information visually onto the 2d presentation of the task becomes functionally redundant, and a lower saccade count would be expected on the gesture trial than on the non-gesture trial. Moreover, we would predict that the function of gesturing is especially relevant (and therefore exploited) for those with lower WM capacity, as those with higher WM capacity may be able to easily project mental simulations using a visual strategy. If this prediction is correct it could provide a functional explanation to why gestures seem especially effective for those with a lower visual working memory capacity (e.g., Marstaller & Burianová, 2013; Pouw et al., under review).

Method

Participants and Design

This study was approved by the Human Research Committee of the University of Wollongong. A total of 20 adults participated in the present study (employees of the Early Start Institute Wollongong), who were unaware of the hypotheses of the study (M_{age} = 34.40, *SD* = 8.63, age range 24-50 years; 5 males).

A within-subjects experimental design was used, meaning that all participants performed two versions of the 4-disc TOH task. Depending on counterbalancing condition participants were instructed not to gesture or to use pointing gestures during the first or second mental solving phase. Whether they first solved the normal TOH and then the inverted TOH or vice versa, was also counterbalanced between subjects. Each physical solving phase was preceded by a mental solving phase of the task for 60 s.

Before the start of the experiment participants reported previous experience with the TOH (yes or no) and one participant reported that he had experience with solving the TOH in the past. We did not exclude this participant, because our within-subjects design should control for possible confounds of skill in relation to the manipulation. Note, however that excluding this participant resulted in the same pattern of findings reported in the result section.

Apparatus and Materials

Eye-tracking equipment

Eye movements were recorded with SMI eye-tracking glasses 2.0 connected via USB to a smart-phone from which the data could be uploaded afterwards. Data were analyzed with SMI BeGaze software (version 3.3). The sampling rate was set at 60 Hz and was bi-ocular. For each participant, before the start of the experiment, a 3-point triangular calibration was performed (distance between participants' eyes and points on the screen: point 1: 175 cm, point 2: 175 cm, point 3: 154 cm; distance between point 1 and point 2: 98 cm, distance between points 1 and 2 with point 3: 56 cm). To verify the accuracy of the calibration, subjects were asked to look at the same points again.

Video screen

All tasks were performed on the computer that projected onto a large LED TV screen, size 167 × 95 cm. The distance between the eyes and the screen was 165 cm.

Visual Patterns Test

The Visual Patterns Test (VPT; Della Sala, Gray, Baddeley, & Wilson, 1997) was a mouse-based task and served as a proxy for visual working memory capacity. We used an adapted version of the VPT (as adapted from and kindly provided by Chu, Meyer, Foulkes, & Kita, 2013). Participants were shown a matrix, in various patterns, wherein half of the cells (i.e., squares of 14 cm × 14 cm) were colored black. Each pattern was displayed for 3 s, after which all the squares turned white. Participants needed to recreate the pattern of black squares by selecting the squares in a non-specific order, which upon selecting would turn black. The VPT consisted of 25 trials, with blocks of 5 trials per difficulty level (from seven to 11 black squares). Before the start of the task participants were provided with 2 practice trials (3 and 4 black squares, respectively). If participants failed to recall all the black squares during a given trial, it was scored as an incorrect response. After five consecutive incorrect responses within one difficulty block of trials the experimenter stopped the task. Performance scores were the proportion of correct responses out of all trials.

Tower of Hanoi

The TOH was programmed in Adobe Flash and consisted of three evenly spaced pegs (distance between pegs: 41 cm, bases: 29.5 × 2.5 cm, peg: 2 × 3.4 cm) with four discs (disc 1: 29 × 4 cm, disc 2: 24 × 4, disc 3: 17 × 4 cm, disc 4: 12 × 4 cm). In the starting position, all discs were stacked on the outer left peg. In the normal rule TOH, the discs decreased in size (i.e., disc 1 to 4), and the inverted rule TOH increased in size (i.e., disc 4 to 1). Discs could be placed on the other pegs during the problem-solving process with the click-and-drag mouse function. The goal of the TOH is to transfer the discs from the left peg to the right peg in the same stacking order, subject to the following rules: 1) only one disc at a time can be moved to another peg, 2) a disc can only be moved if it is on the top of the stack, and 3) only smaller discs can be placed on top of bigger discs (normal TOH setup) or only bigger discs can be placed on top of smaller discs (inverted TOH setup).

8

Procedure

Prior to the experiment participants provided their written consent. Participants were tested individually with the two experimenters present in the room (but they could not see the experimenters during the tasks). They were first presented with the VPT. Participants were instructed on the nature of the task and performed two practice trials before the start of the VPT proper. The

VPT task took approximately 5 min to complete and there were no time restrictions for this task.

Subsequently, participants put on the eye-tracking glasses and the eye tracker was calibrated. After successful calibration, a practice TOH task with two discs was presented to participants and consistent with the counterbalance order, this was a normal TOH or inverted TOH practice task. The experimenter explained the rules of the task (with the third rule depending on assigned condition) and participants then solved the two-disc TOH trial as practice (for both normal and inverted TOH). After each instruction the experimenter verified whether subjects understood the instructions based on whether they solved the practice trial and participants were also asked to verbally repeat the rules.

After the practice trial, participants were informed that before actually solving a similar 4-disc TOH trial, they would be presented with the begin state of the 4-disc TOH trial (i.e., discs placed on the outer-most left peg) and that they should mentally plan the moves in silence for 60 s so they could solve the task as fast as possible directly afterwards. Participants were told that they should rehearse the solving moves repeatedly during this phase. Depending on the counterbalancing condition participants were instructed to think with their hands using pointing-gestures during this mental planning phase in a way that suited them (gesture condition). During this instruction the experimenter made several pointing gestures directed at the TOH as a cue how gestures could be performed. Participants were additionally instructed that they should not gesture directly in front of their face (this was done to ensure that field of vision was not, or only peripherally occluded by gesturing). In the no gesture condition participants were asked not to move their hands during the 60 s of mental solving. Directly after the mental solving phases, participants solved the respective 4-disc TOH.

This cycle (practice task, mental solving, actual solving) was repeated twice. Participants either received the normal task first and the inverted second or vice versa (i.e. counterbalanced between participants), and were instructed either to gesture on the first task and not on the second or vice versa (i.e., counterbalanced between participants). Once participants correctly solved the first problem, they automatically proceeded to the next cycle. When participants were unable to solve the task, they automatically proceeded to the next cycle after 5 minutes. Participants were recorded during the TOH (mental) solving phases with a video camera for the purpose of counting their gestures after the experiment. Finally, participants were debriefed and thanked for their participation.

Scoring and data analysis

Gesture

Participants' video data per task were coded for gesture frequency (for an example see Figure 1). Due to camera malfunction we could not count gestures of two participants. Gestures were defined as any hand movement of one or both hands from one still point to the next, indicating the travel of a disc from one peg to another (see Garber & Goldin-Meadow, 2002). All participants used index-pointing gestures. The first two authors independently counted the gestures, and interrater reliability was high, Pearson's $r = .89$, $p < .001$.

Eye movement data

The number of saccades within the 60 s mental solving phase per task were generated using default settings of the eye-tracking software SMI BeGaze software (version 3.3) for the exact period of 60 seconds.

Performance

For the two problem-solving trials we obtained solving speed and number of solving steps (number of mistakes were not counted by the program). Lower number of solving steps and faster solving speeds reflect a higher performance. For each TOH problem-solving trial the minimal amount steps necessary to solve the task were fifteen steps. As the given period of solving a trial was set at 300 s, participants who did not solve the task in 300 s were not scored on performance.

Figure 1. Example of gesturing during the mental solving phase (1 s per frame).To show where participants look at during gesturing, the last frame is an example of the static Tower of Hanoi presented for 60s during mental problem solving (inverted rules condition).

Results

Three participants had to be excluded due to technical issues with the eye tracking glasses. This resulted in a total sample of 17 participants, (M_{age} = 35.24, *SD* = 9.10, age range 24-50 years; 4 males), wherein counterbalancing resulted in *N* = 4 for gesture-normal setup, *N* = 5 for no gesture-normal setup, *N* = 4 for gesture-inverted setup, and *N* = 4 for no gesture-inverted setup during the first TOH trial (counterbalanced for the second TOH trial). Where TOH performance effects are concerned, an additional 2 participants were not included in the sample as they were not able to solve one of the two TOH trials within 300 s.

Table 1 presents the means, standard deviations and correlations of VPT score, solving steps and solving speed during the solving phase, as well as saccade counts during the mental solving phase. Note, that higher VPT scores were associated with fewer fixations and saccades overall (*p*s < .034). Interestingly, however, when partialling out the correlations per condition (gesture vs. no gesture) we found that this overall significant correlation was primarily carried by the no gesture condition (VPT and saccade count: $r = -.541$, $p = .025$). In the gesture condition there was no significant correlation of VPT with saccade count ($r = -.022$, $p = .933$). Note however, that these correlations did not significantly differ, $p = 0.123$ (see Lee, & Preacher, 2013). These results suggest that visual working memory capacity was more predictive for saccade count in the no gesture condition.

	M (SD)	1.	2.	3.
1. VPT score	.76 (.13)			
2. Solving speed	89.86 (38.97)	-.054		
3. Solving steps	29.37 (12.02)	.195	.828**	
4. Saccade count	138.08 (25.41)	-.517*	.131	.024

Table 1. Overall means and standard deviations, and correlations between VPT score, solving time TOH, solving steps TOH, and saccade count. *Note.* * $p < .05$, ** $p < .01$

8

The mean pointing gesture frequency (which could only be obtained for 15 participants because two had to be excluded due to camera malfunction) during the mental solving phase was 31.87 (SD = 13.11; minimum gesture frequency = 14, maximum = 57). We found no significant correlations between gesture frequency and VPT score, $r = .13$, $p = .638$. Also, the gesture frequency on the task was not significantly correlated with solving speed on the respective trial (which was preceded by gesturing during the mental solving phase), $r = -.33$, $p = .224$, nor was this the case for solving steps, $r = -.28$, $p = .320$. We also checked whether gesture frequency was associated with saccade and fixation count but no significant associations were found, saccade count $r = -.02$, $p = .953$, fixation count $r = -.07$, $p = .802$.

Eye Movements

To test our main hypothesis whether gesturing leads to lower saccade counts during the mental solving phase as compared to not gesturing, and whether this effect was moderated by visual working memory capacity, we performed two separate mixed effects Analyses of Covariance (ANCOVAs) on the number of saccades. For each DV, we examined the within-subjects effect of gesturing versus not gesturing, with VPT score as a covariate. We first checked for between-subjects effects of counterbalancing order of gesture first vs. no gesture first, as well as the order of TOH type (normal vs. inverted) by adding these as between-subjects factors, which there were not: gesture counterbalance order, $F(1, 12) = .22$, $p = .644$, TOH type counterbalance order, $F(1, 12) = .03$, $p = .865$, and interaction, $F(1, 12) = 1.63$, $p = .226$.

The results did reveal a significant relationship on the number of saccades when participants gestured compared to when they did not gesture, $F(1, 12) = 8.34$, $p = .014$, $\eta_p^2 = .41$. Overall, fewer saccades were observed when participants gestured (estimated means saccade count = 124.06, SD = 26.52, 95%CI = 108.36 - 140.00) than when they did not gesture (estimated means saccade count = 132.88, SD = 39.24, 95%CI = 115.34 - 151.79) when controlling for the covariate VPT. Moreover, there was a significant interaction of gesture condition and the VPT regarding the number of saccades, $F(1, 12) = 7.32$, $p = .019$, $\eta_p^2 = .38$. In Figure 2 we have plotted the effect of VPT score on the observed differences of saccade count across gesture condition. As Figure 2 shows, the reduction in saccades when gesturing compared to not gesturing was stronger for participants who scored lower on the VPT.[15, 16]

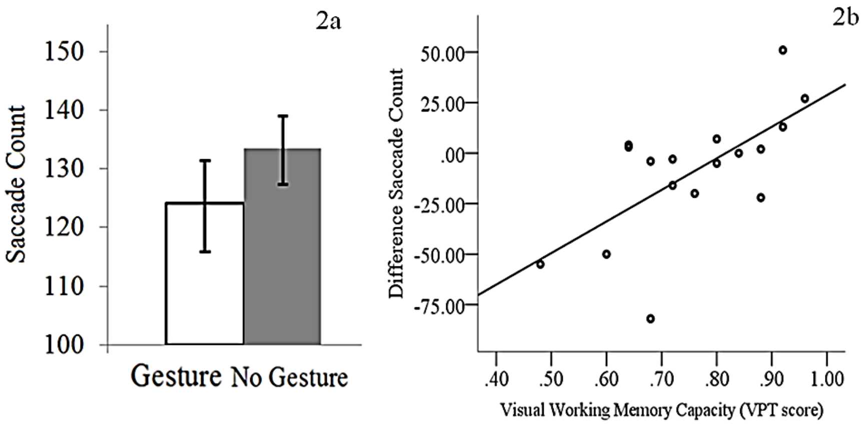


Figure 2a and 2b. On the left the estimated marginal means and standard errors of the ANCOVA for number of saccades during the 60 s are presented. On the right the difference scores are presented in relation to visual working memory

[15] Note that since saccade and fixation frequency closely covary, very similar results are obtained when taking into account fixation frequency. A similar repeated-measures Analysis of Covariance (ANCOVA) was performed with fixation count as the dependent variable. There were no significant between subject-effects of gesture counterbalance order, $F(1, 12) = .05$, $p = .831$, or TOH counterbalance order, $F(1, 12) = .21$, $p = .653$, nor did the interaction of counterbalancing conditions have an effect, $F(1, 12) = 2.48$, $p = .141$. Results revealed significantly lower fixation counts when participants gestured (estimated means = 143.70, *SE* = 5.38, 95%CI = 131.97 – 155.44) compared to when they did not gesture (estimated means = 153.298, *SE* = 6.30, 95%CI = 139.56 – 167.02), $F(1, 12) = 8.29$, $p = .014$, $Eta^2_p = .41$. Also, there was a significant interaction between the number of fixations and gesture and the VPT, $F(1, 12) = 7.22$, $p = .020$, $Eta^2_p = .38$.

[16] As was to be expected given the fixed time available for mental problem solving and the lower fixation count, average fixation duration when gesturing was somewhat higher than when not gesturing, but a similar repeated-measure Analysis of Covariance (ANCOVA) on average fixation duration showed that this difference was not significant. No between-subject effects of gesture counterbalance order were found, $F(1, 12) = 0.831$, $p = .380$, TOH type, $F(1, 12) = .09$, $p = .776$, and its interaction, $F(1, 12) = 0.66$, $p = .433$. Furthermore, average fixation duration was not significantly affected by gesture (estimated means in ms = 300.25, *SE* = 15.16, 95%CI = 267.20–333.25) versus no gesture (estimated means in ms = 284.80, *SE* = 14.96, 95%CI = 252.23 – 317.42), $F(1, 12) = 0.25$, $p = .625$, nor was there an interaction effect of gesture and VPT, $F(1, 12) = 0.14$, $p = .716$.

8

capacity**.** *Note.* On the right plot, a negative difference means that lower saccade counts were observed when participants gestured versus did not gesture during the mental solving phase. The trend shows, that participants with a lower visual working memory capacity were more extremely affected by gesturing, such that a lower saccade count was observed when participants gestured as compared to when they did not gesture.

TOH Performance

For exploratory purposes we assessed whether gesture condition and VPT affected performance of the TOH, using two repeated-measures ANCOVAs with solving time or solving steps as the dependent variable, gesture versus no gesture during the mental solving phase as within-subject factor, counterbalancing variables gesture order and TOH type as between-subject variables, and the scores on the VPT task as the covariate.

Solving time

No effects of between-subject (i.e., counterbalance) factors gesture order, $F(1, 10) = 1.86$, $p = .202$, TOH order, $F(1, 10) = 0.16$, $p = .699$, or their interaction, $F(1, 10) = 1.97$, $p = .191$, were found on solving time. Furthermore, solving time was not affected by whether participants gestured ($M = 94.24$, $SD = 55.80$, 95%CI = 65.48 - 124.48) or did not gesture ($M = 85.48$, $SD = 42.03$, 95%CI = 64.55 - 109.96) during the mental solving-phase, $F(1, 10) = 1.25$, $p = .289$. Also, VPT was not significantly co-varying with observed differences, $F(1, 10) = 1.37$, $p = .346$.

Solving steps

No effects of between-subject (i.e., counterbalance) factors gesture order, $F(1, 10) = 0.44$, $p = .523$, TOH order, TOH, $F(1, 10) = 0.99$, $p = .341$, or their interaction, $F(1, 10) = 0.21$, $p = .655$, were found on the number steps taken to solve the problem. Additionally, solving steps was not affected by whether participants gestured ($M = 30.63$, $SD = 14.68$, 95%CI = 22.27 - 39.01) or did not gesture ($M = 28.13$, $SD = 16.73$, 95%CI = 17.97 - 38.28) during the mental solving-phase, $F(1, 10) = 0.54$, $p = .479$. Also, VPT was not significantly co-varying with observed differences, $F(1, 11) = 0.65$, $p = .437$.

Discussion

Prior research has shown that gesturing may compensate for high working memory load (e.g., Marstaller & Burianová, 2013; Pouw et al., under review). However, it is not yet clear *how* gestures perform this cognitive function. The present study investigated the hypothesis that pointing gestures, by exploiting space, reduce the need for exploiting the visual presentation of the task in the external environment as a way to anchor mental simulations. Consequently, we expected less eye movements to be made when participants gestured during mental problem solving of the Tower of Hanoi (TOH) than when they did not gesture, because gestures can come to "stand-in" for eye movements as an anchor for mental simulations in the external environment. That is, through pointing, gesturers can spatially index mental simulations of moving the discs from one peg to another in peri-personal space, rather than moving the eyes to project imagined disc-movements onto the visual presentation of the task. Given that gestures can compensate for high cognitive load, we expected this effect to be stronger for those individuals who have lower visual working memory capacity (as problem solving places higher demands on their resources).

In line with this hypothesis, our results showed that gesturing lowered saccade counts during mental problem solving, and more strongly so for those with a lower visual working memory capacity. As such, this study makes a novel contribution towards explaining (one of) the mechanism(s) through which gestures may support (mental) problem solving. Whereas eye movements allow for projecting mental simulations in the external environment, gestures do this in exploiting peri-personal space through proprioceptive monitoring and peripheral visual control, thereby offloading visual working memory processes.

An important question is whether we can exclude that the effect of gesture on eye-movements is an epiphenomenon, i.e., functionally irrelevant for mental problem solving? We think that gestures' effect on eye-movements are not likely to be epiphenomenal as there are a host of findings which show that eye-movement patterns are crucial for thinking through the solution space of a problem (Spivey & Dale, 2011) and to visual imagery in general (e.g., Brandt & Stark, 1997; Johansson, Holsanova, & Holmqvist, 2006; Laeng, & Toedorescu, 2002). However, we do not (and cannot) claim (based on the present data) that reduction of saccade count is necessarily beneficial for problem solving as opposed to a more visually dominant strategy. However, given that eye-movements are highly likely to be functionally relevant for mental simulations, and given the present findings that especially those gesturers with a lower visual working memory capacity considerably alter their gaze patterns without significant loss in performance, it is likely that there is some trade-off mechanism present.

But what is the exact nature of this trade-off mechanism? Although this question cannot be definitively answered based on our data, the present study does suggest that the change from a visually dominant strategy to a strategy that exploits sensory consequences of gesture (especially proprioception) may offer a preliminary explanation. Recall that a visually dominant strategy involves moving the eyes in a way that corresponds with mentally moving the discs from one peg to another. This allows a way to anchor mental transformation on a visual presentation of the task (see Figure 1, last frame). This strategy thus involves mental projection onto the external environment, where the external environment offers an anchor or reference that is meaningful to the task (e.g., Kirsh, 2009). Pointing gestures can, we think, fulfil the same function as eye-movements. However, pointing fulfils this function with different and less visually dominant resources. Namely, through pointing peripersonal space is sequentially filled by positions of the hand that are, by physical human nature, monitored through proprioception and/or (peripheral) visual control (e.g., Bremner & Cowie, 2013). The locations that the hand takes in in space during pointing can come to correspond with the mental transformation being made by the gesturer. That is, mentally simulating the move of a disc, corresponds to pointing from one location to another. The reason why we think pointing is not a visually dominant strategy, is that if participants pointing gestures were actively visually tracking their pointing movements then we would not have observed a difference in saccades between gesture vs. no condition, as mental transformation in both cases are visually tracked (albeit in the gesture condition via an external loop). This was not the case. Furthermore, informal inspection of the videos reveals that participants were indeed not looking directly at their hands during gesturing. This leads us to our interpretation that gestures must provide some additional resource for spatially indexing mental transformations. We thus think that next to peripheral vision, proprioception can offer a natural way to monitor the hand in space as to spatially index mental transformations. Finally, although we cannot definitively establish that gestures are indeed propriocepively dominant in this case, it does serve as an additional explanation of why those with a lower visual working memory capacity (a proxy for visual mental imagery ability) are especially likely to reduce their eye-movements. Namely, those problem solvers that are prone to have difficulty projecting/simulating visual transformations on the environment, can reap the benefits of spatially indexing mental transformation in a non-visually dominant way through pointing (using the proprioceptive sense of the hand in space). Findings that gesturing is especially potent for those with a lower working memory capacity (Marstaller & Burianová, 2013; Pouw et al., under review), and is beneficial even when participants cannot see their own hands (Cooperrider et al., 2015), concur with this idea that switching to a non-visually dominant strategy is possible and perhaps potent for some but not all problem solvers.

Another question that could be raised is whether present results exclude a strict motor-based interpretation of gesture, wherein gestures effect should be attributed to re-use (internal simulations) of motor experience (Hostetter & Alibali, 2008). Namely, a strict motor-based interpretation may hold that the *motor-intention to produce a pointing gesture*, rather than the *actual bodily gesture and its sensory consequences*, activates/supports internal motor-simulations which in some way affects gaze-behavior as observed in the present study. This is in contrast to the embedded/extended approach which assumes that any explanation of a cognitive function of gesture must always lay (at least in part) in the sensory consequences of gesturing that are used in some cognitively potent way (Pouw et al., 2014). However, the present study was not designed to differentiate between these interpretations. Future research could focus on distinguishing a strict motor-based interpretation from an interpretation that emphasizes sensory consequences of gesture. This can be done by manipulating gesture intention (as to trigger motor-simulations) versus actual gesture production. If the production of gestures plays no functional role in the present effect, then the intention to gesture should produce the same effect on eye-movements (without loss in problem solving performance). Finally note that the embedded/extended and motor-based approach can also be complementary. Under such a hybrid view, gestures arise out of motor-simulations and have sensory consequences which further affect ongoing simulation-based cognitive processes.

