How seeing iconic gestures facilitates action event memory and verb learning in 3-year-old children

by

Suzanne Aussems

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Department of Psychology

September 2017

THE UNIVERSITY OF WARWICK

Contents

List of Tables

List of Figures

Acknowledgments

First of all, I would thank all the actors in the videos of the GRACE stimuli database. I would also like to thank Natasha Kwok for helping me record, select, and cut the video files that make up the final set. A big thank you goes to Łukasz Walasek and Neil Stewart for helping me run my norming experiments in a web-based environment. And, of course, the development of this database would not be possible without the adult participants who took part in the norming experiments.

Furthermore, I would like to thank the nurseries, parents, caregivers, and children who took part in the memory and verb learning experiments reported in this dissertation. Without them, this research would not be possible. I am really grateful that the children were so sweet, excited, helpful, patient, and attentive during the tasks.

I would also like to express my sincere gratitude to my first supervisor, Professor Sotaro Kita, for his excellent advice on study ideas, scientific articles, grant proposals, and strategic career choices. It was a great pleasure to brainstorm about research ideas. We have probably come up with more ideas than we could possibly investigate in the next three years (without any help). It was also a very useful experience for me to write articles together. I feel that Kita has prepared me for the academic world in the best possible way. We have been very productive over the last three years, and I hope to continue our fruitful collaboration in the years to come.

I would also like to thank my second supervisor, Dr Katherine Messenger, for her invaluable advice on the practicalities of developmental research. Kate's comments on my Early Career Fellowship grant proposal were very useful.

I would like to thank the Institute for Advanced studies for awarding me an Early Career Fellowship (starting date 1st October 2017). This grant was awarded to be under the condition that I would submit my PhD before the starting date of the fellowship. This really gave me a good incentive to finish this dissertation before the end of the third year of my PhD.

I would like to thank the lab members for fruitful discussions and for creating a nice work environment. I would like to mention Priya Silverstein, Stephanie Archer, and Jacob Barker here. I think that their research is really interesting and I wish them all the best in their future studies.

I am also grateful to Neil Stewart and Bodo Winter for their useful comments on mixed-effects modeling.

I would like to express my gratitude to the Department of Psychology at the University of Warwick for funding my PhD research.

I also thank the editors and anonymous reviewers for their comments on the manuscripts on which this dissertation is based, whose suggestions were often very useful in improving my work.

I would also like to mention Łukasz Walasek and Gordon Brown here, who were always interested to hear about my progress and inspired me to submit my work to top-tier journals.

Thank you to my parents Herma and René Aussems, for being so supportive of me moving to the United Kingdom and visiting me during my PhD.

Finally, I would like to thank my friends Chris Emmery, Rick Smetsers, and Bart Joosten, for being there for me during my PhD trajectory, despite the fact that we are all affiliated with different universities and research institutes. I hope that they will successfully finish their PhD research, dissertation, and viva in the near future.

Declarations

This thesis is submitted to the University of Warwick in support of the application for the degree of Doctor of Philosophy. It has been composed by the author and has not been submitted in any previous application for any degree. The work presented (including data collection and data analyses) was carried out by the author. Parts of this dissertation have been published, accepted, submitted, or prepared for publication by the author:

Chapter 3: Aussems, S., Kwok, N., & Kita, S. (in press). GestuRe and ACtion Exemplar (GRACE) Video Database: Stimuli for Research on Manners of Human Locomotion and Iconic Gestures. *Behavior Research Methods.* DOI: 10.3758/s13428-017-0942-2

Chapter 4: Aussems, S., & Kita, S. (in press). Seeing Iconic Gestures while Encoding Events Facilitates Children's Memory of these Events. *Child Development.* DOI: 10.1111/cdev.12988

Chapter 5: Aussems, S., Mumford, K., & Kita, S. (2017). Seeing Iconic Gestures Helps Children Use Prior Action Knowledge for Verb Learning. *Manuscript in preparation.*

Chapter 6: Aussems, S., & Kita, S. (2017). Seeing Iconic Gestures Promotes Lasting Word-Category Knowledge about Verbs in Children. *Manuscript submitted for publication.*

Summary

People naturally produce iconic gestures when they speak. Iconic gestures that depict people's actions may influence the way children process action events. This dissertation investigates experimentally whether seeing such iconic gestures promotes 3-year-old children's action event memory and verb learning.

Chapter 1 introduces the topic and presents an outline of the dissertation.

Chapter 2 provides a working definition of gesture and a literature review on iconicity.

Chapter 3 describes the development and norming of a large database that contains stimuli videos of actions events and iconic gestures. Action videos showed actors moving in unusual manners and iconic gestures depicted how the actors moved.

Chapter 4 examines whether children remember action events differently when they see iconic gestures while encoding these events. Seeing iconic gestures that depicted how actors moved while encoding action events boosted children's memory of actors and their actions. Specifically, children showed better memory for event aspects that were depicted in gesture.

Chapter 5 asks whether prior action knowledge promotes verb learning and whether seeing iconic gestures influences this process. Pre-exposure to unlabeled actions facilitated verb learning when those actions were shown with iconic gestures and when children were shown two actors performing the same actions simultaneously, but children performed better in the iconic-gesture condition.

Chapter 6 investigates whether children learn that verbs typically refer to actions from seeing iconic gestures that depict individual verb meanings. Children who were taught verbs with iconic gestures demonstrated such word-category knowledge about verbs in an immediate and delayed novel verb learning task in which different novel verbs were taught without iconic gestures.

Chapter 7 discusses theoretical and practical implications of the experimental findings. Iconic gestures are meaningful social cues that help children individuate people's actions, encode and remember complex action events, acquire individual verb meanings, and generate word-category knowledge about verbs.

Chapter 1

Introduction

1.1 Prologue

♢

Imagine a mother and her young toddler having breakfast in their kitchen. The toddler is having their morning milk. The mother notices that they are running late for nursery. She asks the child "Come on, *dax* your milk". The child, who does not know the meaning of the word *dax*, has to figure out the correct meaning of the word. Based on how the utterance is structured the child can probably deduce that he or she is required to take an action. Even so, there are many possibilities in this context alone to which the word *dax* could refer. It could mean, "leave", "hand over", "spit out", "drink", or even something completely different than that. This example illustrates referential ambiguity, which occurs when a novel word could refer to multiple referents (Quine, 1960). Children must solve this mapping problem in order to develop adult-like vocabularies.

♢

1.2 Vocabulary Development in Preschool-Aged Children

Children understand and produce quite a few words by the time they reach school age. Understanding the meaning of words prepares children for producing these words in the right contexts. Therefore, production generally lags behind comprehension (Hendriks, 2014). Children build on their receptive vocabulary from as early as six months old. We know this because they start to respond to hearing their own name, words like “mummy” and “daddy”, and to routine expressions such as “bye-bye” and “peek-a-boo” (Clark, 1995). At six to nine months, children start to understand the meaning of many common nouns too (e.g., “milk”, “spoon”) (Bergelson & Swingley, 2012). On average, children typically understand 50 words around their first birthday, but this varies a great deal per individual child (Fenson et al., 2006). This number then gradually increases to approximately 120 words at 15 months and 170 words at 16 months (Fenson et al., 2006). Around 18 months of age, the vocabulary spurt kicks in and children acquire approximately ten novel words per week (McCarthy, 1954; Goldfield & Reznick, 1990).

Studies on productive vocabulary have shown that at the start of the vocabulary spurt children produce approximately 100 words (Mervis & Bertrand, 1995). Between 18 and 24 months of age, children learn how to combine two words into one utterance (Gillibrand, Lam, & O’Donnell, 2011). By the time children are 3 years old, their productive vocabulary typically includes 1000 or more words, and they can combine three or four words into short utterances (Fenson et al., 2006). Three-year-old children are already excellent communication partners, but they still have a long way to go before they possess adult-like vocabularies.

It is important to study vocabulary development in preschool-aged children, because their vocabulary size and skills are major predictors of later academic achievements (Anderson & Freebody, 1979). For instance, the rate of vocabulary growth at 30 months of age predicts children’s vocabulary skills at kindergarten entry (Rowe, Raudenbusch, & Goldin-Meadow, 2012). Moreover, the size of children’s vocabulary at 24 months predicts reading and math skills as well as self-regulation and social behavior in kindergarten (e.g., Morgan, Farkas, Hillemeier, & Scheffner Hammer, 2015). There is also considerable evidence that the vocabulary size of 2-year-old children is a strong predictor of cognitive abilities and language skills at 3 years of age (Feldman et al., 2005) and even at 8 years of age (Marchman & Fernald, 2008). In addition, vocabulary skills at 23 to 31 months are linked to reading competence at age 13 (Rescorla, 2005) and age 17 (Rescorla, 2009). Even though many studies have shown the importance of vocabulary development, children’s vocabulary size varies greatly by the time they enter school.

1.3 What Causes Variability in Vocabulary Size at School Entry?

It is commonly known that children from low socio-economic status (SES) families enter school with smaller vocabularies than children from high-SES families (Hoff, 2006). Early differences in child vocabulary have been related to the way parents from high and low SES families speak to their children (Hoff, 2003). Numerous studies have shown that high-SES parents talk more to their children (Hoff, Laursen, & Tardif, 2002), use more diverse vocabulary (Huttenlocher, Waterfall, Vasilyeva, Vevea, & Hedges, 2010), and form sentences using more complex grammar than low-SES parents do (Huttenlocher, Vasilyeva, Cymerman, & Levine, 2002). However, people do not communicate only through speech but also with their bodies, using the space around them. Words are so essential in communication that nonverbal signals are often overlooked in traditional language learning studies.

Less commonly studied is therefore how SES differences are reflected in parents' gesture use. Rowe and Goldin-Meadow (2009) investigated this question in a seminal study. They recorded 14-month-old children in their homes for 90 minutes, while the children were engaging in normal-day activities with their primary caregivers. All speech and gesture produced by both parents and children were coded to obtain a measure of vocabulary and gesture use. Vocabulary size, child gesture use, and parent gesture use were all operationalized as the number of unique meanings communicated during the session. SES was operationalized as a continuous score based on family income and education level. In a correlation analysis, it was found that parents and children from high-SES families used gesture to communicate a broader range of meanings than parents and children from low-SES families. A follow-up regression analysis showed that parent gesture use mediates the relationship between child gesture use and vocabulary development. That is, if parents produce a broader range of gestures for their children, then children gesture in a more diverse way too, and this predicts vocabulary development. Both parental speech and gesture thus influence children's vocabulary development.

1.4 Nonverbal Scaffolds for Difficult Tasks

The study by Rowe and Goldin-Meadow (2009) showed, for the first time, a possible causal relationship between nonverbal scaffolding and early vocabulary development. The term "scaffolding" refers to guidance provided by adults that narrows the gap between a child's level of ability and the demands of a complex task (D. Wood, Bruner, & Ross, 1976). The goal of scaffolding is to increase the chance of the child succeeding by making the task easier in some way. Nonverbal scaffolding thus refers to guidance provided by means of an adult's gestures to facilitate the demands of a complex task for children.

Even though preschool-aged children already know and produce quite a few words, they still experience difficulties with certain memory- and language-related tasks, which we identify in the next section. A nonverbal scaffold, such as gestures produced by adults that are relevant to the task at hand, could then be introduced as an experimental manipulation to help children overcome their struggle. This could, in turn, help us investigate what type of social cues children rely on in the early stages of cognitive development.

1.5 Encoding Action Events and Verb Learning are Difficult Tasks

We identify two tasks in this section that are difficult for 3-year-old children. First, 3-year-old children find it challenging to encode action events. Specifically, they find it more challenging to encode components of action events that are relevant for verb meaning (e.g., actions) than components of action events that are irrelevant for verb meaning (e.g., objects). For example, Imai, Haryu, and Okada (2005) conducted an event recognition task with 3-year-old Japanese-speaking children, who watched videos of an experimenter performing an action on an object (only the experimenter's hands were in the camera view). The next day, children performed a two-way forced-choice task in which they were asked to point at the video that they had seen the day before. They were tested on their recognition memory of both objects and actions. In object recognition trials, children were presented with two videos on a split screen: one showed the event from the day before and the other showed the female's hands acting on a novel object in the same way as in the target event. In action recognition trials, children were also presented with two videos on a split screen: one showed the event from the day before and the other showed the female's hands performing a different action on the same object as in the target event. Children could pick out the target events above chance level in both action and object recognition memory trials, but they recognized objects significantly more often than actions. This finding suggests that actions, as components of an event that are relevant for verb meaning, may be less well encoded objects, which are irrelevant for verb meaning.