Our study has limitations. First, it should be stressed that the current study is small in scale, and as such definitive conclusions on the precise role of pointing on problem-solving processes should not be drawn from the present data. Especially, the present lack of an effect of gesture on problem-solving performance should be treated with caution as similar studies that did find a beneficial effect investigated this with a larger sample (e.g., Chu & Kita, 2011; Garber & Goldin-Meadow, 2002; Pouw et al., under review). That is, in contrast to our expectation, we did not find beneficial effects of gesturing during the mental problem-solving phase on TOH performance (TOH solving speed and solving steps). This is in contrast to prior findings (Pouw et al., under review), but important differences between the current and prior study lie in the design. First, participants in the prior study had more mental solving time before they physically performed the task: 150 s vs. 60 s in the present study. Second, whereas gesturing was a between-subjects factor in the prior study, it was a within-subjects factor in the present study. As such, even though it is unlikely given that the rule was inversed between tasks and the analysis of order effects revealed no significant differences, we cannot rule out entirely that there were carry-over effects that may have eliminated potential beneficial effects of gesturing on performance (especially since the number of participants per group in the order analyses were based was very small). For example, Chu and Kita (2011) have found that the beneficial effects of gesture can carry over to a subsequent task (similar in nature) when gesturing is prohibited. Additionally, it could be the case that pointing gestures are less beneficial for problem solving performance as compared to co-speech iconic gestures that have been found to co-occur with *verbal* explanations of solving the TOH (e.g., Cook & Tatenhaus, 2009; Cooperrider et al., 2015), wherein participants gesture as-if grasping actual discs. Future research should further investigate whether iconic gestures during actual problem solving may have different effects than pointing-gestures. For example, this can be done by letting participants verbally explain the solution of the TOH (e.g., Cooperrider et al., 2015). Yet there are several reasons why in the present case iconic gestures might not be particularly effective. Firstly, in a previous study (Pouw et al., under review) we have found that pointing gestures, but not iconic gestures, are spontaneously produced during mentally solving a physical Tower of Hanoi task without speech. This suggests that iconic-gestures may be co-dependent on speech production, and not naturally employed during mental problem solving without the additional constraint to verbalize one's thoughts. Furthermore, in this previous study pointing gestures were found to benefit performance on subsequent solving of the TOH when cognitive load is high. Finally, the reason why iconic gestures are held to affect mental problem solving of the TOH is that they offer a correspondence with the actions to be performed on the actual task (Cook & Tatenhaus, 2009; Cooperrider et al., 2015). Yet, in the present case, manipulation of the task was mouse-based, which does not correspond with a grasping action. In sum, although iconic gestures may offer unique or better cognitive support for problem-solving, in the present non-verbal mouse-based task we doubt whether iconic gestures are more potent than pointing gestures.

A second limitation of the present study is that it relied on eye-movement frequency counts, and therefore does not yet illuminate the precise dynamics of pointing and gaze behaviour (i.e., when and how participants use gestures during [mental] problem solving and how this affects their eye movements). The benefit of our mobile eye-tracking device was that it allowed for maintaining natural degrees of freedom in hand movement, which is more difficult to obtain (at present) with remote eye tracking devices. Nevertheless, the higher temporal and spatial resolution that can be obtained with remote eye-tracking devices would allow us to address in more detail how eye movements are affected by gestures in future research.

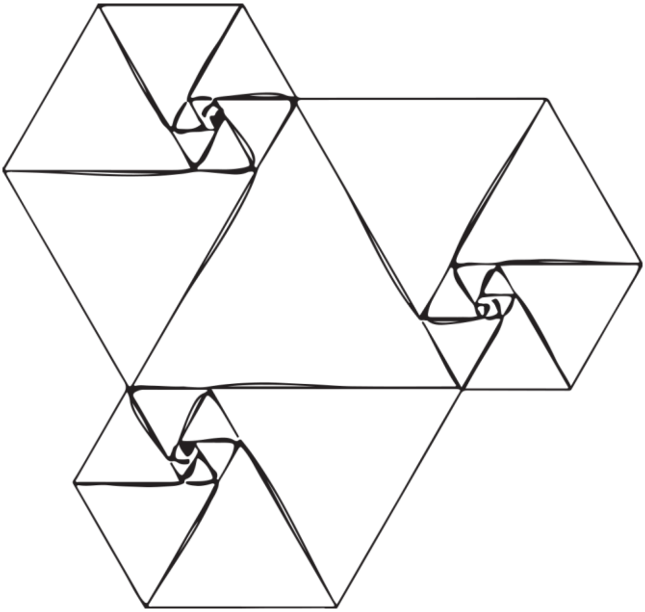
Despite these limitations, this study made a first step towards explaining (one of) the mechanism(s) through which gestures may support (mental) problem solving. Our findings suggest that gesturing may provide a unique embodied resource, exploiting peri-personal space, which may come to stand in for visually dominant strategies when these prove to be insufficient for meeting the cognitive demands imposed by the task. Taking gaze behavior into account in future research, may enhance our understanding of the role that non-communicative pointing gestures play in problem solving processes, for individuals differing in cognitive dispositions.

PART III

Moving Forward

Chapter 9

Gesture as Predictive Action *



*This chapter has been accepted for publication as Pouw, W. T. J. L., & Hostetter, A. (in press). Gesture as predictive action. *Retti, Sapperi, Linguaggi: Italian Journal of Cogntive Sciences*.

Gesture as Predictive Action

Two broad approaches have dominated the literature on the production of speech-accompanying gestures. On the one hand, there are approaches that aim to explain the origin of gestures by specifying the mental processes that give rise to them. On the other, there are approaches that aim to explain the cognitive function that gestures have for the gesturer or the listener. In the present paper we aim to reconcile both approaches in one single perspective that is informed by a recent sea change in cognitive science, namely, Predictive Processing Perspectives (PPP; Clark, 2013b, 2015). We start with the idea put forth by the Gesture as Simulated Action (GSA) framework (Hostetter & Alibali, 2008). Under this view, the mental processes that give rise to gesture are re-enactments of sensori-motor experiences (i.e., simulated actions). We show that such anticipatory sensori-motor states and the constraints put forth by the GSA framework can be understood as top-down kinesthetic predictions that function in a broader predictive machinery as proposed by PPP. By establishing this alignment, we aim to show how gestures come to fulfill a genuine cognitive function above and beyond the mental processes that give rise to gesture.

Introduction

When speakers talk, they often move their hands and arms in a way that mirrors or complements the semantic content of what they are saying. These movements, hereafter referred to simply as gestures, are, in some sense, the epitome of "embodiment" because they are movements of the *body* that are produced in the interest of communication. Yet, to say that gestures are embodied just because they make use of the body is unsatisfactory from the standpoint of Cognitive Science, which uses the term *Embodied Cognition* to mean that cognitive processes make use of perceptual and motor systems, even in situations where such systems would seem to be irrelevant (e.g., Wilson, 2002). To say that gestures are truly embodied then, requires the specification of how, and in virtue of which unique properties (e.g., visual, proprioceptive stimulation), gestures affect cognition (Pouw, de Nooijer, van Gog, Zwaan, & Paas, 2014). That is, we must move beyond descriptive accounts of what gestures are and towards understanding why they are produced and how they come to have facilitative effects on cognition.

So, why are gestures produced? This question has received increasing attention over the past two decades. While space does not allow a detailed description of the many possibilities, a review of the literature reveals two broad types of answers to this question. First, there are answers that are about *origin*. That is, are the processes or representations that underlie gesture unique to gesture or are they similar to those that are involved in speaking or action generation more generally? Second, there are answers that are about *cognitive function*. Once a gesture is produced, what effect does it have for the speaker and for the listener, and how is this effect brought about?

In this paper, we consider both the origin and function of gesture in a single account. We first describe one theory about the origin of speech-accompanying gestures, namely the Gesture as Simulated Action (GSA) framework (Hostetter & Alibali, 2008). Under this view, gestures arise from simulated perceptual and action states that are created as a speaker talks about a present or imagined situation. Inspired by contemporaneous ideas about the embodied nature of language comprehension (Barsalou, 2008; Glenberg & Kaschak, 2002), the goal of the GSA framework was to explain how gestures might originate in an embodied cognitive system that is engaged during speech production. However, in the years since the GSA framework was published, a number of theories have taken hold in Cognitive Science that we believe are compatible with, and nicely complement, ideas presented in the GSA framework, namely, Predictive Processing Perspectives.

Predictive Processing Perspectives (hereon PPP; e.g., Clark, 2013b, 2015a, b, 2015; Den Ouden, Kok, & De Lange, 2012; Friston, 2010; Glenberg & Gallese, 2012; Hohwy, 2013; Lupyan & Clark, 2015), broadly characterized, postulate that the central work of cognitive systems is to engage in predictions. That is, predictions about sensori-motor consequences that emerge during interaction with the environment, and are updated and acted upon in a way that minimizes surprisal or *prediction error* (i.e., the residual discrepancy between what is predicted and what is encountered). We think PPP are promising for furthering our understanding of how gestures emerge from embodied simulations, while also securing a central role for bodily action in these processes. In this paper, we aim to explore how gestures may be thought about as one instantiation of predictive processing; that is, gestures have functions for speakers and thinkers that can be broadly construed as minimizing prediction error, and these functions can be captured by one broad underlying mechanism giving rise to gestures, namely the activation of simulations (i.e., predictions) in the motor and perceptual system. Thus, by considering gestures as a case of predictive processing, we aim to move beyond considering the origin of gesture and its function as separate explanatory quests, and provide a way to understand how origin and function of gesture are intricately related.

Outline

Next, (section 2) we provide an overview of the GSA framework's key tenets and the evidence to date. In section 3, we address the cognitive function of gesture. Finally, we introduce PPP, as a means of understanding how gestures not only arise from sensori-motor predictions (i.e., simulations) in the cognitive system (section 4), but also support these predictive processes during on-going cognitive activity (section 5).

The Gesture as Simulated Action Framework

The Gesture as Simulated Action (GSA) framework (Hostetter & Alibali, 2008) considers gesture production to be the outgrowth of a cognitive system that is actively engaged in simulating motor and perceptual states. Simulations are neural enactments or re-enactments of interactions with the world; when a speaker engages in simulation, the same motor and perceptual areas of the brain are recruited that would be involved in actually performing the action or viewing the scene. This neural activity in the motor and action systems of the brain has the potential to be expressed alongside speech as gesture. The GSA framework proposes three determinants of whether a simulation is actually expressed alongside speech in any particular instance.

First, the production of a gesture depends on how strongly the simulation evokes thoughts of action. Simulations that are closely tied to action are more likely to engage the motor system strongly enough to result in a movement being produced (e.g., gesture) than simulations with weaker ties to action. For example, a speaker who has experience actually making a pattern he is describing is more likely to gesture about the pattern than he is about a pattern he has only viewed (Hostetter & Alibali, 2010). Action can also be evoked in a simulation of a perceptual scene that was not directly acted on. For example, if speakers can easily imagine interacting with what they are describing, they are particularly likely to gesture about it. Chu and Kita (2015) found that speakers gestured less about a mug that had spikes along its handle than they did about a mug with no spikes that more readily afforded grasping. Masson-Carro, Goudbeek, and Krahmer (2015) further show that the affordances of objects directly predict whether the objects are gestured about. Simulating perceptual scenes may also evoke action if the speaker imagines the scene or its objects in motion. For example, speakers frequently gesture when engaged in mental rotation exercises (e.g., Chu & Kita, 2008, 2011) and gesture more when talking about the process of rotating than when describing the end state of the rotation (e.g., Hostetter, Alibali, & Bartholomew, 2011). Finally, even thinking about a static perceptual experience may engage the motor system because of the tight coupling between perception and action. When we perceive an object, we automatically activate processes about how we would use, grasp, or interact with the object (e.g., Tucker & Ellis, 1998). Under the GSA framework, such activation can be expressed as gestures that depict how to interact with the object, or outline the object's shape.

Second, the production of a gesture depends not only on the absolute strength of action activation involved in the simulation, but also whether this activation is strong enough to pass the speaker's current *gesture threshold*. The gesture threshold is conceptualized as the speaker's current resistance to producing a gesture, but can change from moment to moment during speaking. For example, in situations where speakers think a gesture might benefit their listener, they may lower their threshold and gesture more (e.g., Alibali, Heath, & Myers, 2001). Similarly, if speakers consider the information they are conveying to be particularly important to their listener, they gesture more (e.g., Kelly, Byrne, & Holler, 2011), perhaps as the result of maintaining a lower threshold that even weaker action simulations can surpass. Moreover, speakers may adjust their threshold (either consciously or unconsciously) based on the cognitive demands of the speaking situation. Because gestures are known to have a number of beneficial effects (e.g., Goldin-Meadow & Alibali, 2015), a speaker may find it advantageous to lower her threshold to allow even a weak action activation to be expressed as gesture in certain situations. Conversely, even in situations where there is no clear reason to lower one's threshold, simulations that evoke strong activation of action may result in gesturing regardless because the simulation is strong enough to pass even a heightened threshold (see Hostetter, 2014 for some evidence on this point).

Finally, the GSA framework contends that the occurrence of gesture is particularly likely in situations where the articulatory motor system is already activated in the interest of speaking. Because the speaker must engage his or her motor system for speaking, it is difficult to simultaneously inhibit the manual motor system from also expressing the action activation that occurs during simulation. This is corroborated by findings that show that hand and mouth actions are linked from infancy (Iverson & Thelen, 2000) and heavily constrain one-another throughout further adulthood (Gentilucci, Benuzzi, Gangitano, & Grimaldi, 2011). However, while the GSA framework contends that gestures are more likely to occur with speech than in its absence, the framework by no means precludes the occurrence of gestures without speech. Indeed, since the publication of the GSA framework, a number of reports have been published about gestures that occur in the absence of speech (e.g., Chu & Kita, 2011; Delgado, Gómez, & Sarría, 2011). Such co-thought gestures seem to share many characteristics with co-speech gestures (Chu & Kita, 2015). This evidence supports for the idea that gestures are not dependent on language; while they frequently occur with language, the processes that give rise to gesture are more generally rooted in the sensorimotor, rather than linguistic, system.

The cognitive function of gesture

The GSA framework was developed to account for how gestures arise from an embodied cognitive system (Hostetter & Alibali, 2008, p. 495). While not reducing gestures to an epiphenomenon, the issue of how gestures *function* in such a system was left open to further speculation. Consequently, the GSA framework is flexible regarding the possible functions of gesture. Indeed, once a gesture is produced, the GSA framework allows that the movement may have any number of cognitive effects. However, in order to offer a truly embodied account of gesture that considers both their origin and their function, a more detailed specification of how gestures perform their cognitive functions is needed.

What does an embodied account of gesture function entail? Pouw and colleagues (2014; see also Pouw, Van Gog, Zwaan, & Paas, in press) argue that to truly explain the cognitive function of *gesture*, a theory must be able to explicate how this bodily act affects the cognitive system above and beyond neural processes that precede gesturing. That is, it must become clear how the act of gesturing directly affects cognition, which is not accomplished when positing some neural process that generates the gesture as well as its cognitive effect. For example, consider a learner who is attempting to memorize the steps needed to complete a route. The learner may mentally visualize the steps required, and this visualization may lead the learner to gesture about each step and may also lead to improved memory for the steps. However, in order to consider gesture as a causal agent that led to improved memory, it must become clear what additional benefit gesturing brings above and beyond the mental visualization that gives rise to the gesture in the first place.

The idea that the act of gesture might add something to the cognitive toolkit is not new. As the philosopher Andy Clark (2013b) has recently described:

> In gesture, as when we write or talk, we materialize our own thoughts. We bring something concrete into being, and that thing (in this case, the arm motion) can systematically affect our own ongoing thinking and reasoning… … [as such] gesture and overt and covert speech emerge as interacting parts of a distributed cognitive engine, participating in cognitively potent self-stimulating loops whose activity is as much an aspect of our thinking as its result. (Clark, 2013b, p. 263).

In this way of thinking, cognition is not completely brain-bound; rather the physical activity of gesture - in virtue of its co-constitutive role in ongoing cognition - is itself a genuine form of cognition (cf. Clark & Chalmers, 1998). Similarly, McNeill (2005) has argued that gesture and speech exist together in a dialectic, with each influencing and affecting the other. In sum, gestures are not just the result of cognition, they are a critical determinant of cognition (e.g., Goldin-Meadow & Beilock, 2010).

Indeed, there is much research suggesting that gestures affect cognition in a variety of ways (see Goldin-Meadow & Alibali, 2015 for a recent review). For example, speakers who gesture have better memory for what they gesture about than speakers who do not gesture (Cook, Yip, & Goldin-Meadow, 2010). In addition to strengthening the representation being described, gestures also appear to reduce general working memory demands, such that there are more cognitive resources available to devote to a secondary task when speakers gesture than when they do not (Goldin-Meadow, Nusbaum, Kelly, & Wagner, 2001). Gestures appear to affect how speakers solve spatial problems, by influencing the strategy choice (e.g., Alibali, Spencer, Knox, & Kita, 2011) or by focusing attention on perceptual elements of the problem (e.g., Beilock & Goldin-Meadow, 2010). Given these effects on cognitive processing, it is perhaps not surprising that gestures also help speakers communicate, particularly about concepts that are highly spatial or motoric (e.g., Hostetter, 2011). There is some evidence that gestures may actually prime relevant words or ideas in the lexicon (e.g., Krauss, 1998), and that they may help speakers conceptualize what they want to say and package the ideas into the linear stream of speech (e.g., Kita, 2002). In sum, the cognitive functions of gesture are varied, and have been shown in a variety of domains ranging from problem solving to memory to language.

We believe that these varied functions can be explained under a general mechanism suggested by Predictive Processing Perspectives (PPP). Not only are PPP highly compatible with the GSA framework and the suggestion that gestures arise out of embodied neural simulations, but they also provide further explanation for how gestures' function is not reducible to these neural simulations. Rather, action (and thus gesture) is central to the job description of the cognitive

Predictive Processing Perspectives

PPP are a recent sea change in cognitive science (for broad overviews see Clark, 2013a, 2015b; Hohwy, 2013). As PPP are rapidly adapting, they are becoming more divergent from each other (see e.g., Pickering & Clark, 2014). Yet, what unites these models is that they assign a single job description to the cognitive system under which most, if not all, cognitive feats (e.g., perception, action, social cognition, language production and comprehension) can be subsumed. Namely, the cognitive system is engaged in optimizing predictions about the continuous flow of sensory data that perturb the system during the ongoing flux of (potentially hazardous) interactions with the environment. Minimizing prediction error is not some abstract project, but key to the perseverance of life: simply put, "*avoid surprises and you will last longer*" (original emphasis, Friston, Thornton, & Clark, 2012, p. 2).

We will argue that gesture is one special way to optimize predictions. We do this by showing that action-oriented models in PPP (Clark, 2015a, b; Friston, 2009) are compatible with the GSA framework, and are able to clarify the mechanism by which gestures benefit cognition. Two important aspects of PPP will be considered. First, PPP put sensori-motor neural simulations in a broader context of a hierarchical predictive architecture. Second, PPP assign a pivotal role of action within this broader predictive machinery. Thus, by considering gesture as a special case of action, we can use PPP to understand both the cognitive origin and function of gesture.

9

Before introducing basic tenets of PPP, some preliminary remarks are in place. First, we only provide a broad conceptual overview of some key mechanisms of PPP (e.g., Clark, 2015b) that we think are relevant to thinking about gesture, neglecting statistical formalisms (e.g., Bayes Theorem) that ground PPP (see Friston, 2009, 2010; but see Hohwy, 2013 for an approachable introduction). Second, although it is largely undisputed that there is *a* predictive component in many central cognitive processes, such as attention (e.g., Hohwy, 2013), vision (O'Regan & Noë, 2001), action (e.g., Franklin & Wolpert, 2011), and language comprehension and production (e.g., Lupyan & Clark, 2013; Pickering & Garrod, 2013), models within PPP are still highly debated, and at present there is no evidence to decisively choose among competing models. Thus, our account is an attempt to show the preliminary utility of PPP for understanding gesture's production and function, rather than an endorsement of any one view.

Introduction to PPP

Predictive Processing Perspectives (PPP), as presented by Clark (2013a, b; 2015a, b; based on Friston, 2009, 2010), entail that the cognitive system has, and continuously adapts, a body of knowledge (called a 'hierarchical generative model') that allows the agent to self-generate data (called 'predictions' or 'prior expectations') about the world that capture the statistical regularities of incoming sensory data. These predictions mostly run on automatic pilot, and need not be subject of awareness to do their work (although they may be constrained by conscious processing). Importantly, a generative model is never perfect, and its predictions never completely match the incoming sensory input. In fact, these discrepancies between incoming sensory input and the predictions are informative and continuously monitored. These discrepancies, called 'prediction-error', are used to update the generative model in order to issue more precise predictions in the future. Thus, prediction errors are used to calibrate future predictions, and over time enable the generative model to make better predictions about the world.

The generative model is "hierarchical" because predictions are issued on multiple higher and lower order levels. Lower order levels issue fast-changing *sensory* predictions, likely to operate on timescales ranging from hundreds of milliseconds to seconds. For example, when reaching for a mug, lower order haptic predictions are produced about the instant consequences of picking up the mug. Higher order levels are more likely to be abstract and multi-modal, and operate on longer time-scales. That is, these levels are not concerned with one particular sensory consequence, but with complex multi-modal regularities that emerge over longer periods of time. Slower predictions might concern keeping track of a trajectory of a moving object, or for slower predictions still, how particular types of situations generally unfold (e.g., restaurant visits, idle conversations etc.). In each level, the model predicts the output from the level below and compares this predicted output to the actual input received from the level below. This results in a complex multi-layered predictive machinery that works in concert to track relevant small- and large-scale changes that have proved to be relevant to the agent in the past.

In many ways, PPP are a reversal of classical models of perception - wherein the cognitive system passively receives input in a bottom-up fashion. Rather, in PPP, the mind has a more active, anticipatory and self-adaptive role[17]. In PPP, action is an essential part of perception, as active sampling can make perceptual patterns that are predicted come true and allow the agent to make better predictions. For example, I may not know for sure that my coffee cup is empty, but when I grasp the handle without enough muscle tone to account for its filled weight, I sample unexpected proprioceptive feedback (i.e., prediction-error is produced) that informs me to adjust my prediction about the cup's fullness, as well as what action is appropriate. To continue with this example, in some versions of PPP (Clark, 2015b; Friston, 2009; 2010) the proprioceptive consequences of a *full* coffee-mug result in prediction error that activates the motor system to adjust to a grasp that optimally deals with a full cup instead of the present grasp (empty cup). In fact, in such versions of PPP, *all actions* are produced by the motor system to resolve prediction errors by making predictions about the consequences of actions true by actually performing those actions. Simply put, when an action is predicted in a particular context, the consequences of those predicted actions are compared to the present state of the system. This results in prediction-errors that are resolved by acting on those predictions.

Minimizing prediction error with action is called active inference. Active inference reduces prediction error using what Pickering and Clark (2014, p. 451) refer to as "twin strategies." Predictions are altered to fit the environment and the environment and body are altered through action to fit the predictions. Further, action can simplify what we might call the "predictive load" of the generative model (see Clark, 2015a, b). Namely, actions are informative for updating predictions, as active sampling provides information otherwise not (as reliably)

---

[17] Interestingly, computer vision research has yielded productive results by implementing just such an active model of vision (e.g., Rao & Ballard, 1999), and this model has unique explanatory power with regards to persistent optical illusions experienced by humans (for an overview see Hohwy, 2013).

available in a passive mode. As Clark (2015a, p. 15) says, "the course of embodied action to novel patterns of sensory stimulation, may thus acquire forms of knowledge that were genuinely out-of-reach prior to such physical-manipulation-based re-tuning of the generative model." As an illustrative instance, Clark (2015a) points to abacus-training, wherein children are able to learn to perform complex arithmetic by using an abacus, and learn to perform these calculations without an abacus after sufficient training (Stigler, 1984). Learning by acting on an abacus allows the generative model to shape predictions with more reliable inputs (e.g., the results of the actions themselves), and effectively reduces the degrees of complexity of the generative model itself (see Kirsh & Maglio, 1994 for a similar example).