Second, children struggle to learn the meaning of action verbs (i.e., words that describe actions such as *to dax*). Action verbs can either be *transitive* or *intransitive*. Intransitive verbs have two key features. First, they express an action (often a manner of locomotion) such as "*to jump*", "*to run*", "*to sit*", or "*to skip*". They require an actor and an action (e.g., the boy (actor) *jumped*). Second, intransitive verbs do not have a direct object receiving the action, unlike transitive verbs, which refer to actions that require a direct object such as "*to break* the toothpick (direct object)", "*to kick* the ball (direct object)", and "*to buy* a ring (direct object)", where the direct objects are interchangeable. Transitive verbs can also take an instrument (e.g., the boy (actor) *cuts* the paper (direct object) with scissors (instrument)).

Studies on both intransitive and transitive verb learning have shown that children struggle to extend newly learned action verbs beyond the events in which they were originally learned. For example, a study by Kersten and Smith (2002) demonstrated that 3.5- to 4-year-old English-speaking children experienced difficulties in extending newly learned intransitive verbs to events that showed the referent actions performed by novel actors. In their experiment, children were exposed to stimuli videos in which actors (i.e., bug-like creatures) moved around a scene while verbs were taught. Children did not accurately extend the verbs to events in which novel bug-like creatures moved around the scene in the same manner. One could argue that children included actors in their semantic representation of verbs in this study, because the actors were highly unusual bug-like creatures. However, in a study by Imai, Kita, Nagumo, and Okada (2008), who used human actors in their stimuli events, 3-year-old Japanese children also struggled to generalize intransitive verbs to events that showed the referent actions performed by novel actors. During the task, a female experimenter labeled a video of a novel action (e.g., an actor walking across the length of a scene in an unusual way) with a novel verb. Children were then required to extend this novel verb to one of two videos on a split screen. One video showed a novel actor performing the target action and the other video showed the actor from before performing a novel action. Children performed at chance level in this generalization task. Taken together, these studies suggest that around their third birthday, children experience difficulties with individuating intransitive actions (i.e., separating actions from the actors who perform them).

Furthermore, studies on transitive verb learning have shown that children experience difficulties with extending a newly learned verb to events showing the referent actions performed on novel objects (Imai et al., 2005; Imai, Li, et al., 2008). In a study by Imai et al. (2005), 3-year-old Japanese-learning children were taught verbs while watching videos of an experimenter performing actions on objects (e.g., rolling a football-shaped object between the palms of the hands). Only the experimenter's torso was in the camera view. Children were then required to generalize the verbs to one of two videos on a split screen: one video showed the referent actions performed on novel objects (e.g., rolling a pipe-shaped object between the palms of the hands) and the other showed different actions performed on the objects that children had seen when the verbs were taught (e.g., tapping the football-shaped object against the shoulder). Children performed at chance in this generalization task. This study thus shows that 3-year-old children also struggle to individuate transitive actions (i.e., separate actions from the objects acted upon).

Finally, studies on transitive verb learning have shown that children find it difficult to generalize novel verbs to events that show the referent actions but performed with novel instruments (i.e., the means or tools used to execute the action) (Behrend, 1990; Forbes & Farrar, 1993). For example, in the study by Forbes and Farrar (1993), 3-year-old English-learning children were taught novel verbs while watching video clips involving actors depicting novel actions with instruments (e.g., an actor pushing another actor seated in a shopping cart). In a generalization task, children were asked whether the novel verbs could also refer to novel video clips in which one of the semantic components of the actions events had changed (i.e., instrument, actor, direction, continuity). Children were least likely to accept those action events in which the instrument component had changed (e.g., an actor pushing another actor seated in an office chair). Thus, children experience difficulties with separating actions from the instruments used to perform the actions.

All the above verb learning studies show that around their third birthday, children tend to include components of action events in their semantic representation of a verb that are irrelevant for verb meaning (i.e., actors, objects, and instruments), which makes it hard for them to extend newly learned verbs to novel events. Findings from the study by Imai et al. (2005) suggest that this may be the case because children focus too much on the stable component of an action event, rather than on the transient component that is important for verb meaning. However, Imai et al. (2005) only tested children's recognition memory of objects and actions, and not of actors and actions, which is important for investigating children's struggle with intransitive verb learning. We therefore adapted the paradigm of Imai et al. (2005) to investigate children's action recognition memory and actor recognition memory. In this paradigm we also introduced a nonverbal scaffold that could help children to encode action events. Moreover, we adapted the paradigm by Imai et al. (2005) to examine nonverbal scaffolds that may help children to generalize intransitive verbs. As there were no stimuli of intransitive actions openly available, we also developed and normed a stimuli database.

1.6 Using Iconic Gestures to Encode Semantic Information

Let us think back to the example in the prologue in which we used the nonsense verb *dax* to give a feel of the ambiguity problem that children must solve. The mother notices that her child is not following up on her initial request and she repeats it "Come on, *dax* your milk", while producing a hand shape as if holding a cup and tipping up her hand to her mouth as if drinking. The child now understands that the mother is referring to drinking the milk. Her gesture iconically depicted the meaning of the novel word *dax*, that is, the action of drinking from a cup in hand shape and in motion.

Iconic gestures depict key characteristics of a referent (e.g., flapping the arms to represent the wings of a flying bird) (McNeill, 1992). When people speak, they naturally produce iconic gestures, and these gestures often illustrate what is said (McNeill, 1985). In some cases, they closely match parts of speech. For instance, if someone says "Would you like some more *food*?", while bringing one hand to the mouth as if holding food. The same message could be communicated if someone would say "Would you like some *more*?", while producing the same iconic gesture. In the latter example, the gesture complements information conveyed in speech. Without paying attention to the information encoded in iconic gesture (e.g., food in the example), a recipient could have misinterpreted the message. Both examples show that iconic gestures are social cues that carry semantic meaning (McNeill, 1992), and could therefore help to interpret the accompanying speech.

When young children encounter action events, iconic gestures that adults use when talking about these events can help children to encode the action components. For instance, if a mother and her toddler are both looking at the pet dog who sits pretty and the mother says "Look at what the doggy is doing!", while producing a two-handed gesture that depicts the position of the dog's front paws, then this gesture may help the child to identify the action of sitting pretty. Similarly, when a child does not know the meaning of an action label, an iconic gesture that is produced with this label could help them learn the novel verb's meaning (e.g., "Come on, *dax* your milk", while producing a gesture that depicts drinking from a cup).