How is an agent able to flexibly employ the different strategies for prediction error-minimization? For example, in some cases active sampling is not an option, and inference on the basis of present input is more appropriate. PPP employ precision-estimation as a mechanism that allows for the flexibility of predictive strategies. Namely, every prediction and sensory input is given a certain second-order 'weighting' (called a 'precision estimate') on the bases of its predicted accuracy. That is, given the context (e.g., say a misty day, or a dark room), the cognitive system may treat incoming sensory signals as less reliable (i.e., lower precision estimate), which results in relatively higher precision estimates of top-down predictions. This allows the agent to behave according to prior knowledge (e.g., anticipating stop signs on the misty road you are driving on; navigating the dark room based on memory) as a more reliable way to reduce prediction error than relying only on sensory bottom-up information. In contrast, in a completely novel situation, it may be difficult to form top-down predictions with any amount of accuracy. In such situations, action becomes increasingly important as a means of learning the environment. Thus, precision estimates allow the system more flexibility, as in some situations the environment can be used as its own best model, whereas in others, a top-down model for the environment may be more effective (Clark, 2015a, b). Precision estimates allow the system to determine which is best.

We believe that there is synergy between the key concepts of PPP and those of the GSA framework. In the sections that follow, we will explore how each of the three determinants of whether a gesture is produced as proposed in the GSA framework can be explained by PPP. Further, by considering gestures as action in a PPP, some predictions about gesture function naturally emerge.

Action simulations are strongly activated when prediction error is high

The GSA framework holds that gestures arise from action-simulations, wherein the strength of motor activations predicts (in part) the likelihood of overt gestures. The strength of activation is determined in large part by the manual motor-affordances that are solicited by the environment or content of speech (e.g., Chu & Kita, 2015; Masson-Carro et al., 2015).

In PPP, action is produced as a means of resolving the prediction error that exists when an action is predicted but one's body state is different (e.g., static). In order to think about an event that involves action, speakers' cognitive system must predict what actions are involved and what the proprioceptive and visual consequences of those actions would be. Creating such predictions in the absence of overt movement results in high prediction error, as the cognitive system predicts that movement should be occurring but does not receive the kinematic feedback of such movement. To resolve this prediction error, the speaker's motor system may be activated to produce congruent movement. Such movement is recognized as gesture when it occurs alongside speech.

Thus the claim that gestures occur when action simulations are strongly activated in the mind of a speaker is compatible with the basic claim of PPP. For example, PPP can accommodate the idea that when the relevant content of speech is actional (e.g., throwing a ball) gestures are more likely, than when the content of speech is about visual-spatial (e.g., seeing a house) or abstract concepts (e.g., democracy), as action is part of the prediction formed by the cognitive system in the former case. To think (and talk) about throwing a ball without actually producing the corresponding action requires the cognitive system to tolerate a higher amount of prediction error in the motor-system. Under PPP, such a state is not desirable; thus, an action is likely to be produced as a means of resolving the prediction error.

Consider the case of mental rotation, in which participants are asked to imagine the visual consequence of rotating some object a specified amount around its axis. In such a task, the visual information associated with rotation of the object must be predicted top-down by a generative model that captures sensory consequences that co-uccur with such rotations based on previous experience. This requires spatiotemporally fine-grained visual predictions of a moving object. In the terminology of PPP, prediction error in such a situation is high, as the top-down predictions of the object's visual appearance as well as motor-associations following rotation do not match the sensory input of the objects' given starting position. To resolve the error, an action may be initiated, even one which does not actually manipulate the object. Indeed, during mental rotation, either co-occuring with or without speech, participants naturally adopt gestures as-if manipulating the object to be rotated (Chu & Kita, 2008; 2011). In terms of the GSA framework, such gestures occur because the movement involved in rotation is being simulated strongly in the speaker's mind, in order to determine its endstate (see Hostetter, Alibali, & Bartholomew, 2011).

Of course, not all gestures are direct pantomimes of action. Many gestures take a form of outlining or tracing a described object. For example, a speaker might say "it was round" while tracing a circle shape in the air. In such an instance, it is difficult to see how the gesture could be reducing prediction error between a predicted action and the kinematic absence of such action. However, in PPP, predictions are multimodal, meaning that they are not limited to action predictions but can also involve visual predictions. Thinking (and talking) about a ball does not only involve predicting what corresponding actions go along with a ball, but also what the ball looks like. Creating an image of the ball with one's hands could be a way to minimize the prediction error that is inherent to talking about how an object looks without getting sensory input about the object's actual appearance. Indeed, speakers gesture less about objects that are visually present than about objects that are not present (e.g., Morsella & Krauss, 2004). This could be because there is less prediction error involved in talking about an object that is visually present in the environment, so action is not as likely to be initiated. In contrast, when there is no visual object present, gestures make the inferences about the object made in speech become true. Under this view, proprioceptive predictions, that are first inherent to action processing, become multimodally associated with depictive visual-spatial processing (e.g., shape of a house) over development. We speculate that the proprioceptive feedback of an action or gesture comes to activate relevant visual-spatial details as well (see Cooperrider, Wakefield, & Goldin-Meadow, 2015; Pouw, Mavilidi, Van Gog, Zwaan, & Paas, under review, for some evidence on this point).

In sum, the GSA framework proposes that gestures are automatically activated as the result of activation in the sensori-motor system during speaking and thinking. This is congenial to the idea that top-down predictions about sensori-motor events are continuously employed by the cognitive system. Furthermore, that gestures are most likely to be produced when there is a disconnect between the physical and mental environment (e.g., when action is being talked about or when a visual scene is being described that is not visually present) suggests that gestures may emerge precisely when prediction error related to the motor-system is high. In the terms of the GSA framework, action simulations become strongly activated in such situations, and this high activation leads to gesture.

Gesture threshold is adjusted based on precision estimates

Recall that the GSA framework argues that simulations underlying gestures are automatically activated, but that their overt production as a gesture is dependent on a number of contextual factors captured by the 'gesture-threshold'. Speakers can adjust their gesture rate (either consciously or unconsciously) as the result of such things as believing that a gesture will be helpful to either themselves or their listener.

This is similar to the way that predictions and sensory inputs are given precision estimates in PPP, so that the cognitive system can rely more on one or the other in a particular situation. When sensory input is degraded, the cognitive system may favor top-down predictions. When top-down predictions seem insufficient, the cognitive system will seek out sensory input to provide new information through active inference. For example, Tetris players often rotate blocks to decide where to best place them (Kirsh & Maglio, 1994). Producing the rotation movement provides more reliable bottom-up information than top-down predictions (i.e., mental rotations) of where the pieces will fit best.

As mentioned above, a similar process has been observed with the use of gesture during mental rotation tasks (Chu & Kita, 2008, 2011). Most important for our discussion of prediction estimates and the gesture threshold, however, is the finding that participants do not always gesture during such tasks. In cases where the rotational angle is smaller, those participants who generally gesture in more difficult trials may not adopt gestures. At smaller rotation angles, the task is easier, and as such, the precision estimate of top-down visual and motor predictions of the rotation is set to be more reliable (given previous successes in the past) and thus active inference (gesture) is less likely to occur.

This could also explain the findings that those with a lower (as opposed to higher) working memory capacity are more likely to gesture during speech production (e.g., Chu, Meyer, Foulkes, & Kita, 2013; Gillespie, James, Federmeier, & Watson, 2014). For participants with limited working memory systems, top-down predictions are generally more unreliable, leading them to adopt gestures as a means of providing more accurate sensori-motor predictions. In the terms of the GSA framework, such speakers intuit the potential benefit of gesture and thereby set a low gesture threshold so that many of their simulations come to be expressed in gesture. Indeed, Dunn and Risko (2015) found that metacognitive judgments of whether an external rather than internal strategy is more efficient directly predicts how problem solvers approach a task. Thus, precision estimates and reliability judgments may determine whether gestures are produced.

The idea in the GSA framework that speakers can intuit whether a gesture is helpful or not is compatible with the idea in PPP that the cognitive system employs precision estimates as a way to give preference to sensory inputs or top-down predictions. In situations where producing a gesture could help the system visualize the details of the top-down prediction, a gesture is more likely to be produced.

Simultaneous speaking prevents complete inhibition of motor system

In the GSA framework, the final predictor of gesture is whether speech is accompanying the simulation. The GSA framework proposes that because the vocal articulators must be moved during speaking, it is difficult to completely inhibit the motor activity involved in simulation from being expressed as gesture. Although gestures can and do occur without speech (e.g., Chu & Kita, 2015), gestures are typically more prevalent alongside speech than in its absence.

This explanation is in line with the mechanics of PPP. Recall that sensori-motor predictions can be inhibited in situations where top-down predictions are estimated to be more accurate. This is sometimes referred to as "gain control", or the system's ability to gate sensori-motor predictions so that only weak signals are sent to the muscles (e.g., Grush, 2004). However, when the motor system must be involved in the interest of producing speech, it is difficult to completely inhibit all motor signals from being sent to the muscles. In their Action-Based Language theory, Glenberg and Gallese (2012) offer a PPP on language, positing that language learning, comprehension, and production capitalize on systems for motor control. They follow the GSA framework in proposing that gestures are the result of activating relevant actions alongside speech paired with an inability to completely block these movements from being expressed because speaking requires movement of the mouth and vocal articulators. Thus at least one PPP has already offered an account of gesture fully in line with that provided in the GSA framework.

As evidence for this account, consider that the articulatory/oral system and manual system are closely entrained. For example, humans often open and close their mouths during skillful manual manipulation (Darwin, 1998). It has been found that when grasping an object with one's mouth, the size of the mouth opening correspondingly affects the size of index-thumb aperture (Gentiulucci et al., 2001). This is also the case the other way around; the size of the manual grasp of an object affects the size of the aperture of the mouth. Furthermore, when participants had to grasp an object and simultaneously name a syllable printed on it (e.g., "GU", "GA"), the size of the manual grasp aperture affected lip opening as well as voice patterns during syllable expression, showing a clear entrainment between manual action and the articulatory system. In sum, the proposal that gestures are likely to occur alongside speech because motor activity cannot be completely inhibited is compatible with PPP and the existing literature about the mutual entrainment of the oral and manual systems.

Summary

We have shown that the basic tenets put forth by the GSA framework regarding how gestures emerge in the cognitive system are compatible with the claims made by Perspective Processing Perspectives (PPP). Put simply, gesture is produced by prediction errors that reach the motor plant. Namely, when the system predicts some motor-activity, it will produce prediction-errors - as there is no motor-activity yet that matches the predictions - which will in turn activate gestures. This is akin to what the GSA framework calls strong activation of action simulations. Yet there are constraints, on whether the motor system is activated. For example, when precision estimates of incoming motor-sensory signals are low relative to top-down predictions, than prediction will not be quashed by action, as the prediction error that results will be deemed less reliable and top-down predictions will suffice. This is akin to one way in which the GSA framework conceptualizes the gesture threshold, or the idea that action simulations must be strong enough to pass some resistance to gesture. When gesture does not seem useful to the cognitive or communicative situation, action will not be activated strongly enough to be realized as gesture.

Utilizing PPP to think about gesture offers more than just a shift in terminology. On the contrary, in PPP, active inference is a central catalyst for cognition, suggesting that action in the form of gesture may also benefit the cognitive system. As will be explained in the following section, thinking about gesture not just as simulated action, but also as predictive action offers a general explanatory mechanism for how gestures have their facilitative effects on cognition.

Gesture as Predictive Action

In PPP, action can serve as a means of reducing predictive load. That is, by engaging the motor system in action, the cognitive system is able to sample information about the consequences of a particular action that is more precise than the information gleaned from a top-down prediction. We contend that gesture, like action more generally, can have this same effect, by providing the cognitive system with useful sensori-motor information.

We suggest that the act of gesturing provides visual and proprioceptive feedback about the consequences of action that is not available in a static state. Gestures thus provide multimodal information that corresponds to (as they normally co-occur with) the causal consequences of *actually* acting on an object. These consequences thus inform top-down visual and motor predictions with actual kinematic information, which is arguably more reliable than having to predict such consequences completely top-down. What results is a generative model dealing with more reliable externally supplanted (visual, and proprioceptive) information, that allows for less risky (i.e., more accurate) perceptual inferences. This process has a number of potential benefits to the cognitive system.

For example, consider the well-documented finding that gestures reduce working memory demands, as speakers who gesture during a primary task (e.g., explaining their solution to a math problem) are able to perform better on a secondary task (e.g., remembering a string of letters) than speakers who do not gesture (e.g., Goldin-Meadow et al., 2001). In our view, this effect occurs because gestures have reduced the predictive load involved in the primary task. For instance, as speakers describe how to solve a mathematical factoring problem, they use their hands to explore how the numbers move to the relevant positions in the problem space. These gestures provide visual and proprioceptive feedback about where the numbers should be positioned, thereby making it easier for the generative model to operate as the solution is described. As a result, the cognitive system has more resources available to devote to a secondary task (e.g., remembering a list of letters).

The feedback provided by gesture may help problem solving, as well. For example, when speakers are solving a mental rotation problem, moving their hand as they would if they were actually turning the block to be rotated will activate visual information about the end state of that rotation. As they attempt to predict the objects' end state, participants gesture as a form of active inference to determine what the sensorimotor consequences of various amounts of rotation will be, and in doing so, actually provide themselves with information about what those sensorimotor consequences are. The same effect is seen as participants solve the Tower of Hanoi. Producing gestures as-if manipulating the physical apparatus affects problem-solving performance compared to not gesturing (Cooperrider et al., 2015). Such gestures inform top-down predictions with relevant kinematic information. Indeed, when this kinematic information is not relevant to solving the task, performance is hampered (Beilock & Goldin-Meadow, 2010).

A similar effect occurs during mental abacus. Abacus users, after repeated training, learn to do complex arithmetic without the abacus; top-down predictions are doing most of the work in these cases. Interestingly however, abacus users transitioning to do arithmetic without the abacus often use gestures, *as-if* manipulating the beads of an actual abacus (e.g., Hatano & Osawa, 1983). With time, these gestures dissipate, and abacus users learn to do calculations without moving, although they appear to still be using a strategy that involves imagining use of the abacus (see Frank & Barner, 2012 for evidence). Thus, gestures seem to offer some in-between strategy wherein they can supplant the now absent information normally afforded by a physical abacus. Indeed, while the abacus is absent, the affordance of generating proprioceptive and visual consequences that normally occur with acting on an abacus are ready to hand when gesturing. Using this second-hand information afforded by gesture (rather than interaction with the abacus), allows the generative model to deal with a certain amount of uncertainty still present in top-down predictions.

Can this account also explain the effects of gesture on linguistic processing? For example, gesture production has been shown in some circumstances to act as a cross-modal prime that speeds access to corresponding words (Krauss, 1998). In their Action-Based Language Model, Glenberg and Gallese (2012) explain this as occurring because the predictors associated with a physical action and the predictors associated with the articulation of the lexical label for that action are overlapping. We build on this explanation to offer the following account. When speakers are thinking of describing a particular action, they attempt to access the action plan for articulating the correct lexical label. When the precision estimate for accessing this label is low, speakers engage gesture as a means of gathering more information. Because linguistic knowledge is grounded in sensorimotor experience (e.g., Barsalou, 2008; Glenberg & Gallese, 2012), the act of gesture can provide proprioceptive, visual, or kinematic cues that then strengthen activation of the word.

This is especially apparent in the case of gesture-speech mismatches, in which a speaker conveys information in gesture that is not conveyed in the immediately accompanying speech (e.g., Church & Goldin-Meadow, 1986). Such gesture-speech mismatches have been observed in children (and adults) in a wide variety of learning tasks (e.g., solving mathematical equations, balance beam problems, and chemistry problems), and predict children's learning trajectory (for a review see, Goldin-Meadow & Alibali, 2015). When learning a new task, children tend to first produce incorrect solutions in both gesture and speech. With additional learning, it becomes more likely that a correct solution is expressed in either gesture or speech (but not both), before the child finally settles into a stable state where gesture and speech both express a correct strategy (Alibali & Goldin-Meadow, 1993). Thus, it seems that learning does not follow an either-or transition of understanding (i.e., eureka!), but a negotiation of different ways of understanding brought forth through gesture and speech.

From the present perspective, these different "ways" of understanding correspond to the different kinds of predictive processing that govern gesture and speech. Namely, explaining a solution in speech involves predictions that are linear and rule-based (i.e., knowing-that). Speech targets regularities that are present on slower time-scales, which can be applied to several phenomena independent of a single observation (e.g., in a conservation task, knowing that any action could be undone to return to the original state). Yet, these abstract regularities need to be observed and discovered to become articulable. Here, gesture comes into play. As the child thinks about the task without a clear top-down solution in mind (i.e., top-down prediction estimates for speech are low), the child initiates gesture as action to explore the manual affordances of the apparatus. These gestures are governed by the task's predicted manual affordances (e.g., know-how) and not necessarily by representations of an abstract rule (Pouw, Van Gog, et al., in press). The gestures provide proprioceptive and kinematic information about the transformation that goes beyond what the child can see in the stimulus. Through repeated instances of gesturing, invariants can be discovered that become parsed in meaningful sequences that correspond (or not) with segments in speech. Under this view, a stable state is reached when the prediction error between discovered higher order invariants in gesture are resolved with categorical speech predictions that target those invariants.

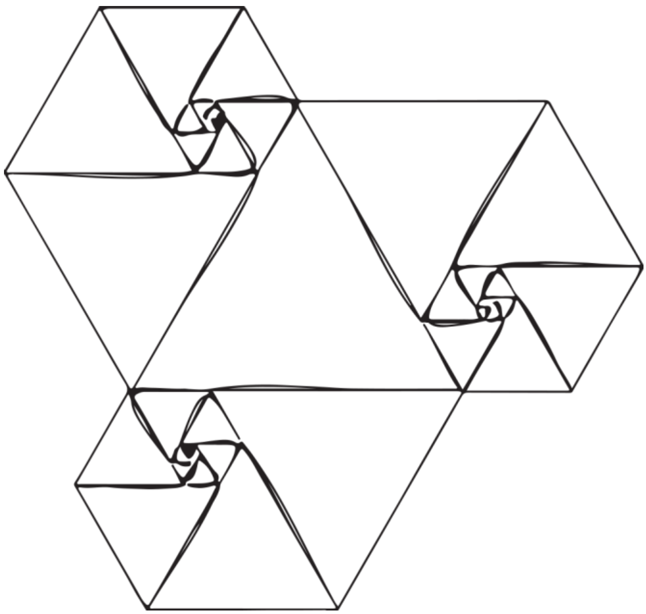
In sum, we believe that many of the documented effects of gesture on cognition and language can be explained by considering gesture as a case of active inference. By engaging action/gesture, the cognitive system creates new bottom-up input that can inform the top-down predictions necessary for a problem solving or language task.

Concluding Remarks

In conclusion, we have offered a preliminary sketch for considering gesture as a case of predictive action. Considering gesture as an example of action in a Predictive Processing Perspective offers a powerful description of how gestures come to have their facilitative effects, as well as how they arise out of anticipatory sensori-motor states (simulations vis-à-vis top-down proprioceptive predictions). Under this view, the distinction between the origin of gesture and their function in the cognitive system is not so clear. Gestures occur because they can have powerful effects on the cognitive system, yet the effects they have are the direct result of their origin as simulated and predictive actions in that cognitive system.

Chapter 10

Summary and General Discussion



The overarching questions guiding the research presented in this dissertation were: If cognition depends directly or indirectly on bodily processes, then what are the implications for learning and problem solving? In other words, what type of behaviors and interventions would we expect to impact learning and problem solving if cognitive processes are dependent on bodily processes?

This dissertation consisted of three parts. Part I focused on the applicability of embodied cognition approaches for improving learning from instructional animations and manipulatives. More precisely, it was concerned with the question of whether manual activity during learning can change the way learners represent problem solutions related to the learned content, thereby improving their learning outcomes. Part II investigated the occurrence of non-communicative hand-gestures and their influence on problem-solving performance. In part III, which includes this discussion chapter, the main ideas in the dissertation are synthesized and interpreted in light of recent developments.

The key findings from each of the three parts are summarized and briefly discussed in relation to the overarching questions of the dissertation.

Part I: Effects of Manual Activity on Learning

Part I focused on whether learning in more bodily engaged ways may result in thinking about the learned content in more productive ways. In **Chapter 2**, a theoretical review paper was presented concerning instructional manipulatives. The early rationale (advocated, for instance, by Montessori [1870-1952] and Pestalozzi [1746-1827]) behind the use of instructional manipulatives is that essential information can be picked up through interaction with the manipulative that is difficult to obtain in more passive modes of learning. Yet, the review of the recent literature on instructional manipulatives showed a more negative sentiment concerning its effective qualities (e.g., McNeil & Jarvin, 2007). Part of this negative sentiment is driven by research that shows digital mouse-based instructional manipulatives are as effective, or even more effective, than classic physical manipulatives. *Digital* manipulatives typically have less degrees of self-guided interactive freedom, lower visual complexity and a smaller range of kinesthetic experiences. According to some researchers studying manipulatives (e.g., T. De Jong et al. 2013, Triona et al. 2005), such findings are surprising from an embodied cognition perspective, which would predict more physical experiences lead to richer knowledge-representations.

However, we came to a different conclusion based on a reassesment of basic findings that ground the embodied cognition approach, and related findings on instructional maniplatives. The most important finding was that perceptual and interactive richness of manipulatives are effective only when such features are *functionally* relevant for the concept that is being learned. To provide one example, it has been shown that chemistry students who have to learn how to translate one molecular scheme into another are aided by manipulating a physical model of the molecular bond (Stull et al., 2012). Indeed, basic research has already shown that expert tetris players can perform on such high levels because they minimize the mental rotation of zoids by rotating the zoids directly rather than mentally (Kirsh & Maglio, 1994). The translation of molecural schemes, too, requires effective (mental) rotations. Importantly, this effect is ill-interpreted when assuming that these findings show that "more" sensori-motor experiences improve learning. Indeed, Stull and colleagues (2013) have shown in a subsequent study that when rotations can be made with a digitially presented 3-D molecular model, this can be as effective as learning with a physical manipulative. And this effect was attributed to the fact that digital rotation enabled learners to rotate the bond on one central axis, reducing additional degrees of freedom that are not productive for thinking about the problem (i.e., reducing the number of axes on which a molecular model can be rotated). This example shows that reducing the richness of kinestethic experience with a manipulative can allow for more productive interactive (i.e., "embodied") information to emerge that is more effective than passive modes of learning. Very simply put, sometime less is more, as less interaction in this case solicits more productive embodied problem solving. We conclude on the basis of these and related findings that embodied cognition fosters understanding of the effectiveness of instructional manipulatives, if it is pinpointed what relevant information can emerge through interaction with the environment that is not available in passive mode. As such the digital vs physical debate can be illuminating in further assessing the functionally relevant information that emerges in interaction.

In **Chapter 3** we attempted to manipulate functionally relevant information in interactional learning. Consider that the central assumption in grounded cognition (applied to manipulatives) reads that physical experiences that are obtained during learning with manipulatives may become functionally re-used when reasoning without such manipulatives (Barsalou, 1999). To use a previous example, we can expect that having rotated the molecular bond on one central axis will result in *mental* rotations on that central axis, thus leading to more productive mental rotations *because* of previous physical experiences (see e.g., Chu & Kita, 2011; Flusberg & Boroditsky, 2011). To further avoid a simplistic reading it should be made clear why some physical experiences (rotating on one bond) are functionally relevant to (learning to) perform some cognitive task. As such, Chapter 3 was designed to gauge the potential benefits of learning through physical interaction by manipulating the meaningfulness of the information that could be gained through interaction.

Participants (76 university students) learned from an interactive instructional animation about the key principles of class-1 levers (represented by a seesaw). All particpants were physically interacting with the instructional animation by pushing with their hands on Nintendo Wii-Board. As such, the mode of interaction was kept constant for all participants, to avoid possible confounding motivational factors related to having to physically interact versus being in some passive learning condition (e.g., Kontra et al., 2015; Skulmowski et al., 2016; Zacharia et al., 2012). Instead, we manipulated the type of information to be gained through interaction in three ways. Firstly, a third of the sample engaged with the instructional animations through the Wii-board, by learning to apply the appropriate force to balance a seesaw depending on the presented counterweights and their relative positions on the beam (as to learn about the principle of mechanical advantage, see chapter 3). Importantly, in this "meaningful" interaction condition there was a lawful correspondence between the physics of class-1 levers and the different forces that the participant learned to apply. In contrast, a third of the sample was assigned to a "non-meaningful" interaction condition, wherein the forces that needed to be applied to balance the seesaw were inconsistently related to the physics of class-1 levers. The final third of the sample was assigned to a "minimal" interaction condition, wherein participants briefly applied a consistent force (with two hands) on the wii-board to make a seesaw balance out. After the learning phase participants performed three test-tasks, each designed to tap into different types of knowledge. Firstly, participants performed a three-choice reaction-time task wherein they needed to judge (as fast as possible) what the end-state of a seesaw would be given the number of counterweights and their positions on the seesaw. This task was designed to tap into automatic, pre-reflective knowledge about the dynamics of the seesaw. In the transfer task, participants had to extend the learned principles to more complex representations of class-1 levers (e.g., judging the endstate of interconnected seesaws). This task thus required a more reflective reasoning process as inferences needed to be made from the more basic principles covered in the instructional animation, and to be applied to a new problem situation. Finally, we assessed participants "motor-knowledge", by letting them judge over several trials the amount of force that needed be applied to balance a seesaw given the amount of counterweights and their positions.

It was expected that having had meaningful physical experiences that lawfully correspond to the physics of the seesaw would result in higer performance on subsequent test-tasks compared to the minimal and non-meaningful conditions. The rationale is that the motor-knowledge that learners acquire to correctly apply force to balance a seesaw becomes mentally reused for further problem solving with class-1 levers. Such motor-knowledge is expected to be functional, as it would allow learners to predict end-states of levers on the basis of the differing motor-associations that are automatically re-experienced, or "reenacted" (Barsalou, 1999). Inversely, providing physical experiences that are inconsistent with class-1 lever physics should be detrimental to subsequent performance (as compared to the minimal and meaningful condition) as the motor-associations that are reexprienced do not map onto the correct physics of the problem, and are therefore not predictive for the solution of the problem.

Yet, the findings showed something different. No effects of instructional condition on performance were obtained on either the reaction-time task or the transfer-task. Interestingly, there was some divergence on the motor-knowledge task. Although a pre-defined ratio did not show significant divergence on this measure, visual inspection of the force-curves of this task clearly indicated that when asked to judge the dynamics of a seesaw motorically, the participants assigned to the non-meaningful condition were indeed less able to motorically differientate between different forces than participants in the meaningful condition. Moreover, even though partcipants in the minimal condition did not learn to differientiate forces, they seemed to apply different forces true to the correct physics of the problem, suggesting that the visual content of the instructional animation may have led to some transfer in the motor-domain. We can highlight three main implications of this study. Firstly, if the current null-results are on track, then it seems that a relatively short physically interactive learning intervention is not readily functionally redeployed in further reasoning, although it does leave traces in motor-knowledge. Secondly, it might be that the visual information obtained from the instructional animation is relatively dominant, as there was some transfer to the motor-domain for participants in the minimal condition. Finally, even regardless of the results, the design of the current study provided a novel way of assessing the functional redeployment of physical experiences.

In **Chapter 4** a different application of basic research related to indirect approaches to embodied cognition was tested for improving learning from non-interactive instructional animations. Again in the context of class-1 levers, but this time with a younger age-sample (74 children between 10-13 years of age), it was assessed whether instructional animations augmented with an analogy to physical principles inherent to the mechanics of the body improves learning as opposed to an otherwise identical instructional animation without such an analogy. This study was informed by previous research that has shown that mental rotations are performed more quickly and accurately if the entities under rotation are human bodies versus abstract (but otherwise identical) objects (Amorim et al., 2006). The rationale behind this improved performance is that humans are intimitaly aware of the spatial properties of their own body, and mental transformation of this engrained spatial body image may sometimes be required to act appropriately on the environment (e.g., will my body fit through this apperture? Can I reach this object? etc.). As such, this practical bodily know-how seems to be functionally reemployed when thinking about spatial transformations, according to an embodied cognition perspective on these findings.

In this study we assessed whether it is possible to improve learning about a physical system by tapping into the knowledge inherent to bodily dynamics. Namely, we reasoned that the principle of mechanical advantage, a key learning principle of class-1 levers, is congruent with physical experiences of placing weights on different positions on the arms. To show this, imagine that you have two equally weighted objects, one of which is placed on the left arm that is extended horizontally from the body at the position of the elbow, and the second weight is is placed on the extended right arm, but in this case the weight is positioned at the hand. Although the weights are of equal mass, the object placed furthest from the shoulder will feel heavier than the object closer to the shoulder (the weight placed at the elbow's positions). Identically, when an object is placed on the left side of a seesaw (analogy: left arm[18]) closer to the fulcrum (analogy: left shoulder), and an object of equal mass is placed on the right side of the seesaw (analogy: right arm) further from the fulcrum (analogy: right shoulder), the seesaw will pivot the the right (analogy: the right arm will feel heavier). This analogy was put to the test. All children learned through an instructional animation wherein the basic principles of class-1 levers were explained (using a seesaw). Half of the participants learned with an instructional animation in which a transparent body was projected over the seesaw. The other half of the sample was not given this body analogy. Afterwards children were tested on a three-choice reaction time task and a (simplified) transfer task similar to the procedure used in Chapter 3. As younger age-samples show large individual differences in cognitive skills (as compared to student-samples), we pseudorandomly assigned children to the condition (control vs. body analogy) by matching for level of general math skill.

The results showed that there was an effect of the body analogy on accuracy on the reaction-time task, but this effect was qualified by an interaction with general math skill. Children with relatively lower general math skills were more likely to learn from the body analogy (i.e., provided more correct answers) as compared to children with similar math skills who did not learn with a body analogy. Similar effects were not found for answer-speed on the reaction-time task, nor for accuracy and time-on-task on the transfer task. Results did however show that children with a relatively lower general math skill showed higher speed-accuracy trade-offs on the transfer task when having learned with a body analogy as opposed to the control animation. Thus, results on both the reaction-time task and the transfer task do not converge on whether the body analogy is more effective, but do converge on the finding that especially those with lower skill in math are likely to be affected by the body analogy, either for better (i.e., increased performance on the RT-task), or for worse (i.e., speed-accuracy tradeoff on the transfer task). We interpret these findings as being potentially promising for educational practices, as it signals that a relatively simple modification that allows learners to map relevant bodily dynamics onto a physical system can improve learning for some children without hampering learning of other children. Yet, although promising, the findings are currently far from applicable yet (for a more thorough discussion see chapter 4). This is mainly because the effect

[18] Coincidentally, in instructional texts the analogy is already implied as the left or right side of the seesaw relative to the fulcrum is referred the left or right *arm* of the seesaw.

of the body analogy seems to meet its boundary condition as soon as the analogy needs to be applied to slightly more complex instantiations of the class-1 lever (i.e., as assessed through the transfer task). Therefore, we recommended further exploration of the the applicability of body-anologies to other physical systems, while being sensitive to the cognitive pre-dispositions of learners.

Part 1: Brief conclusion and suggestions future research

Taken as a whole, the studies in part 1 took a closer look at the relevance of embodied cognition for Educational Psychology. The literature review suggested that more bodily engagement during learning can be effective because of the qualitatively different information it can provide compared to passive modes of learning. Yet, although reframing the relevance of embodied cognition for learning interventions in this way is more theoretically sound than the idea that more bodily engagement is better, further research is needed to pinpoint what qualitative differences matter for learning exactly, if at all[19]. In the case of learning about science concepts, future studies could become more sensitive to the functional information that is afforded by a particular intervention by loosening and tightening the correspondence of perception-action loops with the physics of the learning concept, instead of manipulating only the presence of interaction (active vs. passive learning). Furthermore, more research will be needed to assess the boundary conditions of learning in concrete contexts as it is relevant for children as opposed to adults, and whether bodily learning in concrete contexts transfers to more complex situations.

Part 2: Effects of Manual Activity on Problem Solving

The chapters presented in **Part 2** include a theoretical analysis and a set of experiments focused on the cognitive function of gestures for the gesturer. The main question addressed in this part, is how gestures (i.e., specific bodily acts) are functionally important for problem solving.

---

[19] This assertion that the investigation of the functional nature of learning effects in embodied cognition research is key for its development as a field resonates further with other recent calls for action, for example in the domain of language processing (Zwaan, 2014, 2015; see also Wilson & Golonka, 2013).

The theoretical analysis presented in **Chapter 5**, questioned how gestures can aid cognitive processes of the gesturer. That gestures can aid cognitive processes is well-established. The effects of gestures that occur in problem solving contexts are often accommodated by the idea that gestures enable activation of internally stored knowledge about actions, referred to as motor-simulations (Beilock & Goldin-Meadow, 2010; Chu & Kita, 2008, 2011, 2016). The main theoretical point made in Chapter 5 is that there is a logical objection to this general explanation. Namely, gestures are not effective for spatial problem solving *in virtue* of tapping into pre-existing internal representations of actions). This is because the activation of the motor-representation is classically employed to specify the production of actions, such as a hand-gesture. A gesture is thus logically prior to the activation of a motor-representation (Hostetter & Alibali, 2008). What the exact beneficial cognitive function of the actual physical production of a gesture might be is therefore not clear, or trivial when solely understood as keeping "information in the head" activated (see also Pouw et al., in press; not in this disseratation). Alternatively, a non-trivial account of the cognitive function of gestures would consist of the specification of how gestures-as-bodily-acts offer a particular type of information that cannot be (efficiently) generated (yet) by internal cognitive means.

We tried to offer a preliminary sketch of such a non-trivial account of the function of gesture by assessing how gestures are used in relation to internal cognitive predispositions of the gesturer. Namely, those with relatively lower working memory capacity tend to produce more co-speech gestures (Chu et al., 2015; Gillespie et al., 2014); when cognitive load is increased problem solvers tend to produce more gestures (Smithson & Nicoladis, 2014); and gestures become more effective for problem solving when cognitive load is increased (Marstaller & Buriánova, 2013; see chapter 5 for further evidence). As such we put forth the perspective in which gestures-as-bodily-acts should offer support through creating spatial bodily imagery that is mediated by visual and proprioceptive feedback. From this we predicted that some of the principles that were put forward in chapter 5 could be applied to the phenomenon of co-thought gestures. Co-thought gestures are hand-movements that are often very alike co-speech gestures (e.g., pointing gestures, actional gestures) but occur, in contrast, in the absence of a communicative intention and without speech (e.g., Chu & Kita, 2016). Co-thought gestures have been observed in several problem solving contexts (e.g., solving gear problems, learning routes from a map), but the most prominent work in this domain was conducted in the context of mental rotations

(Chu & Kita, 2008, 2011, 2016; see chapter 4, 5, 6, and 9 for a discussion). This work has shown that co-thought gestures spontaneously emerge during mental rotation tasks, and benefit performance on those tasks.

In **Chapter 6,** this potential role of co-thought gestures was assessed in two novel problem-solving contexts, namely, the Tower of Hanoi and Chess problems. An essential prediction obtained from Chapter 5, and applied in Chapter 6, is that gestures should be more likely to be spontaneously used during mental problem solving, and more likely to positively affect subsequent physical problem-solving performance when the task is more difficult and when visual-spatial working memory capacity is low. In this experiment, spontaneous pointing gestures (but not iconic gestures) were solicited by the two problem solving tasks. Additional to soliciting spontaneous co-thought gestures, we encouraged some of the participants to adopt pointing gestures during the tasks. The Tower of Hanoi task had a mental problem-solving phase before a physical solving phase, and participants completed both a 3-disc and 4-disc (i.e., a more easy and more difficult) version of the task. In this mental problem-solving phase participants were asked to mentally solve the Tower of Hanoi for 2.5 minutes while the physical Tower was in sight in starting position. Immediately aftrewards, they had to physically solve it. A similar mental-problem solving phase followed by a physical solving phase was adopted in the chess task, but this time with 3 trials increasing in difficulty. Importantly, participants were randomly assigned to a no instruction or a gesture instruction condition, and this intervention was applied during the mental solving phases. Those participants in the no instruction condition were allowed to gesture, and therefore could be naturally subdivided in two groups: participants who spontaneously adopted gestures during mental solving of a trial (spontaneous gesture group), and those who did not gesture (no gesture group). Their performance could be compared with participants in the instructed gesture condition (instructed gesture group). Importantly, since we were interested in the question whether gesturing would be especially helpful when visual-spatial cognitive resources are limited (particularly when the task is more difficult), we obtained scores on a visual working memory and a spatial working memory task at the beginning of the study.

The results obtained in Chapter 6 were mixed. First, in contrast to studies that have shown that the likelihood and the rate at which participants spontaneously use gestures is negatively related to working memory capacity (e.g., Chu et al., 2015; Gillespie et al., 2014), we did not find such a relationship. That is, although participants were more likely to spontaneously gesture and to perform more spontaneous gestures during the more complex trials of the Tower of Hanoi and Chess tasks, neither the occurance nor the frequency of spontaneous gestures correlated with the visual or spatial working memory scores obtained prior to the experiment. We did, however, find a performance benefit of gesturing versus not gesturing as a function of working memory capacity. More precisely, on the more complex 4-disc trial of the Tower of Hanoi it was found that only participants with a lower visual working memory capacity who gestured (either when spontaneously adopted or when instructed to do so) performed better than those who did not gesture during the mental problem-solving phase. On the more difficult chess task trials (requiring 4 and 5 solving steps) it was found that participants who were instructed to gesture performed better than those who did not gesture, and these effects were further qualified by an interaction with visual working memory capacity. In the most complex task we again found that those with a lower visual working memory capacity were more likely to benefit from gesture (when instructed) as compared to non-gesturers. However, in the second-most difficult trial of the chess task this relationship was reversed, showing that especially those with a higher visual working memory capacity benefited from gesturing. Interestingly, only visual working memory capacity showed an interaction effect with gesturing on problem-solving performance; no such interaction was found for spatial working memory capacity.

In **Chapter 7** we sought to determine whether the findings obtained with the Tower of Hanoi task with adults in chapter 6, would also extend to a younger age sample (8-12 yrs old). Specifically we assessed whether children with a relatively lower visual working memory capacity would benefit most from gesturing (either when spontaneously adopted or when instructed to do so) during mental problem solving as compared to children with a lower visual working memory capacity who did not gesture. We reasoned that children might be even more likely than adults to gain in performance when gesturing, because their working memory functions are still in development. However, although we found that children naturally adopted gestures during mental problem solving at similar rates as the adults did in the study reported in chapter 6, we did not find any benefit of gesturing vs. not gesturing on their subsequent problem-solving performance. This suggests that co-thought gestures are already spontaneosly used in problem solving contexts by children, but that effects of co-thought gestures may be limited to supporting the current thought processes in children, whereas in adults it seems to have benefits also for the cognitive representation of the task, showing in subsequent performance.

In **Chapter 8** we reported a small-scale study (17 participants, within-subjects design) that aimed to provide additional insight into the functional mechanism of co-thought gestures during mental solving of the Tower of Hanoi. From previous research (e.g., Kirsh, 2009; Patsenko & Altmann, 2010; Spivey & Dale, 2011) it is known that problem solvers who mentally think through problem solving steps do this by visually exploiting a presentation of the task (when such a visual presentation is provided). In the context of the Tower of Hanoi, this manifests itself by eye movements that are being made over the three pegs that seem to reflect the mental problem solving process of placing discs from one peg to another (Patsenko & Altmann, 2010). Interestingly, the same can be said about gestures. The pointing gestures trajecting over the different pegs during mental problem solving of the Tower of Hanoi seem to reflect the mental solving steps that are being made from mentally placing a disc from one peg to another. As such, eye movements and gestures made during mental planning seem to perform the same function, albeit with different means. As foreshadowed by Cappucio and colleagues (2013), it could be that gestures can come to stand in for visual processing demands associated with having to keep track of mental solving steps through eye-movement redirection. This might directly relate to the findings reported in Chapter 6, that those with a relatively lower visual working memory capacity (arguably resulting in higher visual processing load during mental problem solving) were likely to benefit from gesturing. As such, congenial to the theoretical suggestions made in Chapter 5, gestures may offer the gesturer a way to keep track of mental solving steps in physical space, monitored through proprioception (the sense of the body/hands in space). As such we predicted that whether participants gesture or not should affect the degree to which they exploit a visual presentation of the task, as measured by fixations and saccades during mental problem solving directed at the task representation. Indeed, we found that when participants were instructed to gesture (i.e., point) during mental problem solving, they reduced their eye movements compared to when they did not gesture, and this effect was moderated by visual working memory capacity. Congenial to our findings in Chapter 6, especially those with a lower visual working memory capacity were likely to reduce visual exploitation of the task (i.e., reduced fixation and saccades) during mental problem solving when they were instructed to gesture compared to when they were instructed not to gesture. No particular benefits of gesturing during mental problem solving for actual problem-solving were found, but –next to low power of the study when it comes to performance data- this is likely due to the fact that the mental problem-solving phase lasted only 60 s. in this study (as opposed to 150 s. in Chapter 6).

Part 2: Brief conclusion and suggestions future research

One of the implications of the studies presented in Part 2 is that we should distinguish between different degrees of functional relevance of the body for cognition, as it applies to gesture. I have argued, that it is often the case that the activation of a motor-representation is posited as the causal agent of some benefit in cognitive performance, without explication of the functional role of the actual bodily movement itself. The cognitive impotency of a gesture-as-a-bodily-act on such accounts is apparent, as a neural motor-assembly (i.e., motor-representation, motor code, etc.) must be the causal agent of the bodily movement itself, and not viceversa. Thus, although a gesture might always follow an acitivation of a motor-representation, and although a motor-representation is activated because the agent intends to gesture, it can still be that the gesture as such is not adding much to the cognitive process (e.g., improved problem solving). Importantly, whenever we do not posit cognitively potent consequences of gesture-as-bodily-acts when explaining some cognitive effect of gesturing, we are forced to accept that when the intention to gesture is initated by someone who has non-congenital phantom limbs, the described effect of "gesture" must come about in virtue of activating cognitive precursors of gesture. To me, this is an absurd consequence of accepting that activation of motor-representations explains the cognitive function of gesture, although it might of course be very well true.

Admittedly, it is much harder to empirically investigate the juxtaposition that I have cultivated in Part 2 than to theoretically motivate it. Yet, I do think that understanding gestures as producing novel information not already present within the cognitive system does lead to particular research questions that can be empirically addressed. The particular research questions that I have investigated concerned the relation between internal cognitive capacities of the problem solver (working memory system) and the effect gestures have on the problem-solving process and problem-solving performance. Indeed, the findings with adults seem to indicate that gestures may compensate for detriments in performance that normally occur under conditions of limited cognitive resources.

Further research on the cognitive function of gesture might specifically focus on the understudied phenomenon of co-thought gestures (e.g., Chu & Kita, 2015), as investigating these gestures is promising for pinpointing cognitively potent consequences of gestures while excluding communicative functions of gesture. Furthermore, the findings that children 8-12yrs already naturally employ co-thought gestures at similar rates as adults when mentally problem solving further invites research on the cognitive function of those gestures for subsequent physical problem solving (which we failed to detect). Additionally, as shown in Chapter 8, the study of gaze-behavior when problem solvers gesture or not might further help to pinpoint what type of cognitive feats gestures accomplish. As of present I know of no studies that have systematically investigated gaze-coordination during gesturing.

Moving forward

In the final chapter reported in **Part 3** of this dissertation (**Chapter 9**) a preliminary sketch is provided of how we might understand gestures as emerging from action-readiness states (i.e., motor simulations) while maintaining that gestures-as-bodily acts have a non-trivial cognitive function. Such an account is needed because there is evidence that gestures spontaneously emerge in so far as the concepts that are thought about invite concrete actions. For example, mental rotation gestures are less likely to be produced when the objects-to-be-rotated are perceived as less manipulable (Chu & Kita, 2015). Furtheremore, co-speech gestures are more likely to be produced when the concepts that are talked about are manipulable and less likely when they are not (Masson-Carro et al., 2015). By review of a recent grand theory in psychology, called the Predictive Processing Perspective (Clark, 2015; Howhy, 2012), it is suggested that gestures might indeed arise out of action-readiness states, as predicted by embodied simulation theories (Hostetter & Alibali, 2008). However, these action-readiness states take the form kinesthetic predictions that are constantly produced when talking and thinking about manipulative actions. These predictions are imperfect and need to be constantly negotiated (i.e., updated) with information in the world. As such, predictions are put to the test when actuated in gesture, which offers a way to calibrate the kinesthetic predictions from which gesture arises on the basis of the actual kinesthethic information that gestures-as-bodily-acts produce.

References

Adams, F., & Aizawa, K. (2001). The bounds of cognition. *Philosophical Psychology*, *14*, 43-64.

Alibali, M. W. (2005). Gesture in spatial cognition: Expressing, communicating, and thinking about spatial information. *Spatial Cognition and Computation*, *5*, 307-331.

Alibali, M. W., & DiRusso, A. A. (1999). The function of gesture in learning to count: More than keeping track. *Cognitive Development*, *14*, 37-56.

Alibali, M. W., Heath, D. C., & Myers, H. J. (2001). Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen. *Journal of Memory and Language*, *44*, 169-188.

Alibali, M. W., Spencer, R. C., & Kita, S. (2004). *Spontaneous gestures influence strategy choice in problem solving*. Paper presented at the Biennial Meeting of the Society for Research in Child Development, Atlanta, Georgia.

Alibali, M. W., Spencer, R. C., Knox, L., & Kita, S. (2011). Spontaneous gestures influence strategy choices in problem solving. *Psychological Science*, *22*(9), 1138-1144.

Alibali, M., & Goldin-Meadow, S. (1993). Gesture-speech mismatch and mechanisms of learning: What the hands reveal about a child's state of mind. *Cognitive Psychology, 25,* 468-523.

Allen, G. L. (2003). Gestures accompanying verbal route directions: Do they point to a new avenue for examining spatial representations? *Spatial Cognition and Computation*, *3*, 259-268.