1.7 Scope and Research Questions

This dissertation investigates how iconic gestures produced by adults can scaffold children's action event memory and verb learning. Each of the chapters contributes to answering four research questions that are central to this dissertation:

RQ1 What requirements should stimuli meet to investigate action event memory and verb learning with the help of iconic gestures?

RQ2 Do children encode and remember action events differently when their attention is guided by iconic gestures?

RQ3 Do children use prior unlabeled exposures to actions for verb learning and does seeing iconic gestures influence this process?

RQ4 Do children gain word-specific knowledge of verbs and word-category knowledge about verbs from seeing iconic gestures during verb learning?

1.8 Outline of the Dissertation

The Chapters of the dissertation are outlined below. Chapter 2 opens with a brief overview of the field of research in Section 2.1, followed by a working definition of gesture in Section 2.2 and a note on how this working definition should be used in Section 2.3. Subsequently, Section 2.4 defines the dimension of iconicity followed by Section 2.5 on when children become sensitive to iconicity and Section 2.6 on when children use iconicity for word learning. Subsequently, we discuss the developmental stages of spontaneous gesture production in Section 2.7. Finally, Section 2.8 outlines one of the leading views in gesture studies on the relation between gesture, speech, and thought.

Chapter 3 addresses RQ1, by reporting on the development of a large stimuli database that contains videos of intransitive action events and iconic gestures that depict those events. This database was normed across four experiments to investigate the requirements these stimuli should meet for research on action event memory and verb learning. The stimuli videos that received the best ratings were used as experimental materials in the empirical chapters that follow.

Chapter 4 focuses on RQ2 and describes an experiment that examined children's action event memory. This chapter is the precursor for the verb learning chapters that follow, as it presents data on children's action and actor recognition memory, which are key to verb learning and generalization. We virtually take a step back from verb learning in this chapter and investigate what children encode when they see action events, but also how seeing iconic gestures with these action events influences this process.

Chapter 5 discusses RQ3 and presents two experiments in which children were taught novel verbs. We examined what type of prior action knowledge would positively influence children's subsequent verb learning. Importantly, we pre-exposed children to unlabeled actions before labeling novel exemplars in which different actors performed the same actions to see if this influenced their verb learning performance. We also investigated whether seeing iconic gestures that depicted the actions in the unlabeled exemplars influenced this process.

Following up on this research, Chapter 6 addresses RQ4 and describes two experiments that were conducted to investigate the far-reaching and long-term effects on verb learning that seeing iconic gestures may have. Specifically, we examined whether seeing iconic gestures helps children to learn the meaning of individual verbs (word-specific knowledge) and, more importantly, whether seeing iconic gestures that depict individual verb meanings helps children to generate abstract linguistic knowledge about what verbs are (word-category knowledge).

Finally, in Chapter 7, the mechanisms that may underlie the beneficial effect of seeing iconic gestures on action event memory and verb learning in 3-year-old children are discussed. A summary of the research findings and answers to the research questions are provided in Section 7.1. We also discuss the theoretical and practical implications of the findings in Section 7.2. Next, we highlight some of the research output of the dissertation work in Section 7.3, followed by a note on the used analyses in Section 7.4. We then provide recommendations for future research in Section 7.5. This Chapter ends with Section 7.6 in which the main conclusions of the dissertation are presented.

Chapter 2

Background

2.1 The Field of Gesture Studies

The field of gesture studies is a relatively young field of research. The International Society for Gesture Studies (ISGS), which was founded in 2002, is the only international scholarly association dedicated to the study of human gesture (Andrén, n.d.). ISGS is supported by the international peer-reviewed journal *Gesture*, founded by Adam Kendon and Cornelia Müller, which publishes original research on all aspects of human gesture.

Gesture studies as a field is broadly concerned with investigating how people use their hands (and other parts of their body) to communicate (McNeill, 2009). A growing body of research shows that gesture is indispensable from many aspects of human life, including thought, collaborative work, science, art, music, and dance (Andrén, n.d.). Gesture researchers work across a wide range of academic disciplines including (but not limited to) linguistics, psychology, neuroscience, anthropology, communication, computer science, music, and dance.

The diversity of gesture studies, and its interdisciplinary nature, are demonstrated by the range of topics studied. To give just a few examples, neuroscientists investigate the relationship between speech and gesture by investigating whether both modalities are processed by the same areas in the brain (e.g., Özyürek, Willems, Kita, & Hagoort, 2007). Computer scientists model emotional gestures in dance and music performances (e.g., Camurri, Mazzarino, Ricchetti, Timmers, & Volpe, 2004). Ethnolinguists document how gestures differ across cultures (e.g., Enfield, 2001). Cognitive scientists research whether gestures can reveal the strategies in children's repertoire for solving Piagetian conservation problems (e.g., Church & Goldin-Meadow, 1986; Goldin-Meadow, Alibali, & Church, 1993; Kelly & Church, 1998), mathematical equivalence problems (e.g., Alibali & Goldin-Meadow, 1993; Broaders, Cook, Mitchell, & Goldin-Meadow, 2007) and estimation tasks (e.g., Heine et al., 2009).

2.2 A Working Definition of Three Gesture Categories

Gestures can have varying definitions in different contexts. The word gesture captures multiple communicative movements, primarily but not always of the hands and arms (McNeill, 2009). In the case of emotional gestures in dance and music performances (Camurri et al., 2004), one could imagine that facial expressions and body posture are also considered as gestures. Ostensive signals such as eye gaze could serve as gestures too. Imagine someone hinting to another person across the room that they want to leave by gazing at the door. By no means are we denying that these nonverbal behaviors are not gestures, but for the purpose of this dissertation, we define gesture according to Kendon (1982) and McNeill (1992), who formulated the term *gesticulation*. McNeill (2009) explains that "gesticulation is motion that embodies a meaning which can be related to the accompanying speech. It is produced mainly with the arms and hands, but is not restricted to these body parts – the head can take over as a kind of third hand if the anatomical hands are immobilized or otherwise engaged, and the legs and feet too can move in a gesture mode." (p. 299). Gesticulations, or gestures, are defined as the way humans move their hands (and body) when they speak.

Several gesture taxonomies have been proposed over the years, but most researchers have drawn on the gesture taxonomy by Kendon (1982). McNeill (1992) lined up the categories of different gestures on "Kendon's continuum", which he named after Adam Kendon. The three gesture categories that are discussed in this dissertation are *deictic gestures*, *iconic gestures*, and *interactive gestures*. Deictic gestures and iconic gestures are gesticulations that McNeill (1992) described as part of Kendon's continuum. Both are meaningful in the context of speech (McNeill, 1992). However, when McNeill (1992) categorized different gestures, he based his classification on analyses of people's gestures during monologues and not during interactions. Therefore, interactive gestures are underrepresented in Kendon's continuum, and we define this category in our working definition based on work by Bavelas, Chovil, Lawrie, and Wade (1992).

2.2.1 Deictic Gestures

Deictic gestures are generally understood as pointing gestures that indicate the location of real, implied, or imaginary individuals, objects, and directions. *Deixis* is derived from the Greek word *deiknunai*, which means "to show" (Oxford English Dictionary, 1940). The prototypical deictic gesture involves extending the index finger and arm in the direction of the referent, whereas the remaining fingers are curled under the hand with the thumb held down and to the side (Butterworth, 2003). Other extendable body parts than the finger, or held objects, can be used to point too (McNeill, 1992) (e.g., a teacher's pointer in the classroom). In some cultures, people use their extended lips and chin to point at referents (Enfield, 2001).

Deictic pointing gestures locate entities in space. For example, a parent can point at the child's plate at the dinner table while saying the words "Eat your vegetables". Much of the pointing gestures that adults produce are actually not referring to physically present entities but to abstract entities. For example, when an adult says "the film was beautiful from the beginning to end", while producing a pointing gesture from left to right while saying "beginning to end", then the pointing gestures refer to these abstract time concepts. Children generally start pointing at physical objects before their first birthday (Butterworth, 2003), but they do not produce many abstract deictic pointing gestures before age 12 (McNeill, 1992).

2.2.2 Iconic Gestures

Iconic gestures are hand gestures that represent meaning that is closely related to the semantic content of the segments of speech that they accompany (McNeill, 1985). *Iconic* is derived from the Greek word *eikōn*, which means "likeness, image" (Oxford English Dictionary, 1750). Iconic gestures communicate semantic information in the form of visual representations that look similar to their referential meaning. As such, they present images of concrete, objects, actions, people, or events spoken about (McNeill, 2009). For instance, when someone says "the TV chef chopped the onion in only a few seconds", while producing a rapid upward and downward hand movement with an open palm oriented sideways to depict the action of chopping. Importantly, iconic gestures are referential symbols, which function via their formal and structural resemblance to events or objects (McNeill, 1992; Peirce, 1955).

2.2.3 Interactive Gestures

Interactive Gestures often refer to the speaker or addressee rather than to the topic of conversation, and they help maintain the conversation as a social system. They fulfil a discourse function (Bavelas et al., 1992). Common features of interactive gestures are i) citing a previous contribution (e.g., highlighting a topic that has been mentioned before), ii) seeking agreement, understanding, or help (e.g., flicking the hand outwards, towards the addressee, with palm facing up and index finger pointing at the addressee as if to say "you know"), iii) the delivery of new versus shared information, and iv) coordinating turn taking. No two interactive gestures are exactly alike, but the primary characteristic of such gestures is that they convey no concrete information about the topic of conversation.

The category of interactive gestures subsumes *beat gestures*, which McNeill (1992) lined up on Kendon's continuum. Beat gestures are motor gestures associated with the coherence of speech through rhythm, for example, marking initiation of new discourse or introduction of new topics of conversation (McNeill, 1992). The gestures look like repetitive strokes of the hand that have an emphatic quality, drawing attention to what is being said with the beat emphasizing important part of speech. Since they do not refer to entities, real or abstract, they have more of a discourse function than deictic gestures or iconic gestures. McNeill (1992) notes that "of all gestures, beats are the most insignificant looking" (p. 15).

2.3 Dimensions Rather than Fixed Gesture Categories

It is important to note here that McNeill explains gesture taxonomies should not propose fixed categories, but rather that gestures may exhibit "dimensions" of multiple categories simultaneously (e.g., deixis, iconicity, and "temporal highlighting" for beat gestures, and iconicity) (McNeill, 2000, 2005). For instance, there is a fuzzy distinction between pointing gestures and iconic gestures (Goodwin, 2003). A pointing gesture can be used to trace the outline of what is being pointed at, and in this way, mimic the shape of a deictic point and create an iconic display at the same time. We will see later on that *iconicity* in iconic gestures can also be interpreted along the dimension of *deixis*, and this will help us to explain some of the findings presented in Chapter 4. But first, let us take a closer look at the dimension of *iconicity*.

2.4 What is Iconicity?

Iconicity is the similarity or resemblance between a form and its meaning (linguistic or otherwise) (Meir & Tkachman, 2014). Additionally, researchers have defined iconicity as a semiotic relation that comes in kinds (Peirce, 1955; Ahlner & Zlatev,

2010) and a substance that comes in degrees (Kunihira, 1971; Perry, Perlman, & Lupyan, 2015). Importantly, iconicity is the perceived analogy between the form of the sign and its meaning, as opposed to arbitrariness. In an arbitrary sign, the association between form and meaning is based solely on convention; there is nothing in the form of the sign that resembles aspects of its meaning (Meir & Tkachman, 2014). Most words in most languages are arbitrary signs, which rely on cultural conventions (de Saussure, 1916; Hockett, 1960). English speakers know that the word "painting" refers to a picture or design executed in paint that you could hang on the wall, but Dutch speakers agreed on the word "schilderij" for the same thing, which has nothing in common with the English word. Gestures can also be arbitrary. *Emblematic gestures* are highly conventionalized gestures such as waving the hand for "bye-bye". Just like language, emblematic gestures may differ per culture. For instance, the Dutch gesture for "the food is delicious" is shaking your hand next to your face with an open palm facing your cheek, whereas in Italian it is pressing your index finger slightly in your cheek while twisting it clockwise and anticlockwise.

In the previous section, we have seen that gestures can be iconic, but in some cases, language can be iconic too. In the early stages of vocabulary development, children hear quite a few iconic words that directly imitate features of the referents, such as *onomatopeia* (Bredin, 1996). Examples of onomatopeia can be found across languages, for instance, when expressing the sounds of animals in spoken words (e.g., "meow" "woof-woof", "chirp-chirp", and "quack-quack"). Because these words sound like what they mean, they show commonalities between different languages. In Dutch, animal sounds for cats, dogs, birds, and ducks are "miauw", "waf-waf", "tjirp-tjirp", and "kwak-kwak", respectively.