Amorim, M.-A., Isableu, B., & Jarraya, M. (2006). Embodied spatial transformations: "Body analogy" for the mental rotation of objects. *Journal of Experimental Psychology: General*, *135*(3), 327-347.

Anderson, M. L. (2008). On the grounds of (X)-grounded cognition. In P. Calvo & T. Gomila (Eds.), *Handbook of Cognitive Science: An Embodied Approach* (pp. 423-435). New York: Elsevier.

Anderson, M. L., Richardson, M. J., & Chemero, A. (2012). Eroding the boundaries of cognition: Implications of embodiment. *Topics in Cognitive Science*, *4*(4), 717-730.

Andres, M., Seron, X., & Oliver, E. (2007). Contribution of hand motor circuits to counting. *Journal of Cognitive Neuroscience*, *19*, 563-576.

Antle, A. N. (2012). Exploring how children use their hands to think: An embodied interactional analysis. *Behaviour & Information Technology*, *32*, 938-954.

Antle, A. N., Droumeva, M., & Ha, D. (2009). Hands on what?: Comparing children's mouse-based and tangible-based interaction. In *Proceedings of the 8th International Conference on Interaction Design and Children* (pp. 80-88).

Antonenko, P., Paas, F., Grabner, R., & van Gog, T. (2010). Using electroencephalography to measure cognitive load. *Educational Psychology Review*, *22*(4), 425-438.

Arbib, M. A., Gasser, B., & Barrès, V. (2014). Language is handy but is it embodied? *Neuropsychologia, 55*, 57-75. In Aydede, M. (eds). *Cambridge Handbook of Situated Cognition*. Cambridge: Cambridge University Press.

Ayres, P., & Paas, F. (2007a). Can the cognitive-load approach make instructional animations more effective? *Applied Cognitive Psychology, 21*, 811-820.

Ayres, P., & Paas, F. (2007b). Making instructional animations more effective: A cognitive load approach. *Applied Cognitive Psychology, 21*, 695-700.

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, *4*, 829-839.

Ball, D. L. (1992). Magical hopes: Manipulatives and the reform of math education. *American Educator, 16*(2), 14-18.

Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience, 7*(1), 66-80.

Barrós-Loscertales, A., González, J., Pulvermüller, F., VenturaCampos, N., Bustamante, J. C., Costumero, V., Parcet, M. A., & Ávila, C. (2011). Reading "salt" activates gustatory brain regions: fMRI evidence for semantic grounding in a novel sensory modality. *Cerebral Cortex, 22*, 2554-2563.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*(4), 577-660.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59,* 617-645.

Bates, E., Camaioni, L., & Volterra, V. (1975). The acquisition of performatives prior to speech. *Merrill-Palmer Quarterly*, 21, 205-226.

R

Beilock, S. L. & Goldin-Meadow, S. (2010). Gesture changes thought by grounding it in action. *Psychological Science*, *21*, 1605-1610.

Bivall, P., Ainsworth, S., & Tibell, L. A. (2011). Do haptic representations help complex molecular learning? *Science Education*, *95*(4), 700-719.

Black, J. B. (2011). Embodied cognition and learning environment design. in D. Jonassen, & S. Land (Eds.), *Theoretical foundations of student-centered learning environments*, eds. (pp. 128-223). New York: Routledge.

Boncoddo, R., Dixon, J. A., & Kelley, E. (2010). The emergence of a novel representation from action: Evidence from preschoolers. *Developmental Science*, *13*, 370-377.

Borst, J. P., Buwalda, T. A., van Rijn, H., & Taatgen, N. A. (2013). Avoiding the problem state bottleneck by strategic use of the environment. *Acta Psychologica*, *144*(2), 373-379.

Boysen, S. T., Berntson, G. G., Shreyer, T. A., & Hannan, M. B. (1995). Indicating acts during counting by a chimpanzee (Pan troglodytes). *Journal of Comparative Psychology*, *109*, 47-51.

Brandt, S. A., & Stark, L. W. (1997). Spontaneous eye movements during visual imagery reflect the content of the visual scene. *Journal of Cognitive Neuroscience*, *9*(1), 27-38.

Bredo, E. (1994). Reconstructing educational psychology: Situated cognition and Deweyian pragmatism. *Educational Psychologist*, *29*(1), 23-35.

Bremner, A. J., Cowie, D. (2013). Developmental origins of the hand in the mind, and the role of the hand in the development of the mind in Z. Radman (Ed.) *The hand, an organ of the mind: What the manual tells the mental*, (pp. 255-268). Cambridge, MA: MIT Press.

Brooks, N. B. (2014). *Movements for doing, thinking, and speaking: The relationship between motor experience and gesture in novice mental abacus users* (unpublished doctoral dissertation). The University of Chicago, Chicago.

Brooks, R. (1991a). Intelligence without representation. *Artificial Intelligence Journal*, 47, 139-160.

Brown, M. C., McNeil, N. M., & Glenberg, A. M. (2009). Using concreteness in education: Real problems, potential solutions. *Child Development Perspectives 3*(3), 160-164.

Brucker, B., Ehlis, A. C., Häußinger, F. B., Fallgatter, A. J., & Gerjets, P. (2015). Watching corresponding gestures facilitates learning with animations by activating human mirror-neurons: An fNIRS study. *Learning and Instruction*, *36*, 27-37. 10.1016/j.learninstruc.2014.11.003

Calvo, P., & Gomila, T. (2008). *Handbook of Cognitive Science: An Embodied Approach*. San Diego: Elsevier.

Cappuccio, M. L., Chu, M., & Kita, S. (2013). Pointing as an instrumental gesture: Gaze representation through indication. *Humana Mente: Journal of Philosophical Studies, 24*, 125-149.

Carlson, R. A., Avraamides, M. N., Cary, M., & Strasberg, S. (2007). What do the hands externalize in simple arithmetic? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(4), 747-756.

Cary, M., & Carlson, R. A. (1999). External support and the development of problem-solving routines. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(4), 1053-1070.

Chandler, P. (2009). Dynamic visualisations and hypermedia: Beyond the "wow" factor. *Computers in Human Behavior*, *25*(2), 389-392.

Chase, W. G., & Ericsson, K. A. (1982). Skill and working memory. *The Psychology of Learning and Motivation*, *16*, 1-58.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*, 55-81.

Chawla, P., & Krauss, R. M. (1994). Gesture and speech in spontaneous and rehearsed narratives. *Journal of Experimental Social Psychology*, *30*, 580-601.

Chemero, A. (2009). *Radical embodied cognitive science*. Massachusetts: The MIT Press.

Chu, M., & Kita, S. (2008). Spontaneous gestures during mental rotation tasks: Insights into the microdevelopment of the motor strategy. *Journal of Experimental Psychology: General*, *137*, 706-723.

Chu, M., & Kita, S. (2011). The nature of gestures' beneficial role in spatial problem solving. *Journal of Experimental Psychology: General*, *140*(1), 102-115.

Chu, M., & Kita, S. (2016). Co-thought and co-speech gestures are generated by the same action generation process. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 24*(2), 257-270.

Chu, M., Meyer, A., Foulkes, L., & Kita, S. (2013). Individual differences in frequency and salience of speech-accompanying gestures: The role of cognitive abilities and empathy. *Journal of Experimental Psychology: General, 143*, 694-709.

Church, R. B., & Goldin-Meadow, S. (1986). The mismatch between gesture and speech as an index of transitional knowledge. *Cognition*, *23*(1), 43-71.

Clark, A. (2005). Beyond the flesh: Some lessons from a mole cricket. *Artificial Life*, *11*(2), 233-244.

Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension.* New York: Oxford University Press.

Clark, A. (2013a). "Gesture as thought", in *The hand, an organ of the mind: What the manual tells the mental*, ed. Z. Radman (pp. 255-268). Cambridge, MA: MIT Press.

Clark, A. (2015a). Radical Predictive Processing. *The Southern Journal of Philosophy*, *53*(1), 3-27.

Clark, A. (2015b). *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford: Oxford University Press.

Clark, A., & Chalmers. D. (1998). The extended mind. *Analysis, 58*(1), 7-19.

Clark, A., (2013b). Whatever Next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(03), 181-204.

Clements, D. H. (2000). 'Concrete' Manipulatives, Concrete Ideas. *Contemporary Issues in Early Childhood*, *1*(1)*,* 45-60.

Cook, S. W., & Tanenhaus, M. K. (2009). Embodied communication: Speakers' gestures affect listeners' actions. *Cognition*, *113*, 98-104.

Cook, S. W., Yip, T. K., & Goldin-Meadow, S. (2010). Gesturing makes memories that last. *Journal of Memory and Language*, *63*(4), 465-475.

Cook, S. W., Yip, T. K., & Goldin-Meadow, S. (2012). Gestures, but not meaningless movements, lighten working memory load when explaining math. *Language and Cognitive Processes*, *27*(4), 594-610.

Cooperrider, K., Wakefield, E., & Goldin-Meadow, S (2015). More than meets the eye: Gesture changes thought, even without visual feedback. *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Cutica, I., & Bucciarelli, M. (2013). Cognitive change in learning from text: Gesturing enhances the construction of the text mental model. *Journal of Cognitive Psychology*, *25*(2), 201-209.

Cutica, I., Ianì, F., & Bucciarelli, M. (2014). Learning from text benefits from enactment. *Memory & Cognition*, *42*(7), 1-12.

Daponte, P., De Vito, L., Picariello, F., & Riccio, M. (2014). State of the art and future developments of the augmented Reality for measurement applications. *Measurement*, *57*, 53-70.

Darwin, C. (1998) *The Expression of the emotions in man and animals*. London: Harper Collins.

De Bock, D., Deprez, J., Van Dooren, W., Roelens, M., & Verschaffel, L. (2011). Abstract or concrete examples in learning mathematics? A replication and elaboration of Kaminski, Sloutsky, and Heckler's study. *Journal for Research in Mathematics Education*, *42*(2), 109-126.

De Jaegher, H., & Di Paolo, E. (2007). Participatory sense-making. *Phenomenology and the Cognitive Sciences*, *6*(4), 485-507.

De Jaegher, H., Di Paolo, E., & Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends in Cognitive Sciences*, *14*(10), 444-447.

De Jong, T., Linn, M. C., & Zacharia, Z. C. (2013). Physical and virtual laboratories in science and engineering education. *Science*, *340*(6130), 305-308.

De Koning, B. B., & Tabbers, H. K. (2011). Facilitating understanding of movements in dynamic visualizations: An embodied perspective. *Educational Psychology Review*, *23*(4), 501-521.

De Koning, B. B., & Van der Schoot, M. (2013). Becoming part of the story! Refueling the interest in visualization strategies for reading comprehension. *Educational Psychology Review*, *25*, 261-287.

De Nooijer, J. A., Van Gog, T., Paas, F., & Zwaan, R. A. (2013). Effects of imitating gestures during encoding or during retrieval of novel verbs on children's test performance. *Acta Psychologica*, *144*, 173-179.

De Nooijer, J. A., van Gog, T., Paas, F., & Zwaan, R. A. (2014). Words in action: Using gestures to improve verb learning in primary school children. *Gesture*, *14*(1), 46-69.

Delgado, B., Gómez, J. C., & Sarriá, E. (2011). Pointing gestures as a cognitive tool in young children: Experimental evidence. *Journal of Experimental Child Psychology*, *110*(3), 299-312.

Delgado, B., Gomez, J.C., & Sarria, E. (2009). Private pointing and private speech: developing parallelisms. In A. Winsler, C. Fernyhough & I. Montero (Eds.), *Private speech, executive function and the development of verbal self-regulation* (pp. 153-162). Cambridge: Cambridge University Press.

Della Sala, S., Gray, C., Baddeley, A., & Wilson, L. (1997). *The Visual Patterns Test: A new test of short-term visual recall*. Bury St. Edmunds, England: Thames Valley Test.

DeLoache, J. S. (1987). Rapid change in the symbolic functioning of very young children. *Science*, *238*(4833), 1556-1557.

DeLoache, J. S. (1991). Symbolic functioning in very young children: Understanding of pictures and models. *Child Development*, *62*(4), 736-752.

DeLoache, J. S. (2000). Dual representation and young children's use of scale models. *Child Development*, *71*(2), 329-338.

DeLoache, J. S. (2004). Becoming symbol-minded. *Trends in Cognitive Sciences*, *8*(2), 66-70.

Den Ouden, H. E., Kok, P., & De Lange, F. P. (2012). How prediction errors shape perception, attention, and motivation. *Frontiers in psychology*, *3*, 1-12.

Dienes, Z. P. (1973). *The Six Stages in the Process of Learning Mathematics*. Slough: National

Dijkstra, K., Eerland, A., Zijlmans, J., & Post, L. S. (2014). Embodied cognition, abstract concepts, and the benefits of new technology for implicit body manipulation. *Frontiers in Psychology*, *5*, 1-8.

Dixon, J. A., & Dohn, M. C. (2003). Redescription disembeds relations: Evidence from relational transfer and use in problem solving. *Memory & Cognition*, *31*, 1082-1093.

Dourish, P. (2004). *Where the action is: The foundations of embodied interaction*. Massachusetts: MIT press.

Dove, G. (2010). On the need for embodied and dis-embodied cognition. *Frontiers in Psychology*, *1*, 129-141.

Driver, J., & Spence, C. (1998). Attention and the crossmodal construction of space. *Trends in Cognitive Sciences*, *2*(7), 254-262.

Droll, J. A., & Hayhoe, M. M. (2007). Trade-offs between gaze and working memory use. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(6), 1352-1365.

Duit, R., & Treagust, D. F. (2012). How can conceptual change contribute to theory and practice in science education? In B. J. Fraser, K. Tobin & C. J. McRobbie (Eds.), *Second International Handbook of Science Education* (pp. 107-118). Dordrecht, The Netherlands: Springer.

Dunn, T. L., & Risko, E. F. (2015). Toward a metacognitive account of cognitive offloading. *Cognitive Science*, *40*, 1080-1127.

Finkelstein, N. D., Adams, W. K., Keller, C. J., Kohl, P. B., Perkins, K. K., Podolefsky, N. S., Reid, S., & LeMaster, R. (2005). When learning about the real world is better done virtually: A study of substituting computer simulations for laboratory equipment. *Physical Review Special Topics-Physics Education Research*, *1*(1), 1-8.

Flanagan, R. (2013). Effects of learning from interaction with physical or mediated devices. *Cognitive Processing*, *14*, 213-215.

Flusberg, S. J., & Boroditsky, L. (2011). Are things that are hard to physically move also hard to imagine moving? *Psychonomic Bulletin & Review*, *18*(1), 158-164.

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

Frank, M. C., & Barner, D. (2012). Representing exact number visually using mental abacus. *Journal of Experimental Psychology: General*, *141*(1), 134-149.

Franklin, D. W., and Wolpert, D. M. (2011). Computational mechanisms of sensorimotor control. *Neuron*, *72*, 425-442.

Friston, K. (2009). The free-energy principle: a rough guide to the brain? *Trends in cognitive sciences*, *13*(7), 293-301.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127-138.

Friston, K., Thornton, C., & Clark, A. (2012). Free-energy minimization and the dark-room problem. *Frontiers in Psychology*, 1-7.

Fu, W. T. (2011). A dynamic context model of interactive behavior. *Cognitive Science*, *35*(5), 874-904.

Fyfe, E. R., McNeil, N. M., Son, J. Y., & Goldstone, R. L. (2014). Concreteness fading in mathematics and science instruction: A systematic review. *Educational Psychology Review*, *26*(1), 9-25.

Galati, A., Weisberg, S., Newcombe, N., & Avraamides, M.N. (2015). Individual differences in spatial ability influence the effect of gesturing on navigation and spatial memory. In G. Ferré, & M. Tutton (Eds.) *Proceedings of Gesture and Speech in Interaction - 4th edition (GESPIN 4)* (pp. 119-124).

Gallagher, S. (2009). Philosophical antecedents to situated cognition. In Robbins, P. and Aydede, M. (eds). *Cambridge Handbook of Situated Cognition* (35-51). Cambridge: Cambridge University Press.

Garber, P. R., & Goldin-Meadow, S. (2002). Gesture offers insight into problem-solving in adults and children. *Cognitive Science, 26*, 817-831.

Garber, P. R., Alibali, M. W., & Goldin-Meadow, S. (1998). Knowledge conveyed in gesture is not tied to the hands. *Child Development, 69*, 75-84.

Gaschler, R., Vaterrodt, B., Frensch, P. A., Eichler, A., & Haider, H. (2013). Spontaneous usage of different shortcuts based on the commutativity principle. *PloS One*, *8*(9), 1-13.

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology*, *40*(2), 177-190.

Gentilucci, M., Benuzzi, F., Gangitano, M., & Grimaldi, S. (2001). Grasp with hand and mouth: a kinematic study on healthy subjects. *Journal of Neurophysiology*, *86*(4), 1685-1699.

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton-Mifflin.

Gillespie, M., James, A. N., Federmeier, K. D., & Watson, D. G. (2014). Verbal working memory predicts co-speech gesture: Evidence from individual differences. *Cognition*, *132*(2), 174-180.

Glenberg, A. M. 2008. Embodiment for education. In P. Calvo & T. Gomila (Eds.), *Handbook of Cognitive Science: An Embodied Approach* (pp. 355-372). New York: Elsevier.

Glenberg, A. M., & Gallese, V. (2012). Action-based language: A theory of language acquisition, comprehension, and production. *Cortex*, *48*(7), 905-922.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, *9*(3), 558-565.

Glenberg, A. M., Goldberg, A. B., & Zhu, X. (2011). Improving early reading comprehension using embodied CAI. *Instructional Science*, *39*(1), 27-39.

Glenberg, A. M., Gutierrez, T., Levin, J. R., Japuntich, S., & Kaschak, M. P. (2004). Activity and imagined activity can enhance young children's reading comprehension. *Journal of Educational Psychology*, *96*(3), 424-436.

Glenberg, A. M., Havas, D., Becker, R., & Rinck, M. (2005). Grounding language in bodily states. In D. Pecher & R. A. Zwaan, (Eds.) *Grounding cognition: The role of perception and action in memory, language, and thinking*. (pp. 115-128). Cambridge University Press.

Glenberg, A., Willford, J., Gibson, B., Goldberg, A., & Zhu, X. (2011). Improving reading to improve math. *Scientific Studies of Reading*, *16*, 316-340.

Golding-Meadow, S. & Beilock, S. L. (2010). Action's influence on thought: The case of gesture. *Perspectives on Psychological Science, 5*, 664-674.

Goldin-Meadow, S. (1998). The development of gesture and speech as an integrated system. *New Directions for Child and Adolescent Development*, *79*, 29-42.

Goldin-Meadow, S. (2003). *Hearing gesture: How our hands help us think.* Cambridge, MA: Harvard University Press.

Goldin-Meadow, S., & Alibali, M. W. (2002). Looking at the hands through time: A microgenetic perspective on learning and instruction. In N. Granott & J. Parziale (Eds.), *Microdevelopment: Transition processes in development and learning* (pp. 80-105). Cambridge, England: Cambridge University Press.

Goldin-Meadow, S., & Alibali, M. W. (2012). Gesture's role in learning and development. In P. Zelazo (ed.), *Oxford Handbook of Developmental Psychology*. (pp.953-973). Oxford: Oxford University Press.

Goldin-Meadow, S., & Alibali, M.W. (2015). Gesture's role in speaking, learning, and creating language. *Annual Review of Psychology*, *123*, 448-453.

Goldin-Meadow, S., & Beilock, S. L. (2010). Action's influence on thought: The case of gesture. *Perspectives on Psychological Science*, *5*(6), 664-674.

Goldin-Meadow, S., & Momeni-Sandhofer, C. (1999). Gestures convey substantive information about a child's thoughts to ordinary listeners. *Developmental Science*, *2*, 67-74.

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining math: Gesticulation lightens the load. *Psychological Science*, *12*, 516-522.

Goldstone, R. L, & Sakamoto, Y. (2003). The transfer of abstract principles governing complex adaptive systems. *Cognitive Psychology*, *46*(4), 414-466.

Goldstone, R. L., & Barsalou, L. W. (1998). Reuniting perception and conception. *Cognition*, *65*(2-3), 231-262.

Goldstone, R. L., & Son, J. Y. (2005). The transfer of scientific principles using concrete and idealized simulations. *The Journal of the Learning Sciences*, *14*(1), 69-110.

Gonzalez J., Barros-Loscertales A., Pulvermuller F., Meseguer V., Sanjuán, A. (2006). Reading cinnamon activates olfactory brain regions. *Neuroimage*, *32*, 906-912.

Graham, J. A., & Argyle, M. (1975). A cross-cultural study of the communication of extra-verbal meaning by gesture. *International Journal of Psychology*, *10*, 57-67.

Gray, W. D, & Fu, W. (2004). Soft constraints in interactive behavior: The case of ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head. *Cognitive Science* *28*(3), 359-382.

Gray, W. D., Sims, C. R., Fu, W-T, & Schoelles, M. J. (2006). The soft constraints hypothesis: A rational analysis approach to resource allocation for interactive behavior. *Psychological Review*, *113*, 461-482.

Grush, R. (2004). The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and Brain Sciences*, *27*(03), 377-396.

Han, I., & Black, J. B. (2011). Incorporating haptic feedback in simulation for learning physics. *Computers & Education*, *57*(4), 2281-2290.

Haselen, G. L. V., van der Steen, J., & Frens, M. A. (2000). Copying strategies for patterns by children and adults. *Perceptual and Motor Skills*, *91*(2), 603-615.

Hatano, G., & Osawa, K. (1983). Digit memory of grand experts in abacus-derived mental calculation. *Cognition*, *15*(1), 95-110.

Hatano, G., Miyake, Y., & Binks, M. G. (1977). Performance of expert abacus operators. *Cognition*, *5*, 47-55.

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, *41*, 301-307.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York: Guilford Press.

Hayhoe, M. M., Pook, P. K., & Rao, R. P. N. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, *20*, 723-767.

Healy, & R. W. Proctor (Vol. Eds.), *Handbook of Psychology: Vol. 4. Experimental Psychology* (pp. 147-176). New York: Wiley.

Hegarty, M., Kriz, S., & Cate, C. (2003). The role of mental animations and external animations in understanding mechanical systems. *Cognition and Instruction, 21*, 325-360.

Hegarty, M., Mayer, S., Kriz, S., & Keehner, M. (2005). The role of gestures in mental animation. *Spatial Cognition and Computation*, *5*, 333-356.

Hoetjes, M., Krahmer, E., & Swerts, M. (2014). Does our speech change when we cannot gesture? *Speech Communication*, *57,* 257-267.

Höffler, T. N., & Leutner, D. (2007). Instructional animation versus static pictures: A meta-analysis. *Learning and Instruction*, *17*(6), 722-738.

Hohwy, J. (2013). *The Predictive Mind*. Oxford: Oxford University Press.

Holler, J., & Stevens, R. (2007). The effect of common ground on how speakers use gesture and speech to represent size information. *Journal of Language and Social Psychology*, *26*, 4-27.

Hostetter, A. B. (2011). When do gestures communicate? A meta-analysis. *Psychological Bulletin, 137*, 297-315.

Hostetter, A. B. (2014). Action attenuates the effect of visibility on gesture rates. *Cognitive Science*, 1-14.

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review*, *15*(3), 495-514.

Hostetter, A. B., & Alibali, M. W. (2010). Language, gesture, action! A test of the Gesture as Simulated Action framework. *Journal of Memory and Language, 63,* 245-253.