2.5 When do Children Become Sensitive to Iconicity?

One important question addressed in the following sections is when children become sensitive to iconicity (i.e., the similarity between the form of a sign, linguistic or otherwise, and its meaning), and perhaps also, what their level of understanding of symbolic representations is at different ages. Although this dissertation focuses on iconicity in gestures, we discuss some of the literature on iconicity in spoken language in the following sections too. Reviewing this literature could help us better understand when children become sensitive to iconicity, when they start using iconicity for word learning, and what their level of symbolic understanding is at different stages of development. To this aim, we review empirical, observational, and cross-linguistic studies.

2.5.1 Sensitivity to Sound Symbolism

The idea that people might not map sounds to shapes arbitrarily was first tested in 1929. Köhler (1929) visually presented adult participants with a spiky and a curvy shape, while they heard either the word "takete" or the word "baluma". When participants were asked "Which object is the *takete* and which is the *baluma*?", they paired "takete" with the spiky shape and "baluma" with the curvy shape. Ramachandran and Hubbard (2001) showed a similar pattern when the words "kiki" and "bouba" were used (i.e., "kiki" was paired with the spiky shape and "bouba" with the curvy shape). Somehow, adults show a strong preference for pairing spiky shapes with words that have unrounded vowels like "takeke" and "kiki" and curvy shapes with words that have rounded vowels like "baluma" and "bouba". Since then, this basic paradigm has been adapted to test if children differentiate between congruent and incongruent sound-shape mappings too.

In one of the first studies on children's sensitivity to sound symbolic words (Irwin & Newland, 1940), 9-year-old American children were presented with several sets of nonsense words and drawings, including Köhler's (1929) "takete" and "baluma" stimuli. Children consistently mapped words with rounded vowels to curvy shapes and words with unrounded vowels to spiky stimuli. Thus, this study shows that, by age 9, children are sensitive to sound-shape mappings.

Children learning Japanese are also sensitive to sound symbolism. In a study by (Imai, 2008), Japanese-speaking 25-month-olds, 3-year-olds, adults, and native English-speaking adults were asked to match six newly created Japanese sound-symbolic words to different manners of walking (e.g., *batobato* for walking with heavy steps). Each sound-symbolic word (e.g., *batobato* was shown with two videos: one video showed a congruent manner of walking (e.g., heavy steps) and the other video showed an incongruent manner of walking (e.g., light steps). Participants were asked to select the manner of walking that best matched the heard word. Even though the sound-symbolic words were completely novel, Japanese adults picked out the congruent pairs in all instances. Japanese-speaking 25-month-olds, 3-year-olds, and native English-speaking adults also consistently picked out the congruent pairs, and reliably above chance level. This study thus suggests that Japanese children as young as 25 months old can detect sound symbolism between words and actions, and this effect is not confined to the language spoken, as English adults could detect this sound symbolism too.

Furthermore, cross-cultural studies have shown that children across cultures also appreciate the same sound-shape mappings. Davis (1961) presented native Swahili-speaking children between the ages of 8 and 14 years old with abstract drawings of curvy and spiky shapes. Children were asked to match the words "takete" and "uloomu" to the drawings. Children matched the word "takete" to the spiky shape and the word "uloomu" to the curvy shape in a highly consistent manner and their choices were similar to those of a control group of native English-speaking children. Thus, this study suggests that the properties of sound symbolism may be universal.

The finding that sound-object correspondences are not perceived as completely arbitrary has also been demonstrated in younger native English-speaking children. In a study by Maurer, Pathman, and Mondloch (2006), native English-speaking adults and 2.5-year-old children were presented with a two-way forced choice task that tested whether they consistently map words with rounded vowels to curvy shapes and words with unrounded vowels to spiky shapes. For child participants, the task was set up as a game in which they had to help a puppet find his favorite toys which all had funny names. An experimenter showed the children two pictures at a time, always one picture of a curvy shape and one of a spiky shape. The experimenter then asked "Which one do you think is *bamu*?". Both children and adults chose curvy objects for words that had rounded vowels and spiky objects for words that had unrounded vowels more than chance would predict. In fact, the scores for children and adults did not differ, suggesting that native English-speaking 2.5-year-old children differentiate between congruent and incongruent sound-shape mappings as consistently as adults do.

Because adults and toddlers have had considerable experience with the word-referent mappings that exist in their environment, it is unclear whether the sound-symbolic mappings found in previous studies are the result of this experience or the product of natural habit. A more recent study shows that infants as young as 4 months old can distinguish between congruent and incongruent sound-shape mappings (Ozturk, Krehm, & Vouloumanos, 2012). In a preferential looking task, Ozturk et al. (2012) visually presented children with a curvy object and a spiky object while a word was played through a speaker. The word "kiki" formed a congruent pair with the spiky object and the word "bubu" formed a congruent pair with the curvy object. Each infant was presented with congruent and incongruent word-object pairs. On average, children looked longer at incongruent pairings (i.e., "bubu" with the spiky object or "kiki" with the curvy object) than at congruent pairings (i.e., "bubu" with the curvy object or "kiki" with the spiky object). Thus, this study suggests 4-month-old infants expect that certain sounds refer to objects with a certain shape. As such, sound symbolism may reduce referential ambiguity in word learning contexts.

More importantly, showing sensitivity to sound-shape mappings in prelinguistic infants (i.e., infants who have not acquired any words yet) (e.g., Ozturk et al., 2012), opens the possibility that sound symbolism may be a crucial factor in helping children achieve the basic understanding that words map to referents. The idea that sound symbolism helps children to gain such referential insight for speech sounds, and thereby reduces referential ambiguity in mapping words to referents (Quine, 1960), lies at the core of the sound symbolism bootstrapping hypothesis for language acquisition and evolution (Imai & Kita, 2014).

2.5.2 Sensitivity to Iconic Gestures

In the previous section, we have seen that children become sensitive to iconicity in spoken language before their first birthday, and that they consistently map sound-symbolic words to congruent shapes from 4 months of age onward, and do so in an adult-like manner by 2.5 years of age. In this section, we review empirical studies which show that children become sensitive to iconicity in gesture much later than to iconicity in spoken language. Generally, children become sensitive to iconicity in gesture around their second birthday.

Findings from a gesture comprehension task revealed that children develop a basic understanding of iconicity in gesture by 18 months of age. In a study by Namy (2008), an experimenter taught 14-, 18-, -22, and 26-month-old children actions on novel objects (e.g., using an ink roller) to subsequently test their understanding of iconic gestures derived from those actions. During test trials, the experimenter placed two objects that one would use in different ways in front of the child and elicited a choice by producing the target gesture, saying, "Which one can you get? [gesture] Can you get it? [gesture]". Children handed over the objects whose use was depicted in iconic gesture reliably above chance at 18 months, 22 months, and 26 months, but not at 14 months. Thus, this study provides tentative evidence that children have a basic understanding of iconic gestures and use those gestures to interpret speech at 18 months of age.

However, findings from another gesture and speech comprehension task suggest that children cannot accurately interpret the meaning of iconic co-speech gestures until age 3. In a study by Stanfield, Williamson, and Özçalışkan (2014), 2-4-year-old children were presented with gesture and speech combinations composed of iconic gestures that depicted an action characteristic of an object and a verbal description (e.g., a female experimenter moved her hand towards her mouth with her fingers and thumb extended in a U-shape as if eating a sandwich while she said "I am eating"). The experimenter then placed a pair of pictures depicting two different objects (e.g., a bowl of cereal vs. a sandwich) on the table and asked the child to choose the picture that matched the description (e.g., "What did I eat?"). Three-year-olds and 4-year-olds, but not 2-year-olds, selected the picture that best matched information conveyed by the gesture and speech combination. This study thus shows that children cannot fully comprehend subtle differences in meaning conveyed by iconic gestures by age 2, but they can by age 3. This is not completely in line with results by Namy (2008) that were described in the previous paragraph. However, in the study by Namy (2008), the hand shapes of the iconic gestures were very similar across objects (e.g., holding an ink roller vs. holding a hammer), but the motion of the iconic gestures was very distinct (e.g. pressing on a surface while going forward and backward vs. swinging up and down). In the study by Stanfield et al. (2014), the manner of motion was the same for depicting actions on the objects in the pictures in iconic gesture (e.g., bringing the hand to the mouth), but the hand shape was different (e.g., as if holding a sandwich vs. as if holding a spoon). The different stimuli may thus have led to the discrepancy between age groups in these studies. It may be the case that children can indeed recognize manners of motion in iconic gestures by 18 months of age, but they cannot recognize subtle differences in meaning conveyed by the hand shape of iconic gestures until age 3.

The study by Stanfield et al. (2014) showed that 3-year-old children can interpret information conveyed by iconic gesture, but not whether they can integrate information from both iconic gesture and speech. After all, children could get the correct answer in their task if they ignored the speech and just paid attention to iconic gesture. A study by Sekine, Sowden, and Kita (2015) shows that 3-year-old children can integrate information from iconic gesture and speech, but only if the gestures are produced live by an adult. In their study, 3- and 5-year-old Japanese children and adults were presented with either a video of a female producing an iconic gesture (throwing a basketball with two hands), a spoken sentence ("He is throwing"), or both. They were then required to select one of four photographs that best matched the message. The photographs showed a gesture-speech integration match (throwing a basketball with two hands), a verbal-only match (throwing a softball with one hand), a gesture-only match (opening a door with two hands), and an unrelated foil (taking a picture with a camera). The choices of 5-year-olds and adults, but not those of 3-year-olds, clearly showed that they integrated information from speech and gesture (i.e., they picked the gesture-speech integration matches). In a follow-up experiment, different 3-year-old children were presented with the same speech and gesture stimuli as in the first task, but this time they were produced live by an experimenter. When presented live, 3-year-olds could integrate information from speech and gesture. Taken together, these studies show that children develop a good understanding of iconic co-speech gestures (live) produced by adults by age 3.

It is noteworthy that in all the studies described in this section, iconic gestures depicted action characteristics of objects and children had to imagine these objects to correctly interpret the gestures. That is, the iconic gestures often depicted how an object is used, without the hands representing the object's physical features or attributes. For example, one could produce an iconic gesture for the object "tennis ball" by producing a hand movement as if throwing the ball, but one could also shape the hand like a fist as if the hand takes the physical shape of the tennis ball. These two gestures present children with different representations of a tennis ball. One could argue that the direct mapping of the fist to the physical shape of the ball may be easier to understand because it is less imaginative, and the indirect mapping of the hand shaped as if throwing the ball may be more difficult to understand because it is more imaginative. On the other hand, if the iconic gesture depicts the use of the object it places a meaning in a context and this may be easier to understand for children than just an iconic shape gesture. Nevertheless, experimental studies which investigated children's comprehension of these two types of iconic gestures (i.e., handling objects vs. object attributes) indicate that there may be differences in comprehension at age 2.5, but that children are equally good at understanding both types of iconic gestures at age 3 (Hodges, Özçalışkan, & Williamson, 2015), and have an adult-like ability to recognize the meaning of iconic gestures at age 5 (Tolar, Lederberg, Gokhale, & Tomasello, 2008).

2.6 The Role of Iconicity in Early Word Learning

In this section, we ask at which stages in development children can use iconicity in spoken language and gesture for word learning. We present evidence from empirical studies that iconicity in spoken language and gesture can help children to map words to their referents, but also that iconicity in each modality can help children to form a semantic representation that is ready for generalization.

2.6.1 Word Learning with the Help of Sound Symbolism

Researchers commonly use associative learning tasks to test if sound symbolism facilitates children's word leaning. During such tasks, children are first habituated to word-referent mappings, followed by test trials that present children with two options on a split screen (the target meaning and a distractor) while they hear a target word. Children's looking times to each of these options can reveal if children retained the mappings they were trained on (i.e., if children look longer at the target meanings than at the distractors when the words are heard). In a study by Imai et al. (2015), Japanese 14-month-old infants were either trained on congruent or incongruent sound-shape mappings. The stimuli looked like the orginal stimuli used by (Köhler, 1929), but the novel words used were "moma" for the curvy shape and "kipi" for the spiky shape. After children habituated to either congruent or incongruent mappings, their knowledge of word-referent mappings was measured using looking times to curvy and spiky shapes on a split screen while they heard "Kipi (or moma)! Which is the kipi (or moma)?". Infants in the congruent condition looked longer at the target objects on test trials than infants in the incongruent condition. This study thus suggests that sound symbolism boosts learning and retention of word-referent mappings.