Hostetter, A. B., Alibali, M. W., & Bartholomew, A. E. (2011). Gesture during mental rotation. In *Proceedings of the 33rd Annual Conference of Cognitive Science Society. Cognitive Science Society, Austin, TX* (pp. 1448-1453).

Hostetter, A. B., Alibali, M. W., & Bartholomew, A. E. (2011). Gesture during mental rotation. In *Proceedings of the 33rd Annual Conference of Cognitive Science Society* (pp. 1448-1453). Austin, TX: Cognitive Science Society.

Hostetter, A. B., Alibali, M. W., & Kita, S. (2007). I see it in my hands' eye: Representational gestures reflect conceptual demands. *Language and Cognitive Processes, 22*, 313-336.

Hurley, S. L. (2002). *Consciousness in action*. Harvard University Press

Hutchins, E. (1995a). *Cognition in the wild*. Cambridge: MIT press.

Hutchins, E. (1995b). How a cockpit remembers its speeds. *Cognitive Science*, *19*, 265-288.

Hutchins, E. (2005). Material anchors for conceptual blends. *Journal of Pragmatics*, *37*(10), 1555-1577.

Hutchins, E., & Nomura, S. (2011). Collaborative construction of multimodal utterances. In J. Streeck, C. Goodwin, & C. Lebaron (Eds.), *Embodied interaction: Language and body in the material world* (pp. 29-43). Cambridge: Cambridge University Press.

Iverson, J. M. (1998). Gesture when there is no visual model. *New Directions for Child Development*, *79*, 89-100.

Iverson, J. M., & Thelen, E. (1999). Hand, mouth, and brain: The dynamic emergence of speech and gesture. *Journal of Consciousness Studies*, *6*(11), 19-40.

Johansson, R., Holsanova, J., & Holmqvist, K. (2006). Pictures and spoken descriptions elicit similar eye movements during mental imagery, both in light and in complete darkness. *Cognitive Science*, *30*(6), 1053-1079.

Johnson-Glenberg, M. C., Birchfield, D. A., Tolentino, L. & Koziupa, T. (2014). Collaborative embodied learning in mixed reality motion-capture environments: Two science studies. *Journal of Educational Psychology*, *106*(1), 86-104.

Jones, M. G., Minogue, J., Tretter, T. R., Negishi, A., & Taylor, R. (2006). Haptic augmentation of science instruction: Does touch matter? *Science Education*, *90*(1), 111-123.

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review*, *19*(4), 509-539.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, *38*(1), 23-31.

Kaminski, J. A, Sloutsky, V. M., & Heckler, A. F. (2009). Transfer of mathematical knowledge: The portability of generic instantiations. *Child Development Perspectives*, *3*(3), 151-155.

Kaminski, J. A., & Sloutsky, V. M. (2009). The effect of concreteness on children's ability to detect common proportion. In *Proceedings of the XXXI Annual Conference of the Cognitive Science Society* (pp. 335-340).

Kaminski, J. A., Sloutsky, V. M., & Heckler, A. F. (2008). Learning theory: The advantage of abstract examples in learning math. *Science*, *320*(5875), 454-455.

Kaminski, J. A., Sloutsky, V. M., & Heckler, A. F. (2009). The devil is in the superficial details: Why generic instantiations promote portable mathematical knowledge. *Child Development Perspectives*, *3*, 151-155.

Kaminski, J. A., Sloutsky, V. M., & Heckler, A. F. (2013). The cost of concreteness: The effect of nonessential information on analogical transfer. *Journal of Experimental Psychology: Applied*, *19*(1), 14-29.

Kaschak, M. P., Zwaan, R. A., Aveyard, M., & Yaxley, R. H. (2006). Perception of auditory motion affects language processing. *Cognitive Science*, *30*, 733-744.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*, 773-795.

Kastens, K. A., Liben, L. S., & Agrawal, S. (2008). Epistemic actions in science education. In C. Freksa, N. S. Newcombe, P. Gardenfors, & S. W. Wölfl (Eds.), *Proceedings of the International Conference on Spatial Cognition VI: Learning, Reasoning, and Talking About Space* (pp. 205-215). Heidelberg: Springer-Verlag.

Kelly, S. D. (2001). Broadening the units of analysis in communication: Speech and nonverbal behaviours in pragmatic comprehension. *Journal of Child Language*, *28*, 325-349.

Kelly, S. D., Barr, D. J., Church, R. B., & Lynch, K. (1999). Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and memory. *Journal of Memory and Language*, *40*, 577-592.

Kelly, S. D., McDevitt, T., & Esch, M. (2009). Brief training with co-speech gesture lends a hand to word learning in a foreign language. *Language and Cognitive Processes*, *24*(2), 313-334.

Kelly, S., Byrne, K., & Holler, J. (2011). Raising the ante of communication: Evidence for enhanced gesture use in high stakes situations. *Information*, *2*(4), 579-593.

Kendon, A. (1994). Do gestures communicate? A review. *Research on Language and Social Interaction*, *27*, 175-200.

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

Kiefer, M., & Trump, N. M. (2012). Embodiment theory and education: The foundations of cognition in perception and action. *Trends in Neuroscience and Education*, *1*, 15-20.

Kirsh, D. (1995). Complementary strategies: Why we use our hands when we think. *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society* (pp. 212-217).

Kirsh, D. (1995). The intelligent use of space. *Artificial Intelligence*, *73*(1), 31-68.

Kirsh, D. (2009). Projection, problem space and anchoring. In *Proceedings of the 31st annual conference of the cognitive science society* (pp. 2310 - 2315). Hillsdale: Lawrence Erlbaum Associates.

Kirsh, D. (2010). Thinking with external representations. *AI & Society*, *25*(4), 441-454.

Kirsh, D., & Maglio, P. (1994). On distinguishing epistemic from pragmatic action. *Cognitive Science*, *18*, 513-549.

Kita, S. (2000). How representational gestures help speaking. In D. McNeill (Ed.), *Language and gesture* (pp. 162 -185). Cambridge, UK: Cambridge University Press.

Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and language*, *48*, 16-32.

Klahr, D., L. Triona, M., & Williams, C. (2007). Hands on what? The relative effectiveness of physical versus virtual materials in an engineering design project by middle school children. *Journal of Research in Science Teaching*, *44*(1), 183-203.

Kontra, C., Lyons, D. J., Fischer, S. M., & Beilock, S. L. (2015). Physical experience enhances science learning. *Psychological Science*, *26*(6), 737-749.

Krauss R. M., Chen Y., & Gottesmann R. F. (2000) Lexical gestures and lexical access: A process model, in D. McNeill. (*Ed.*) *Language and Gesture*, Cambridge University Press, Cambridge.

Krauss, R. M. (1998). Why do we gesture when we speak? *Current Directions in Psychological Science*, *7*, 54-60.

Krauss, R. M., & Hadar, U. (1999). The role of speech-related arm/hand gestures in word retrieval. *Gesture, Speech, and Sign*, 93-116.

Laeng, B., & Teodorescu, D. S. (2002). Eye scanpaths during visual imagery reenact those of perception of the same visual scene. *Cognitive Science*, *26*(2), 207-231.

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

Lakoff, G., & Núñez, R. E. (2000). *Where mathematics comes from: How the embodied mind brings mathematics into being*. New York: Basic books.

Landy, D., & Goldstone, R. L. (2007). How abstract is symbolic thought? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 720-733.

Lee, I. A., & Preacher, K. J. (2013, September). Calculation for the test of the difference between two dependent correlations with one variable in common [Computer software]. Available from http://quantpsy.org.

Lindgren, R., & Johnson-Glenberg, M. (2013). Emboldened by embodiment: Six precepts for research on embodied learning and mixed reality. *Educational Researcher*, *42*(8), 445-452.

Logan, T., Lowrie, T., & Diezmann, C. M. (2014). Co-thought gestures: Supporting students to successfully navigate map tasks. *Educational Studies in Mathematics, 87*, 87-102.

Lu, M. T. P. (2011). *The effect of instructional embodiment designs on Chinese language learning: The use of embodied animation for beginning learners of Chinese characters*. Doctoral dissertation, Columbia University.

Lupyan, G., & Clark, A. (2015). Words and the World Predictive Coding and the Language-Perception-Cognition Interface. *Current Directions in Psychological Science*, *24*(4), 279-284.

Macedonia, M., & Klimesch, W. (2014). Long-Term effects of gestures on memory for foreign language words trained in the classroom. *Mind, Brain, and Education*, *8*(2), 74-88.

Macedonia, M., & Knösche, T. R. (2011). Body in mind: How gestures empower foreign language learning. *Mind, Brain, and Education*, *5*(4), 196-211.

Macedonia, M., & von Kriegstein, K. (2012). Gestures enhance foreign language learning. *Biolinguistics*, *6*(3-4), 393-416.

Macken, L., & Ginns, P. (2014). Pointing and tracing gestures may enhance anatomy and physiology learning. *Medical Teacher*, *36*(7), 596-601.

Manches, A. D., & O'Malley, C. (2012). Tangibles for learning: a representational analysis of physical manipulation. *Personal and Ubiquitous Computing*, *16*(4), 405-419.

Manches, A., O'Malley, C., & Benford, S. (2010). The role of physical representations in solving number problems: a comparison of young children's use of physical and virtual materials. *Computers & Education*, *54*(3), 622-640.

Markman, A., & Gentner, D. (1993). Structural alignment during similarity comparisons. *Cognitive Psychology*, *25*, 431-467.

Marley, S. C., & Carbonneau, K. J. (2014). Theoretical perspectives and empirical evidence relevant to classroom instruction with manipulatives. *Educational Psychology Review*, *26*(1), 1-7.

Marley, S. C., Levin, J. R., & Glenberg, A. M. (2007). Improving Native American children's listening comprehension through concrete representations. *Contemporary Educational Psychology*, *32*(3), 537-550.

Marley, S. C., Szabo, Z., Leven, J. R., & Glenberg, A. M. (2011). Investigation of an activity-based text-processing strategy in mixed-age child dyads. *The Journal of Experimental Education*, *79*(3), 340-360.

Marstaller, L., & Burianová, H. (2013). Individual differences in the gesture effect on working memory. *Psychonomic Bulletin & Review, 20*, 496-500.

Martin, A. (2007). The representation of object concepts in the brain. *Annual Review of Psychology*, *58*, 25-45.

Martin, T., & Schwartz, D. L. (2005). Physically distributed learning: Adapting and reinterpreting physical environments in the development of fraction concepts. *Cognitive Science*, *29*(4), 587-625.

Martin, T., Lukong, A., & Reaves, R. (2007). The role of manipulatives in arithmetic and geometry tasks. *Journal of Education and Human Development*, *1*, 1-10.

Masson, M. E. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behavior Research Methods*, *43*, 679-690.

Masson-Carro, I., Goudbeek, M., & Krahmer, E. (2015). Can you handle this? The impact of object affordances on how co-speech gestures are produced. *Language, Cognition and Neuroscience*, *31*(3), 430-440.

McNeil, N. M., & Fyfe, E. R. (2012). "Concreteness fading" promotes transfer of mathematical knowledge. *Learning and Instruction*, *22*, 440-448.

McNeil, N. M., & Jarvin. L. (2007). When theories don't add up: Disentangling he manipulatives debate. *Theory into Practice*, *46*(4), 309-316.

McNeil, N. M., Alibali, M. W., & Evans, J. L. (2000). The role of gesture in children's comprehension of spoken language: Now they need it, now they don't. *Journal of Nonverbal Behavior*, *24*, 131-150.

McNeill, D. (1985). So you think gestures are nonverbal? *Psychological Review*, *92*, 350-371.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago Press.

McNeill, D. (2008). *Gesture and thought*. University of Chicago Press.

Melinger, A., & Kita, S. (2007). Conceptualisation load triggers gesture production. *Language and Cognitive Processes*, *22*(4), 473-500.

Morris, D., Tan, H., Barbagli, F., Chang, T., & Salisbury, K. (2007). Haptic feedback enhances force skill learning. In EuroHaptics Conference, 2007 and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems. *World Haptics 2007*.Second Joint (pp. 21-26).

Morsella, E., & Krauss, R. M. (2004). The role of gestures in spatial working memory and speech. *The American Journal of Psychology*, 117, 411-424.

Moyer, P. S., Bolyard, J. J., & Spikell, M. A. (2002). What are virtual manipulatives? *Teaching Children Mathematics*, *8*, 372-377.

Nathan, M. J. (2008). An Embodied Cognition perspective on symbols, gesture, and grounding instruction." In M. De Vega, A. M. Glenberg, & A. C. Graesser (Eds.), *Symbols and embodiment: Debates on meaning and cognition*, (pp. 375-396). New York: Oxford University Press.

Nathan, M. J. (2012). Rethinking formalisms in formal education. *Educational Psychologist*, *47*(2), 125-148.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs: Prentice-Hall.

Norman, D. A. (1988). *The psychology of everyday things*. New York: Basic books.

O'Malley, C., & Stanton-Fraser, D. (2004). *Literature review in learning with tangible technologies* (Report 12). Bristol: Nesta FutureLab Series.

O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, *24*, 939-1031.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, *9*(1), 97-113.

Olympiou, G., & Zacharia, Z. C. (2012). Blending physical and virtual manipulatives: An effort to improve students 'conceptual understanding through science laboratory experimentation. *Science Education*, *96*(1), 21 - 47.

Olympiou, G., & Zacharia, Z. C. (2012). Making the invisible visible: enhancing students' conceptual understanding by introducing representations of abstract objects in a simulation. *Instructional Science*, 1-22.

Osman, M. (2010). Observation can be as effective as action in problem solving. *Cognitive Science*, *32*(1), 162-183.

Paas, F. G., & Van Merriënboer, J. J. (1993). The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *35*(4), 737-743.

Paas, F., & Sweller, J. (2012). An evolutionary upgrade of cognitive load theory: Using the human motor system and collaboration to support the learning of complex cognitive tasks. *Educational Psychology Review*, *24*(1), 27-45.

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*, *38*(1), 63-71.

Page, M. (1990). *Active learning: historical and contemporary perspectives*. Unpublished manuscript, University of Massachusetts, Amherst. ERIC Document Reproduction

Park, B., & Brünken, R. (2014). The rhythm method: A new method for measuring cognitive load-An experimental dual-task sudy. *Applied Cognitive Psychology*, *29*(2), 232-243.

Patsenko, E. G., & Altmann, E. M. (2010). How planful is routine behavior? A selective-attention model of performance in the Tower of Hanoi. *Journal of Experimental Psychology: General*, *139*(1), 95-116.

Pecher, D., & Zwaan, R. A. (2005). *Grounding cognition: The role of perception and action in memory, language, and thinking*. Cambridge: Cambridge University Press.

Peelen, M. V., & Downing, P. E. (2007). The neural basis of visual body perception. *Nature Reviews Neuroscience*, *8*(8), 636-648.

Pickering, M. J., & Clark, A. (2014). Getting ahead: forward models and their place in cognitive architecture. *Trends in Cognitive Sciences*, *18*(9), 451-456.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*, 04, 329-347.

Pine, K. J., & Messer, D. J. (2000). The effects of explaining another's actions on children's implicit theories of balance. *Cognition and Instruction, 18*, 37-54.

Pine, K. J., Lufkin, N., & Messer, D. (2004). More gestures than answers: Children learning about balance. *Developmental Psychology*, *40*(6), 1059-1067.

Ping, R., & Goldin-Meadow, S. (2010). Gesticulation saves cognitive resources when talking about nonpresent objects. *Cognitive Science*, *34*, 602-619.

Pouw, W. & Hostetter, A. (in press). Gesture as Predictive Action. *Reti, Saperi, Linguaggi*: *Italian Journal of Cognitive Sciences*.

Pouw, W. T. J. L., De Nooijer, J. A., Van Gog, T., Zwaan, R. A., & Paas, F. (2014a). Toward a more embedded/extended perspective on the cognitive function of gestures. *Frontiers in Psychology*, *5*(539), 1-14.

Pouw, W. T. J. L., Eielts, E. C., van Gog, T., Zwaan, R. A., & Paas, F. (in press). Does (non-)meaningful sensori-motor engagement promote learning with animated physical systems? *Mind, Brain & Education, 10*(2), 91-104.

Pouw, W. T. J. L., Mavilidi, M., Van Gog, T., & Paas, F. (2016). Gesturing during mental problem solving reduces eye movements, especially for individuals with lower visual working memory capacity. *Cognitive Processing, 17*(3), 269-277.

Pouw, W. T. J. L., Van Gog, T., & Paas, F. (2014b). An embedded and embodied cognition review of instructional manipulatives. *Educational Psychology Review*, *26*(1), 51-72.

Pouw, W. T. J. L., Van Gog, T., Zwaan, R. A., & Paas, F (in press). Are gesture and speech mismatches produced by an integrated gesture-speech system? A more dynamically embodied perspective is needed for understanding gesture-related learning. *Behavioral and Brain Sciences.*

Pouw, W. T. J. L., van Gog, T., Zwaan, R. A., & Paas, F. (2016). Augmenting instructional animations with sensori-motor information to help children learn about physical systems. *Frontiers in Psychology*, 7 (860), 1-11.

Pouw, W. T. J. L., & de Jong, H. L. (2015). Rethinking situated and embodied social psychology. *Theory & Psychology*, *25*(4), 411-433.

Pulvermüller, F., Moseley, R., Egorova, N., Shebani, Z., & Boulenger, V. (2014). Motor cognition-motor semantics: Action perception theory of cognition and communication. *Neuropsychologia*, *34*(3), 602-619.

Radman, Z. (2013). *The hand, an organ of the mind*. Massachusetts: MIT press.

Rao, R.P.N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra classical receptive field effects, *Nature Neuroscience 2, 1,* 79-87.

Rauscher, F. H., Krauss, R. M., & Chen, Y. (1996). Gesture, speech, and lexical access: The role of lexical movements in speech production. *Psychological Science*, *7*, 226-231.

Resnick, L., & S. Omanson (1987). *Learning to understand arithmetic.* Pittsburgh: Erlbaum.

Rimé, B., Schiaratura, L., Hupet, M., & Ghysselinckx, A. (1984). Effects of relative immobilization on the speaker's nonverbal behavior and on the dialogue imagery level. *Motivation and Emotion*, *8*, 311-325.

Risko, E.F., Medimorec, S., Chisholm, J.D., & Kingstone, A. (2013). Rotating with rotated text: A natural behavior approach to investigating cognitive offloading. *Cognitive Science*, *38*(3), 537-564.

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, *3*(2), 131-141.

Robbins, T. W., Anderson, E. J., Barker, D. R., Bradley, A. C., Fearnyhough, C., Henson, R., Hudson, A. D., & Baddeley, A. D. (1996). Working memory in chess. *Memory & Cognition*, *24*(1), 83-93.

Roux, F.-E., Boetto, S., Sacko, O., Chollet, F., & Tremoulet, M. (2003). Writing, calculating, and finger recognition in the region of the angular gyrus: A cortical stimulation study of Gerstmann syndrome. *Journal of Neurosurgery*, *99*, 716-727.

Rupert, R. D. (2009). *Cognitive systems and the extended mind*. Oxford: Oxford University Press.

Ryle, G. (1945). Knowing how and knowing that: The presidential address. In *Proceedings of the Aristotelian Society* (Vol. 46, pp. 1-16).

Sarama, J., & Clements, D. H. (2009). 'Concrete' computer manipulatives in mathematics education. *Child Development Perspectives, 3*(3), 145-150.

Scheiter, K., Gerjets, P., & Schuh, J. (2010). The acquisition of problem-solving skills in mathematics: How animations can aid understanding of structural problem features and solution procedures. *Instructional Science*, *38*(5), 487-502.

Schiff, R., & Vakil, E. (2015). Age differences in cognitive skill learning, retention and transfer: The case of the Tower of Hanoi Puzzle. *Learning and Individual Differences*, *39*, 164-171.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences*, *36*(04), 393-414.

Schönborn, K. J., Bivall, P., & Tibell, L. A. (2011). Exploring relationships between students' interaction and learning with a haptic virtual biomolecular model. *Computers & Education*, *57*(3), 2095-2105.

Schwartz, D. L., & Black, T. (1999). Inferences through imagined actions: Knowing by simulated doing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 116-136.

Schwartz, D. L., & Martin, T. (2006). Distributed learning and mutual adaptation. *Pragmatics & Cognition*, *14*(2), 313-332.

Semenza, C., & Goodglass, H. (1985). Localization of body parts in brain injured subjects. *Neuropsychologia, 23*(2), 161-175.

Seyfeddinipur, M., & Kita, S. (2001). Gesture as an indicator of early error detection in self-monitoring of speech. In *ISCA Tutorial and Research Workshop (ITRW) on Disfluency in Spontaneous Speech.*

Shaer, O., & Hornecker, E. (2010). Tangible user interfaces: Past, present, and future directions. *Foundations and Trends in Human-Computer Interaction*, *3*(1-2), 1-137.

Shapiro, L. A. 2011. *Embodied cognition*. New York: Routledge.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*, 701-703.

Sherman, J., & Bisanz, M. (2009). Equivalence in symbolic and nonsymbolic contexts: Benefits of solving problems with manipulatives. *Journal of Educational Psychology*, *101*(1), 88-100.

Sigrist, R., Rauter, G., Riener, R., Wolf, P. (2013). Augmented visual, auditory, haptic, and multimodal feedback in motor learning: A review. *Psychonomic Bulletin & Review, 20*, 21-53.

Skulmowski, A., Pradel, S., Kühnert, T., Brunnett, G., & Rey, G. D. (2016). Embodied learning using a tangible user interface: The effects of haptic

perception and selective pointing on a spatial learning task. *Computers & Education*, *92*, 64-75.
Sloutsky, V. M, Kaminski, J. A., & Heckler. A. F. (2005). The advantage of simple symbols for learning and transfer. *Psychonomic Bulletin & Review*, *12*(3), 508-513.
Smithson, L., & Nicoladis, E. (2014). Lending a hand to imagery? The impact of visuospatial working memory interference upon iconic gesture production in a narrative task. *Journal of Nonverbal Behavior*, *38*(2), 247-258.
Snow, J. C., Pettypiece, C. E., McAdam, T. D., Mclean, A. D., Stroman, P. W. Goodale, M. A., & Culham, J. C. (2011). Bringing the real world into the fMRI scanner: Repetition effects for pictures versus real objects. *Scientific Reports*, *1*(130), 1-10.
So, W. C., Ching, T. H. W., Lim, P. E., Cheng, X., & Ip, K. Y. (2014). Producing gestures facilitates route learning. *PloS One*, *9*(11), 1-21.
So, W. C., Shum, P. L. C., & Wong, M. K. Y. (2015). Gesture is more effective than spatial language in encoding spatial information. *The Quarterly Journal of Experimental Psychology*, *68*(12), 2384-2401.
Son, J. Y., Smith, L. B., Goldstone, R. L., & Leslie, M. (2012). The importance of being interpreted: grounded words and children's relational reasoning. *Frontiers in Psychology*, *3*(45), 1-12.
Spanjers, I. A., Van Gog, T., & Van Merriënboer, J. J. (2010). A theoretical analysis of how segmentation of dynamic visualizations optimizes students' learning. *Educational Psychology Review*, *22*(4), 411-423.
Spivey, M. J., & Dale, R. (2011). Eye movements both reveal and influence problem solving. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *The Oxford handbook on eye movements* (pp. 551-562). Oxford: Oxford University Press.
Stephen, D. G., Dixon, J. A., & Isenhower, R. W. (2009). Dynamics of representational change: entropy, action, and cognition. *Journal of Experimental Psychology: Human Perception and Performance, 35*, 1811-1832.
Stevanoni, E., & Salmon, K. (2005). Giving memory a hand: Instructing children to gesture enhances their event recall. *Journal of Nonverbal Behavior*, *29*, 217-233.

Stigler, J. W. (1984). "Mental abacus": The effect of abacus training on Chinese children's mental calculation. *Cognitive Psychology*, *16*(2), 145-176.

Stull, A. T, Barrett, T., & Hegarty. M. (2013). Usability of concrete and virtual models in chemistry instruction. *Computers in Human Behavior*, *29*(6), 2546-2556.

Stull, A. T., Hegarty, M., Dixon, B., & Stieff, M. (2012). Representational translation with concrete models in organic chemistry. *Cognition & Instruction*, *30*(4), 404-434.

Svensson, H. (2007). Embodied simulation as off-line representation. Licentiate thesis, University of Linköping/University of Skövde, Sweden.

Svensson, H. (2007). *Embodied simulation as off-line representation* (unpublished doctoral dissertation), Linköping: Linköping University.

Symes, E., Ellis, R., & Tucker. M. (2007). Visual object affordances: Object orientation. *Acta Psychologica*, *124*(2), 238-255.

Tallis, R. (2003). *The hand: A philosophical enquiry into human being*. Edinburgh; Edinburgh University Press.

Thomas, L. E., & Lleras, A. (2007). Moving eyes and moving thought: On the spatial compatibility between eye movements and cognition. *Psychonomic Bulletin & Review*, *14*(4), 663-668.

Tomasello, M. (2008). *Origins of Human Communication*. Boston: MIT Press.

Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A new look at infant pointing. *Child Development*, *78*, 705-722.

Triona, L. M., & Klahr, D. (2003). Point and click or grab and heft: Comparing the influence of physical and virtual instructional materials on elementary school students' ability to design experiments. *Cognition and Instruction*, *21*(2), 149 - 173.

Triona, L. M., Klahr, D., & Williams, C. (2005). Point and click or build by hand: Comparing the effects of physical vs. virtual materials on middle school students' sbility to optimize an engineering design. In *Proceedings of the 30th Annual Conference of the Cognitive Science Society (pp. 2202-2205).*

Trofatter, C., Kontra, C., Beilock, S., & Goldin-Meadow, S. (2015). Gesturing has a larger impact on problem-solving than action, even when action is accompanied by words. *Language, Cognition and Neuroscience*. 30(3), 251-260.

Tucker, M., & Ellis, R. (1998). On the relations between seen objects and components of potential actions. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 830-846.

Tversky, B., Morrison, J. B., & Betrancourt, M. (2002). Animation: Can it facilitate? *International Journal of Human-Computer Studies*, *57*(4), 247-262.

Uttal, D. H., Scudder, K. V. & DeLoache, J. S. (1997). Manipulatives as symbols: A new perspective on the use of concrete objects to teach mathematics. *Journal of Applied Developmental Psychology*, *18*(1), 37-54.

Valenzeno, L., Alibali, M. W., & Klatzky, R. (2003). Teachers' gestures facilitate students' learning: A lesson in symmetry. *Contemporary Educational Psychology*, *28*, 187-204.

Van der Ven, S. H. G., Van der Maas, H. L. J., Straatemeier, M., & Jansen, B. R. J. (2013). Visuospatial working memory and mathematical ability at different ages throughout primary school. *Learning and Individual Differences*, *27*, 182-192.

van Elk, M., van Schie, H., & Bekkering, H. (2014). Action semantics: A unifying conceptual framework for the selective use of multimodal and modality-specific object knowledge. *Physics of Life reviews*, *11*(2), 220-250.

Van Gog, T., & Rummel, N. (2010). Example-based learning: Integrating cognitive and social-cognitive research perspectives. *Educational Psychology Review, 22,* 155-174.

Van Gog, T., Paas, F., Marcus, N., Ayres, P., & Sweller, J. (2009). The mirror neuron system and observational learning: Implications for the effectiveness of dynamic visualizations. *Educational Psychology Review*, *21*(1), 21-30.

Vernadakis, N., Gioftsidou, A., Antoniou, P., Ioannidis, D., & Giannousi, M. (2012). The impact of Nintendo Wii to physical education students' balance compared to the traditional approaches. *Computers & Education*, *59*(2), 196-205.

Wagner, S. M., Nusbaum, H., & Goldin-Meadow, S. (2004). Probing the mental representation of gesture: Is handwaving spatial? *Journal of Memory and Language*, *50*, 395-407.

Warburton, E. C., Wilson, M., Lynch, L., & Cuykendall, S. (2013), The cognitive benefits of movement reduction: evidence from dance marking. *Psychological Science, 24*(9), 1732-1739.

Wesp, R., Hesse, J., Keutmann, D., & Wheaton, K. (2001). Gestures maintain spatial imagery. *The American Journal of Psychology, 114*, 591-600.

Wheeler, M. (2007). *Reconstructing the cognitive world: The next step*. Cambridge: MIT Press.

Wheeler, M. (2013). Is cognition embedded or extended? The case of gestures, in Radman, (Ed.), *The hand, an organ of the mind: What the manual tells the mental,* Z.. MIT Press.

Wiebe, E. N., Minogue, J., Gail Jones, M., Cowley, J., & Krebs, D. (2009). Haptic feedback and students' learning about levers: Unraveling the effect of simulated touch. *Computers & Education*, *53*(3), 667-676.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, *9*(4), 625-636.

Wilson D. A., & Golonka, S. (2013). Embodied cognition does not mean what you think it means. *Frontiers in Psychology*, *4*, 263-275.

Winn, W. (2003). Learning in artificial environments: Embodiment, embeddedness and dynamic adaptation. *Technology, Instruction, Cognition and Learning*, *1*(1), 87-114.

Zacharia, Z. C., & Constantinou, C. P. (2008). Comparing the influence of physical and virtual manipulatives in the context of the Physics by Inquiry curriculum: The case of undergraduate students' conceptual understanding of heat and temperature. *American Journal of Physics*, *76*, 425-430.

Zacharia, Z. C., & Olympiou, G. (2011). Physical versus virtual manipulative experimentation in physics learning. *Learning and Instruction*, *21*(3), 317-331.

Zacharia, Z. C., Loizou, E., & Papaevripidou. M. (2012). Is physicality an important aspect of learning through science experimentation among kindergarten students? *Early Childhood Research Quarterly*, *27*(3), 447-457.

Zago, L., Pesenti, M., Mellet, E., Crivello, F., Mazoyer, B., & Tzourio-Mazoyer, N. (2001). Neural correlates of simple and complex mental calculation. *Neuroimage*, *13*, 314-327.

Zook, N. A., Davalos, D. B., Delosh, E. L., & Davis, H. P. (2004). Working memory, inhibition, and fluid intelligence as predictors of performance on Tower of Hanoi and London tasks. *Brain and Cognition*, *56*(3), 286-292.

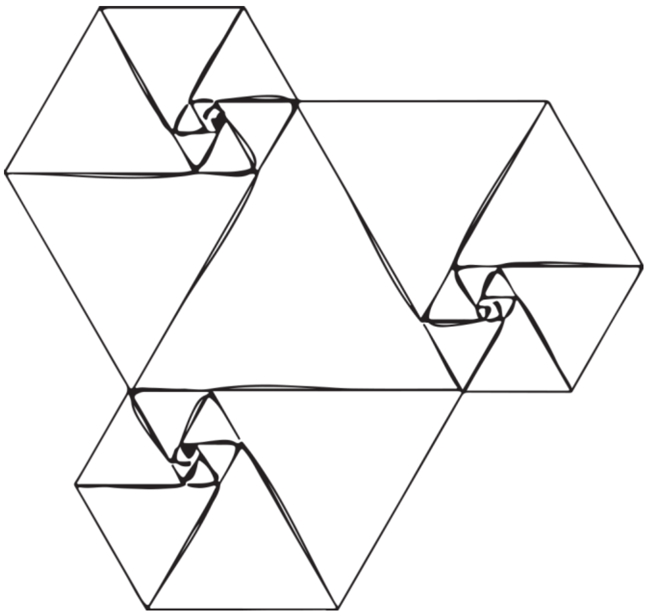
Zurina, H., & Williams, J. (2011). Gesturing for oneself. *Educational Studies in Mathematics Education*, *77*(2-3), 175-188.

Zwaan, R. A. (2014). Embodiment and language comprehension: Reframing the discussion. *Trends in Cognitive Sciences*, *18*(5), 229-234.

Zwaan, R. A. (2015). Situation models, mental simulations, and abstract concepts in discourse comprehension. *Psychonomic Bulletin & Review*, *23*(4), 1028-1034.

Nederlandse Samenvatting

Summary in Dutch



Korte introductie: Belichaamde cognitie

Het onderzoek gebundeld in deze dissertatie neemt theorieën over 'belichaamde cognitie' als uitgangspunt voor interventies om leren en probleemoplossen te verbeteren. Volgens deze benadering van cognitie zijn lichaamsbewegingen, en de mentale-neurale mechanismes die bewegingen ondersteunen, niet alleen een consequentie van cognitieve processen maar hebben ze ook een centrale rol in de totstandkoming en structurering van het denken. De belichaamde benadering van cognitie kan worden gekenschetst door het volgende simpele voorbeeld, dat als metafoor kan dienen voor cognitie (m.a.w. bewust en onbewust denken) in den brede. Stel dat een psycholoog wil begrijpen hoe een tekenaar uit het hoofd een bepaalde tekening produceert op een blad, zeg een landschap. Een klassieke psycholoog zal theorieën opstellen over de mentale voorstellingen en processen die het tekenen hebben gestuurd. Immers, de tekenaar moet zijn bewegingen sturen op zo'n manier dat het landschap dat zij 'voor ogen heeft' op het blad verschijnt. Deze sturing komt voort uit mentale voorstellingen van de tekenaar. Een psycholoog met sympathieën voor een belichaamd perspectief op cognitie, zal het proces echter anders interpreteren. Om Bredo (1994, p. 28) te parafraseren:

> Zij tekent, reageert op het getekende, tekent wat meer, enzovoort. De vooruitzichten voor de tekening veranderen naarmate de tekening evolueert, en verschillende mogelijkheden dienen zich aan, waardoor de ontwikkeling van de tekening een proces is van wederkerigheid tussen denken en handelen. Dit proces is daarom niet gedetermineerd door de mentale voorstellingen."
>
> (Bredo, 1994, pp. 28, vrij vertaald door de auteur)

Deze herdefiniëring van tekenen (en andere activiteiten) als een belichaamd proces heeft belangrijke consequenties voor het bestuderen van cognitie. De klassieke psycholoog zal zich niet geroepen voelen om de kenmerken van het potlood, de handgrip van de tekenaar, met andere woorden, het handelende proces van het tekenen zelf, te bestuderen. Immers, dit zijn strikt genomen niet-psychologische processen. Vanuit een belichaamde benadering van cognitie ziet men echter het handelend proces - de *wederkerigheid* tussen het potlood, blad, penstreken, en denken - als centraal voor hoe de tekening tot stand komt.

De Cognitieve Functie van Handbewegingen in Leren en Probleem Oplossen

In de theoretische en empirische studies gebundeld in deze dissertatie, onderzocht ik de rol van lichaamsbewegingen, in specifiek handbewegingen, in leren en probleem oplossen. Meer specifiek, werd in **Deel I** onderzocht of en hoe het manipuleren van de leeromgeving en het uitnodigen van handelend denken kan bijdragen aan leren. In **Deel II en III** werd onderzocht of en hoe handgebaren kunnen bijdragen aan het oplossen van problemen.

Als cognitie afhankelijk is van lichamelijke, handelende processen, wat voor implicaties zou dit dan hebben voor leerprocessen en het bevorderen daarvan? In **Deel I** start de verkenning van deze vraag door middel van een overzicht van de wetenschappelijke literatuur met betrekking tot het leren met fysieke objecten (ofwel “Instructional manipulatives”; **Hoofdstuk 2**). Men kan hierbij bijvoorbeeld denken aan het gebruik van blokjes tijdens rekenen om optellen en delen te leren begrijpen, aan het gebruik van fysieke modellen om moleculaire structuren te bestuderen, of aan het ontdekken van de werking van een hefboom middels een schaalmodel en gewichten. Leren met fysieke objecten kent haar oorsprong bij pedagogen als Montessori (1870-1952) en Pestalozzi (1746-1827). Het devies is dat fysieke interactie met objecten het leren beter ondersteunt dan het passief bestuderen van teksten, afbeeldingen, of abstracte formules.

Waarom deze fysieke leermodellen effectief zijn kan worden samengevat aan de hand van twee ideeën eigen aan belichaamde cognitie. Ten eerste, fysieke objecten kunnen er voor zorgen dat mentale processen worden vervangen door handelende processen waardoor het leren wordt vergemakkelijkt. Een goed voorbeeld hiervan is de bevinding dat studenten een beter begrip krijgen van de verschillende representaties van moleculaire structuren wanneer zij fysieke modellen actief kunnen roteren (Stull e. a., 2012). Dergelijke rotaties zouden normaliter *mentaal* moeten worden voorgesteld, wat het leerproces voor beginnende leerlingen te moeilijk kan maken omdat het maken van mentale rotaties veel van het werkgeheugen vergt. Het uitbesteden van delen van het probleem (mentale rotaties) aan handelend denken (fysieke rotaties van het model) kunnen het leren over gerelateerde aspecten (representaties van moleculaire structuren) dus verbeteren. Ten tweede, zijn fysieke objecten niet alleen belangrijk voor het denken *met* die objecten maar ondersteunen ze juist ook het toekomstig denken *zonder* deze objecten. Een goed voorbeeld hiervan is het leren rekenen met een telraam (ofwel *Abacus*; Hatano e.a., 1977; Hatano & Osawa, 1983). In landen als Japan, wordt er intensief gebruik gemaakt van het telraam om kinderen complexe berekeningen te leren uitvoeren. Het kind leert hierbij eerst de berekeningen uit te voeren met het telraam. Echter, wanneer expertise zich ontwikkelt kunnen de kinderen de complexe berekeningen uitvoeren zonder het telraam - dus mentaal. Echter, het telraam is op een psychologisch interessante manier nog steeds present. Kinderen die net mentale berekeningen leren te doen zonder het telraam, produceren namelijk handbewegingen *alsof* ze nog steeds het telraam manipuleren. Dit laat zien, dat de sprong van het fysiek rekenen met het telraam, naar het mentaal rekenen geen verandering lijkt te brengen in hoe ze de berekeningen aanpakken. Hier is het handelend denken met het telraam geïnternaliseerd in het mentale proces. Zonder lichamelijke ervaringen had dit effectieve mentale proces niet tot stand kunnen komen. Deze en vergelijkbare bevindingen met betrekking tot belichaamde cognitie komen in hoofdstuk 2 aan bod, en vormen de basis voor de onderzoeksvragen die in de volgende hoofdstukken worden behandeld.

In **Hoofdstuk 3** tracht ik uit zoeken welke informatie er kan worden gewonnen uit interactie met objecten tijdens het leren. Eerder onderzoek heeft voornamelijk gekeken naar de effecten van actieve interactie met fysieke of virtuele objecten in de leeromgeving in vergelijking met een passieve leerconditie. Alhoewel dit de effectiviteit van actieve versus passieve manieren van leren kan vaststellen, gaat dit voorbij aan de vraag *waarom* interactie effectief zou zijn. Bijvoorbeeld, het kan zijn dat het leerresultaat verbetert omdat interactie tijdens het leren de motivatie en concentratie van de leerling verhoogt in plaats van dat het bewegen van het object intrinsieke informatie oplevert over het onderwerp van studie. In deze context is er een experiment ontwikkeld waarbij jongvolwassenen leerden hoe een hefboom werkt aan de hand van een interactieve animatie. Studenten konden actief experimenteren met een wipwap (een type hefboom) en bestuderen hoe de wipwap reageert op verschillende krachten die werden uitgeoefend op de armen. Hun taak was de wipwap in evenwicht te brengen, wat ze deden door middel van het drukken met de handen op een Nintendo Wii Balance Board. Drie vormen van interactie werden met elkaar vergeleken. In de ‘congruente’ conditie was de druk die de studenten op het BalanceBoard moesten geven om de wipwap in de animatie in balans te brengen in overeenstemming met de afgebeelde objecten in de animatie (bijvoorbeeld: als er op de rechterarm één blokje stond, hoefde er links minder kracht gegeven te worden om de wipwap in balans te brengen dan wanneer er 2 blokjes stonden). In de 'incongruente' conditie was de druk die nodig was om de wipwap in balans te brengen niet consistent (de ene keer moest er harder gedrukt worden wanneer er één blokje stond dan wanneer er twee stonden, etc.). In de 'minimale' interactie conditie tenslotte, konden studenten de wipwap in balans brengen door een korte druk op het BalanceBoard.

De verwachting was dat studenten die op een congruente manier interacteerden met de animatie beter zouden scoren op verschillende taken die kennis over de geleerde concepten maten dan studenten die niet konden achterhalen hoeveel kracht nodig was om de hefboom in balans te brengen (minimale interactie conditie) of studenten die verkeerde krachten kregen aangeleerd (incongruente conditie). Er werden echter geen verschillen in leerprestaties gevonden tussen de condities. Dit was opmerkelijk, omdat een motorische test uitwees dat studenten in de congruente conditie het in balans brengen van de wipwap wel degelijk beter 'in de vingers hadden' dan studenten in de incongruente conditie. We concluderen op basis van deze bevindingen dat het weten hoe een hefboom reageert op basis van interacties niet altijd invloed hoeft te hebben op het latere redeneren over hefbomen. Deze bevinding lijkt in te druisen tegen belichaamde perspectieven op leren.

In **Hoofdstuk 4** wordt eveneens een experimentele studie beschreven naar het leren over het hefboomprincipe met behulp van een instructieve animatie. In dit experiment werd onderzocht of het uitnodigen tot handelend denken kan bijdragen aan het leren van basisschoolkinderen (8-12 jaar). Dit werd gedaan door een lichaamsanalogie te bieden voor de werking van de hefboom. Stelt u zich voor dat u uw armen horizontaal uitstrekt, en er een object van 2kg wordt geplaats op uw linkerarm ter hoogte van de hand, en een object van 2kg wordt geplaatst op de rechterarm ter hoogte van de elleboog. De objecten zullen niet even zwaar voelen, alhoewel ze even zwaar zijn. Dit is een klassiek hefboomeffect (mechanisch voordeel genaamd), waarbij de afstand van een last tot het draaipunt (schoudergewricht) mede bepaald hoeveel kracht er wordt uitgeoefend (hoe verder van het draaipunt, hoe meer kracht). In dit experiment werd er nagegaan of kinderen het principe van mechanisch voordeel van een wipwap beter zouden begrijpen als hen werd geleerd dat dit principe zowel op de wipwap als op het eigen lichaam betrekking heeft. Het idee is dat handelend denken, het voorstellen van de effecten van kracht op de armen, kan bijdragen aan het begrip van de werking van de hefboom omdat deze lichamelijke kennis impliciet bekend is bij kinderen, in tegenstelling tot de meer abstracte werking van een hefboom. De helft van de groep kinderen werd geleerd over het mechanisch principe van de wipwap door middel van een animatie, zonder een analogie. De andere helft van de groep kinderen kreeg dezelfde animatie te zien, maar dit keer met een bovenlichaam met uitgestrekte armen geprojecteerd over de wipwap, wat tot het maken van de analogie uitnodigde. De bevindingen lieten ziet dat kinderen over het algemeen minder fouten maakten op een testtaak, in het bijzonder kinderen die een lage CITO-score hadden op rekenen. Alhoewel het bemoedigend is dat een dergelijke kleine interventie het leren van kinderen met lagere vaardigheden kan bevorderen zonder het leren van kinderen met hogere vaardigheden te schaden, lieten de bevindingen ook zien dat wanneer een taak iets complexer wordt gemaakt door hefbomen met elkaar te koppelen (en er transfer van kennis nodig is), het voordelige effect van de analogie verdween. We concluderen daarom dat lichaamsanalogieën slechts een beperkt lijken te hebben op leren en dat verder onderzoek naar de specifieke grenzen van deze effecten nodig is (voor wie en onder welke omstandigheden het effectief is) voordat deze methode kan worden toegepast in het onderwijs.

In **Deel II** wordt de cognitieve functie van handgebaren voor het oplossen van problemen onderzocht. Eerder onderzoek heeft laten zien dat het gebruik van handgebaren voordelige effecten kan opleveren voor probleemoplossing. Bijvoorbeeld, wanneer proefpersonen mentale-rotatietaken moesten oplossen, produceerden ze spontaan handgebaren die de mimiek hadden van het manipuleren en roteren van de objecten die mentaal moeten worden geroteerd. Handgebaren lijken dus niet alleen een communicatieve functie te hebben, maar ontstaan ook in niet communicatieve situaties, ter ondersteuning van ruimtelijk denken. Een grote uitstaande vraag in gebarenonderzoek is echter *hoe* gebaren probleemoplossing ondersteunen.

Hoofdstuk 5 biedt een kritisch overzicht van theoretische verklaringen voor de functie gebaren in niet-communicatieve cognitieve processen. Een vaak geponeerde verklaring is dat handgebaren, bijvoorbeeld gebaren die ontstaan tijdens het mentaal roteren, voortkomen uit een mentale simulatie van manipulaties op objecten. Het idee is dat we kennis hebben over hoe we objecten kunnen manipuleren, en dat gebaren deze kennis tot uiting brengen met als gevolg een betere prestatie. In Hoofdstuk 5 wordt beargumenteerd dat het uiten van intern aanwezige kennis niet de enige drijvende kracht achter handgebaren kan zijn. Een dergelijke verklaring schiet tekort omdat deze negeert dat gebaren ook kunnen *bijdragen* aan het denken (een stelling die ik centraal schaar onder belichaamde cognitie). Waarom zou het cognitief systeem (moeizame) fysieke processen (handbewegingen) in gang zetten wanneer de kennis die nodig is om het proces mentaal uit te voeren al intern aanwezig is? Daarom wordt er in Hoofdstuk 5 een perspectief uitgewerkt waarin handgebaren (die worden gemonitord door het visuele en proprioceptieve systeem) cognitieve processen beïnvloeden (denk aan het voorbeeld m.b.t. tekenen aan het begin van deze samenvatting). In een dergelijke interpretatie is het de visuele en lichamelijke feedback die gebaren opleveren waardoor cognitie en probleemoplossen ondersteund worden.

Een specifieke hypothese die uit dit alternatieve perspectief voortvloeit, wordt getoetst in **Hoofdstuk 6**. Namelijk, uit eerder onderzoek naar handgebaren die ontstaan tijdens het praten, is gebleken dat wanneer mensen meer handgebaren gebruiken tijdens het praten wanneer hun werkgeheugencapaciteit relatief beperkt is (Chu e. a., 2014). Vanuit het huidige perspectief kan dit worden verklaard doordat gebaren het werkgeheugen ontlasten door informatie tijdelijk lichamelijk 'vast te houden'. In het experiment gerapporteerd in Hoofdstuk 6 is onderzocht of jongvolwassenen inderdaad vaker en meer handgebaren maakten tijdens het mentaal oplossen van een logische puzzel (de Toren van Hanoi) en een schaakprobleem, wanneer de taken relatief complexer waren zij een relatief beperktere visuele werkgeheugencapaciteit hadden. Ook onderzochten we of gebaren maken tijdens het mentaal oplossen, de latere prestatie tijdens het daadwerkelijk oplossen, verbeterde. Voor aanvang van de probleemoplossingstaken werden participanten getest op hun ruimtelijk-visuele werkgeheugencapaciteit. De taken vereisten van de participanten dat ze voor ieder probleem het probleem eerst mentaal doorliepen om vervolgens het probleem daadwerkelijk op te lossen. Tijdens het mentaal doorlopen van de problemen, hetgeen in stilte gebeurde, vond de experimentele manipulatie plaats. De helft van de participanten werd expliciet geïnstrueerd om wijsgebaren te maken tijdens het mentaal probleem oplossen. De andere helft kreeg geen gebaar-instructie. Een deel van deze groep die geen instructie kreeg, produceerde spontaan wijsgebaren die de locaties markeerden van de mentale stappen, terwijl een ander deel van de groep geen handgebaren gebruikte tijdens het mentaal doorlopen van het probleem.

Onze voorspelling was dat handgebaren vaker spontaan gebruikt zouden worden en dat gebaren (spontaan of geïnstrueerd) een positief effect zouden hebben op de mentale probleemoplossing (wat tot uiting komt in de prestatie tijdens het daadwerkelijk oplossen kort daarna) wanneer de taken moeilijker waren (en dus meer druk uitoefenen op intern werkgeheugen) en voor proefpersonen die relatief lagere ruimtelijk-visuele werkgeheugencapaciteit hadden. Hoewel participanten meer spontaan gingen gebaren bij moeilijke taken, was deze toename niet gerelateerd aan hun werkgeheugencapaciteit. Echter, zoals verwacht, presteerden proefpersonen die spontaan of geïnstrueerd gebaarden tijdens het mentaal probleem oplossen beter op het daadwerkelijk oplossen van de problemen dan proefpersonen die niet gebaarden, maar alleen bij moeilijkere problemen en voor individuen met een lagere werkgeheugencapaciteit. In dit geval lijkt het idee dat handgebaren effectief zijn omdat ze het werkgeheugen kunnen ontlasten en het mentaal probleemoplossen ondersteunen, stand te houden.

In **Hoofdstuk 7** trachten we de hoofdbevinding van hoofdstuk 6, namelijk dat handgebaren effectief zijn voor individuen met relatief lagere werkgeheugencapaciteit, te repliceren bij kinderen (8-12 jaar). Kinderen van die leeftijd zijn een interessante groep om het effect van gebaren te onderzoeken omdat hun werkgeheugencapaciteit nog in ontwikkeling is, waardoor de eventuele ondersteunende rol van handgebaren van extra belang kan zijn. Het uitgevoerde experiment was exact hetzelfde als het experiment uit Hoofdstuk 6, maar beperkte zich tot de Toren van Hanoi taak. De bevindingen van dit experiment laten zien dat kinderen, met vergelijkbare intensiteit als volwassenen, spontaan gebaren tijdens het mentaal doorlopen van de problemen. Dit is interessant, aangezien er weinig bekend is over de ontwikkeling van dit type handgebaren bij kinderen. De bevindingen repliceerden echter niet onze voorgaande resultaten: er waren geen effecten van spontaan of geïnstrueerd gebaren tijdens mentaal oplossen op de prestatie tijdens het daadwerkelijk oplossen van de taak. Deze bevindingen suggereren dat, alhoewel gebaren natuurlijk ontstaan bij kinderen tijdens mentaal probleem oplossen, het effect van deze gebaren op latere prestatie niet duidelijk te identificeren is, dit in tegenstelling tot jongvolwassenen.

Hoofdstuk 8 betreft een verdiepend onderzoek naar de vraag *hoe* wijsgebaren die worden gebruikt bij mentaal doorlopen van de Toren van Hanoi de cognitieve processen ondersteunen. Uit voorgaand onderzoek is gebleken dat de oogbewegingen van participanten die de taak mentaal doorlopen, synchroon lopen met de mentale stappen die zij maken (Patsenko & Altmann, 2010): Wanneer zij een stuk mentaal verplaatsen naar een andere positie, verplaatsen zij ook hun blikveld naar die nieuwe positie. De wijsgebaren lijken een vergelijkbare functie te hebben. Namelijk, de wijsgebaren markeren de mentale verplaatsingen van de stukken. Deze observatie is interessant in het licht van de voorgaande bevinding dat specifiek de individuen met een lagere visuele-ruimtelijke werkgeheugencapaciteit voordeel hebben van het maken van dit type gebaren. Wellicht bieden de gebaren een alternatieve strategie; door een plaats in de ruimte te markeren kunnen zij misschien als vervanging dienen voor de meer visueel/cognitief belastende strategie van het verplaatsen van het blikveld. Om dit te testen hebben we volwassenen de Toren van Hanoi mentaal laten doorlopen, eenmaal met het gebruik van wijsgebaren, en eenmaal zonder wijsgebaren (volgorde was willekeurig). Tijdens deze procedure werden de oogbewegingen van de participanten gemeten en voorafgaand aan het experiment hebben we eveneens ruimtelijk-visuele werkgeheugencapaciteit gemeten. De bevindingen lieten zien dat participanten die gebaren produceerden tijdens het mentaal probleem oplossen, minder hun blikveld verplaatsten vergeleken met participanten die geen handgebaren produceerden. Echter dit effect van gebaren op verminderde verwisseling van het blikveld was alleen sterk aanwezig voor diegenen met een lager ruimtelijk-visuele werkgeheugen. Deze bevindingen suggereren dat gebaren effectief zijn voor het ondersteunen van mentaal probleemoplossen omdat zij de ruimtelijke posities kunnen markeren op een non-visuele manier, namelijk via het lichamelijk verplaatsen van de handen. Deze handelingen kunnen als cognitieve vervanging dienen voor de meer visueel/cognitief belastende strategie, namelijk het verwisselen van het blikveld. Dit heeft voornamelijk voordeel voor diegenen met een lager ruimtelijk-visueel werkgeheugen.

Tot slot wordt er in **Hoofdstuk 9** een theoretisch kader geboden waarin de veschillende perspectieven op de cognitieve oorsprong en functie van handgebaren met elkaar verzoend worden. Er wordt een perspectief geboden waarin erkend wordt dat gebaren hun cognitieve oorsprong hebben in intern aanwezige kennis over manipulatief handelen (m.a.w. het simulatie perspectief op handgebaren), terwijl er tevens vanuit wordt gegaan dat de lichamelijke uiting van deze kennis in de vorm van handgebaren nieuwe informatie tot stand brengt die niet intern efficiënt genereerbaar is (m.a.w. beantwoordend aan de kritische noot in hoofdstuk 5). Deze verzoening wordt gestaafd aan de hand van een overzicht van recente cognitieve theorieën die poneren dat het cognitief systeem als functie heeft om een model van de omgeving te kalibreren door het toetsen van voorspellingen over de omgeving (Clark, 2015; Howhy, 2012). Gebaren, volgens deze theorie, zijn een manier om zulke voorspellingen te kalibreren aan de hand van de visuele en proprioceptieve informatie die de gebaren genereren.

Curriculum Vitae

Wim T. J. L. Pouw, born the 4th of October 1987, Murcia (Spain), obtained his Bachelor degree in Psychology in 2009, Master degree (research master) in Social Psychology in 2011 (cum laude), and a second master degree in Theoretical/Philosophical Psychology in 2012, at the VU University Amsterdam. During his Master studies he was employed as a KNAW "Akademie" research-assistant at the Department of Sociology at the VU University Amsterdam where he investigated the role of collective action in newly build neighborhoods. The theses and projects pursuit in the master studies and research assistantship led to a publication in *Theory and Psychology* (Pouw & Looren de Jong, 2015), and co-authored publications in *Acta Psychologica* and *Journal of Social and Political Psychology* (IJzerman, Gallucci, Pouw et al., 2012; Van Stekelenburg, Petrovic, Pouw et al., 2013). After his studies he was admitted as a PhD student for a 3 year project under supervision of Professors Tamara van Gog, Rolf Zwaan, and Fred Paas at the Erasmus University Rotterdam (EUR). The project concerned the investigation of supporting role of human movement in learning and problem solving. During this project he has published in journals such as *Applied Cognitive Psychology*, *Cognitive Processing*, *Behavioral and Brain Sciences*, and *Frontiers in Psychology*.

Currently Wim Pouw is employed at the EUR as a Postdoctoral reseacher under supervision of Prof. Fred Paas, with whom he is continuing his investigation on the cognitive function of gesture, as well as venturing into new domains such as haptic mental imagery. In 2017, the Netherlands Organisation of Scientific Research (NWO) has awarded Wim Pouw a Rubicon grant for a 2-year research project which applies human movement analyses to gesture, under supervision of Prof. James Dixon at the University of Connecticut.

Publications*

**order of acceptance*

Van Wermeskerken, M., Fijan, N., Eielts, E. C., & Pouw, W. (2016). Observation of depictive- versus tracing gestures selectively aids verbal versus visual-spatial learning in primary school children. *Applied Cognitive Psychology, 30,* 806-814. doi: 10.1002/acp.3256

Pouw, W. T. J. L., van Gog, T., Zwaan, R. A., & Paas, F. (2016). Augmenting instructional animations with sensori-motor information to help children learn about physical systems. *Frontiers in Psychology*, *7*, 1-11. doi: 10.3389/fpsyg.2016.00860

Pouw, W. & Hostetter, A. (2016). Gesture as predictive action. *Reti, Saperi, Linguaggi*: *Italian Journal of Cognitive Sciences*. *Reti, Saperi, Linguaggi: Italian Journal of Cognitive Sciences*, *3*, 57-80. doi: 10.12832/83918 Invited contribution to special issue "Language, Communication, and Embodied Cognition".

Pouw, W. T. J. L., Eielts, E. C., van Gog, T., Zwaan, R. A., & Paas, F. (2016). Does (non-)meaningful sensori-motor engagement promote learning with animated physical systems? *Mind, Brain & Education*, *10*, 91-104. doi: 10.1111/mbe.12105

Pouw, W. T. J. L., Van Gog, T., Zwaan, R. A., & Paas, F (in press). Are gesture and speech mismatches produced by an integrated gesture-speech system? A more dynamically embodied perspective is needed for understanding gesture-related learning. *Behavioral and Brain Sciences* (Commentary on Goldin-Meadow and Brentari, in press).

Pouw, W. T. J. L., Mavilidi, M., Van Gog, T., & Paas, F. (2016). Gesturing during mental problem solving reduces eye movements, especially for individuals with lower visual working memory capacity. *Cognitive Processing, 17*, 269-277. doi: 10.1007/s10339-016-0757-6

Pouw, W. T. J. L., & de Jong, H (2015). Rethinking situated and embodied social psychology. *Theory and Psychology, 25*(4), 411-433. doi:10.1177/0959354315585661

Pouw, W. T. J. L., De Nooijer, J. A., Van Gog, T., Zwaan, R. A., & Paas, F. (2014a). Toward a more embedded/extended perspective on the cognitive function of gestures. *Frontiers in Psychology*, *5*, 359. doi: 10.3389/fpsyg.2014.00359

Pouw, W. T. J. L., Van Gog, T., & Paas, F. (2014b). An embedded and embodied cognition review of instructional manipulatives. *Educational Psychology Review*, *26*(1), 51-72. doi: 10.1007/s10648-014-9255-5

Van Stekelenburg, J., Petrovic, I., Pouw, W. T. J. L., Limburg, N., & Nederlof, N. (2013). From correlation to causation: The cruciality of a collectivity in the context of collective action. *Journal of Social and Political Psychology*, *1*(1), 161-187 doi:10.5964/jspp.v1i1.38

IJzerman, H., Gallucci, M., Pouw, W. J. T. L., Weiβgerber, S. C., Van Doesum, N. J., & Williams, K. D. (2012). Cold-blooded loneliness: Social exclusion leads to lower finger temperatures. *Acta Psychologica, 140(3), 283-288.* http://dx.doi.org/10.1016/j.actpsy.2012.05.002

Articles Under review/Under revision

Pouw, W. T. J. L., Aslanidou, Kamermans, K., & Paas, F. (Under review). Is ambiguity detection in haptic imagery possible? Evidence for Enactive imaginings.

Kamermans*, K., Pouw, W. T. J. L.*, Mast, W. F. & Paas, F. (Under review). Ambiguity detection in visual imagery is possible without visual cues. ** Shared first authorship.*

Pouw, W. T. J. L., Eielts, C., van Gog, T., Zwaan, R. A., & Paas, F. (Under revision). When gestures are put into thought: The role of co-thought gestures, visual working memory, and task complexity in problem solving.

Pouw, W. T. J. L., Eielts, C., van Gog, T., Zwaan, R. A., & Paas, F. (Under review). The role of Co-thought Gestures, Visual Working Memory, and Task Complexity in Problem-solving: A replication study with children (8-12 yrs). Preregistered at OSF: https://osf.io/dreks/

Conference papers, poster presentations and presentations (*Presenting author[s])

Pouw, W. T. J. L.*, Kamermans, K., Mast., W. F., & Paas, F. (2017). Ambiguity detection in visual imagery is possible without visual cues. 2nd Biennial International Convention of Psychological Science (ICPS), poster presentation, March 23-25th, 2017.

Pouw, W. T. J. L.* (2016, May). The cognitive function of co-thought gestures in problem solving. Invited presentation, Centre for the Ecological Study of Perception and Action, University of Connecticut, USA.

Pouw, W. T. J. L. *, Eielts, C., Van Gog, T., Zwaan, R.A., & Paas, F. (2015, September). Embodied and embedded problem solving in primary school children and adults. Invited presentation at *Symposium Embodied Cognition and Multimedia Learning* (chairs: Van Gog, T., Zwaan, R. A., & Paas, F.), Erasmus University Rotterdam, The Netherlands.

Pouw, W. T. J. L.*, Van Gog, T., Zwaan, R.A., & Paas, F. (2015, October). Soliciting embedded and embodied cognition in learning and problem solving. Invited presentation at *Embodied Design* (chairs: Bakker, M. F., S. Shayan, & P. Leseman), Freudenthal Institute for Science and Mathematics Education, Utrecht University, The Netherlands.

Pouw, W. T. J. L., Eielts, C.*, Van Gog, T., Zwaan, R.A., & Paas, F. (2015, August). The cognitive benefits of co-thought gestures in physical and virtual problem-solving contexts in relation to working memory load. Presentation at *16th Biennial Conference EARLI*, Cyprus.

Pouw, W. T. J. L., Eielts, C.*, Van Gog, T., Zwaan, R.A., & Paas, F. (2015, June). When gestures are put into thought: the role of co-thought gestures, working memory, and difficulty in problem solving. Presentation at the *8th Cognitive Load Theory Conference*, Colarado, USA.

Pouw, W. T. J. L.*, Van Gog, T., Zwaan, R.A., & Paas, F. (2014, April). Instructional manipulatives for learning concrete and abstract content: An embodied interaction (re)view. Presentation at the symposium *Theoretical Perspectives and Empirical Evidence Relevant to Classroom Instruction with Manipulatives* (chairs: Marley, S, & Carbonneau, K. J.), *AERA Annual Meeting*, Philadelphia, USA.

Pouw, W. T. J. L.*, Van Gog, T., Zwaan, R.A., & Paas, F. (2014, June). Embedded embodied cognition: A review on manipulatives and gesture. Invited presentation at *Embodied Cognition in Education and Training* (Chairs: Bakker, M. F. van der Schaaf, S. Shayan, & P. Leseman), *Freudenthal Institute for Science and Mathematics Education*, Utrecht University, The Netherlands.

Pouw, W. T. J. L.*, Van Gog, T., Zwaan, R.A., & Paas, F. (2014). Augmenting instructional animations with sensori-motor information. Poster Presented at the *6th Conference of the International Society for Gesture Studies*, San Diego, USA.

Pouw, W. T. J. L.*, Eielts, C.*, Van Gog, T., Zwaan, R.A., & Paas, F. (2013, February). Embodied know-how: Does motor-involvement promote learning outcomes in science education. Poster presented at the *7th International Conference on Tangible, Embedded and Embodied Interaction*, Barcelona, Spain.

Van Stekelenburg, J.* , Petrovic, I. , Pouw, W. T. J. L., Limburg, N. and Nederlof, N. , (2012, July). The Evolution of Collective Action in Developing Neighborhoods. Paper presented at *The Annual Meeting of the ISPP 35th Annual Scientific Meeting,* Chicago, *USA.*

IJzerman, H., Gallucci, M., & Pouw*, W. T. J. L., Van Doesum, N. J., Vetrova, M., Williams, K. D. (2011, July). Understanding relations through experience: Social exclusion leads to lower bodily temperatures. Presented at *General meeting of the European Association for Social Psychology*, Stockholm, Sweden.

Van Leeuwen*, E., Grabo, A., & Pouw, W. T. J. L. (2011, July). Dealing with deviants. Presented at *13th Jena workshop on Intergroup Processes*, Friedrich Schiller University Jena, Germany.

Van Stekelenburg, J., Pouw*, W. T. J. L., Petrovic*, I., Limburg*, N., & Tessel, P. (2011, May). Evolution of protest in developing neighborhoods. Presented at the *VU Network Institute symposium*, VU University Amsterdam, The Netherlands.

Pouw, W. T. J. L.*, Van Doesum, N. J.*, Vetrova, M., Wei ßgerber, C. S. & IJzerman, H. (2011, December). De fysieke representatie van sociale relaties: Buitensluiting en lichaamstemperatuur. Presented at the *Annual ASPO Conference*, Enschede, The Netherlands.

Dankwoord (Acknowledgements)

Wie ik hier noem en perongeluk vergeet te noemen verdienen een langere lofzang dan ik hier zal geven. Ik hoop deze schuld te vereffenen in onze voortdurende (wetenschappelijke) kameraadschap en verbondenheid.

Tamara, ik wil jou bedanken voor werkelijk van alles, maar vooral voor je onvoorwaardelijke, scherpe en onvermoeibare begeleiding. Je hebt mij het gevoel gegeven dat ik er niet alleen voor stond. Kortom, een ideale begeleider.

Rolf, jouw spitsvondige visies op de Psychologie zou ik niet hebben willen missen van dichtbij. Ik heb enorme bewondering voor je theoretisch denkvermogen en empirische grondigheid.

Fred, jouw mentorschap heeft veel voor mij betekend, en doet dit nog steeds, en ik wil je bedanken voor al je geboden kansen en de geboden handvatten om deze kansen ook zelf te creëren. Op goede dagen hoop ik dat ik je academische werklust zal evenaren.

Charly, jij was cruciaal voor het eindresultaat van dit proefschrift. Je genealiteit en audididactisch vermogen gaat mij vaak te boven - als een regenboog niet te vangen ('insiders joke'). Dankjewel voor je vriendschap en dat je zo op zijn nodige tijd mij de dingen liet relativeren.

Ook wil ik andere mensen bedanken voor hun vriendschap. Gerdien, ik voelde mij direct thuis toen ik bij je op de kamer kwam, alsook een verdrietige vondeling na kamerverhuizing. Gelukkig weet ik dat we elkaar nog blijven treffen. Jesper, het was mij een eer om samen te dienen voor de intergalactische zaak. Margina, bedankt dat wij altijd even plots konden vergaderen over de existentiele zaak, met een zuur wijntje in de linkerhand (rechts voor gebaren). Hans van der Braak, wat een geluk dat je mijn buurman was, ik zal mijn gesigneerde kopie van *Prometheus Complex* koesteren en zo nu en dan weer oppakken. Kim, je nuchterheid en humor is verfrissend, ik ben blij dat je mijn kamergenoot bent. Hugo ik wil jouw bedanken voor je vreugdige en jeugdige visie op de Psychologie en Filosofie, stelling 11 zal je bekoren denk ik. Bij HLP en contreien heb ik genoten om in de aanwezigheid te zijn van Martine, Gertjan, Lisette, Sofie, Vincent, Judith, Mario, Lysanne, Marit, Tim, Bas (willekeurig volgorde). Bij B&C bedank ik graag Jan, Anita, en Jacqueline voor de uren aan discussies over "de goede zaak" (embodied cognition), ik heb veel van jullie geleerd. Diane en Rene, vergeef me de uitdrukking -

“academic power couple par excellence” – bedankt voor de gesprekken over de wetenschap en het leven.

Natuurlijk wil ik mijn gehele famillie bedanken, in specifiek Tante Wil en mijn vader Wim Pouw, voor hun rotsvaste steun en toeverlaat. Roos en Stefan, bedankt dat jullie mij altijd trots hebben laten voelen en voor de broodnodige gezelligheid.

Voor mijn verblijf in Wollongong wil ik Shirley Agostinho bedanken voor al haar hulp in het uitvoeren van onderzoek aldaar. Daniel Hutto voor zijn intellectuele gastvrijheid. Myrto en Yvonne voor hun vriendschap onder de Australische zon.

Ik wil de commissie bedanken voor het doorploeteren van dit proefschrift, ik ben mij terdege van bewust dat ik soms wat korter van stof kan zijn. Arthur Bakker wil ik bedanken voor zijn aanmoediging in de vorm van uitnodigingen voor praatjes te Utrecht en het delen van zijn breed geïnteresseerde visie op de drie E’s.

Het onderzoek in dit proefschrift is afhankelijk van veel mensen. In specifiek de kinderen, volwassen proefpersonen en scholen, en ook musea (Wollongong Science Centre) die hebben meegewerkt. Alsook de vele studenten die ik heb mogen begeleiden en die mij hebben begeleid in het uitvoeren van het onderzoek.

Ik zal jullie maar kort noemen, maar zonder kameraden is het leven weinig waard: Albert, Wouter, Dane, Etienne – ik ben blij dat wij elkaar altijd zullen treffen.

Tot slot wil ik richten tot twee van mijn favoriete mensen. Zus (Lucinda), kind van het licht, mij is bijgebleven wat jij in jouw dissertatie schreef - “met jou voelt het leven lichter en als we dat willen zwaarder”. Terwijl het leven wat meer gewicht aanneemt in tegenstelling tot de onbezonnen studententijd, wordt jouw verlichting steeds belangrijker voor mij. Ik ben blij te weten dat jij en Sjors, altijd bereikbaar zijn, waar we ons ook mogen begeven vanwege een of ander avontuur.

Azaria. Ik draag dit proefschift op aan jou. Hoe een hoopje moleculen toevallig heeft besloten zich zo te organiseren, is voor mij een autopoietisch wonder.

The human hand has been held to be of crucial importance for understanding (the evolution of) human intelligence. Next to the unique morphological characteristics of the human hand (e.g.,opposing thumbs relative to other digits) that allow humans to craft the environment in more flexible and sensitive ways than other animals, the use of the hands by humans is also truly unique. Notably, humans often move their hands in signifying ways that mirror or complement the content of speech and thought. To set these unique but ubiquitous types of hand-movements apart from the hands' most concrete manipulations, these hand-movements are referred to as gesticulation, or gestures in short. This dissertation is about the various ways that manual activity, such as gestures, but also actions on the environment, can support learning and problem solving.