Learning the novel name for an entity is one step of the word learning process, but an important next step is to generalize the newly learned name to novel events showing the referent meaning. In a study by Imai, Kita, et al. (2008), Japanese 3-year-old children's ability to generalize newly learned verbs to novel events was tested. Children were either taught novel sound-symbolic verbs (words like "chokachoka" for fast walking with small steps) or novel non-sound-symbolic verbs (e.g., nonsense like "chimoru" which follow Japanese phonotactical rules) while they viewed a training phase in which an actor performed a movement. After the training phase, children were required to generalize each newly learned verb by pointing at one of two novel events that they thought corresponded to its meaning: one showed the referent action performed by a novel actor and the other showed the actor from the training phase performing a novel action. Children who were taught novel sound-symbolic verbs, but not children who were taught novel non-sound-symbolic verbs, successfully generalized the newly learned verbs to novel events that showed the referent actions. However, it could be the case that the children in the novel sound-symbolic verb condition matched the sounds of those verbs to the actions at the test stage, without any consideration of which test event the verb learned in the training phase could be generalized to. In a follow-up experiment, a different group of children were taught novel verbs that did not sound-symbolically match the movement in the training phase, but the distractor movement in the test phase (i.e., the same actor as in the training phase performing a novel action). Children in the control experiment performed at chance in the generalization task. This ruled out the possibility that the 3-year-olds in the novel sound-symbolic verb condition of the first experiment simply matched the sound to the action without being engaged in verb learning. Thus, this study shows that sound symbolism facilitates early verb learning.

Imai et al.'s (2008) Japanese sound-symbolic verbs even facilitate verb learning in children who are learning a language from a different language family than Japanese. A more recent study by Kantartzis, Imai, and Kita (2011) showed that English-speaking 3-year-old children performed better in a verb generalization task with Japanese sound-symbolic verbs than with non-sound-symbolic verbs, just like Japanese-speaking children. This study thus demonstrates the universal nature of non-arbitrary links between sound and meaning. Similarly, Yoshida (2012) found that both English and Japanese 2-4-year-old children both successfully generalized novel sound-symbolic verbs (compared to novel non-sound-symbolic verbs) to events showing the referent actions in a novel verb learning task. Some researchers argue that sound symbolism should be recognized as a general property of language, based on the argument that sound-symbolic mappings are also consistently present across many different languages (Perniss, Thompson, & Vigliocco, 2010). Recognizing a non-arbitrary property of language goes against more traditional views of language which assume that language is a symbol system, and words point to their meanings by convention (de Saussure, 1916; Hockett, 1960).

2.6.2 Word Learning with the Help of Iconic Gestures

Children can use iconic gestures for word learning between the ages of 18 and 24 months old. In a case-study by McGregor, Rohlfing, Bean, and Marschner (2009), 18-24-month-old children participated in one of three training phases to enhance their understanding of the preposition *under*. As this is an existing English word, the children were pre-tested on their understanding of the word. After the pre-test, the experimenter placed pairs of objects (e.g., a toy boat and a bridge) on a table in front of the children, instructing them to "Put the boat *under* the bridge". With this request, children either saw an iconic gesture for under (e.g., the experimenter held her right hand over her left hand then moved the right hand *under* the left), a picture of objects in the under relationship (e.g., a photograph of a boat *under* a bridge), or no support at all. Children were tested on their understanding of the word *under* immediately after the training phase, and after two or three days' delay. The stimuli involved object pairs from the training phase and novel object pairs for testing generalization of the word *under*. Only children who saw an iconic gesture that depicted the relation *under* showed a significant increase in generalization performance from pretest to delayed posttest. As this study had no control group in which children saw a non-iconic gesture, we do not know whether the effect of seeing iconic gestures on word learning can be attributed to information conveyed by iconic gesture or to the engaging function of seeing a gesture. Moreover, many of the kids who took part in this study showed a good understanding of the word *under* in the pre-test, so engaging the children in the task could have led to increased performance.

Children also retain more words from a foreign language when these words are taught with iconic gestures that depict the word meanings. In a study by Tellier (2008), 5-year-old native French-speaking children were taught English common nouns (e.g., "house") and verbs (e.g., "swim") while either seeing iconic gestures or pictures that depicted the word meanings. Both groups could point out the meaning of the learned words in a picture naming task, but the iconic-gesture group outperformed the picture-group in an immediate and delayed word production task. Thus, this study shows that iconic gestures help children to gain an understanding of word meaning, but also lead to better production of those words, which is also a crucial step of the word learning process. However, children in the iconic-gesture group were asked to reproduce the iconic gestures they saw with each word, so we do not know how much of the effect of iconic gesture on word production can be attributed to seeing iconic gestures, rather than to producing iconic gestures. In fact, it is hard to ascribe the effects of iconic gesture to the gesturally depicted meanings, because there was no movement of the hands or body in the picture-group, meaning that a trace in motor memory could have led to the found effects.

Furthermore, seeing iconic gestures influences how children interpret novel verb meanings. In a study by Goodrich and Hudson Kam (2009), 3- and 4-year-olds and adults participated in a novel verb learning task. An experimenter placed two toys that operated in different ways in front of the participants. Participants were shown how a puppet named Sam could operate the different toys. Then, the experimenter said "One of the toys lets my puppet Sam (*meek* + gesture). The other toy lets Sam (*dack* + gesture). Which toy lets Sam go (*meek*-ing)?". The iconic gestures produced by the experimenter depicted the actions that the puppet Sam could perform on the toys (e.g., rolling down a ramp was indicated with the index finger producing a downward spiral movement). Note that the experimenter did not gesture when the children were asked to pick a toy. Both groups of children and adults successfully picked out the toys that corresponded to the meaning of the novel verb. Thus, this study shows that by age 3, children can learn verb meanings from iconic gestures, because the iconic gestures were the only source of information that children could use to interpret the novel verb meanings.

So far, we have seen that iconic gestures can help children to map a word to its referent, but the next study shows that 3-year-old children can also generalize newly learned words to novel events when those words were taught while children saw the word meanings depicted in iconic gestures. In an experimental study by Mumford and Kita (2014), 3-year-old English-speaking children were children were taught made-up verbs (e.g., "to dax") that could be interpreted as manner verbs (e.g., "to push") or change-of-state verbs (e.g., "to break"), while the experimenter produced either manner gestures, end-state gestures, or no gesture at all. Manner gestures depicted the manual action of creating a shape (e.g. sprinkling sand onto a piece of paper in the shape of a square) and end-state gestures traced the outline of the final shape that was being created (e.g., a square). Children who saw manner gestures when the verbs were taught were more likely to generalize the newly learned verb to an event showing the same manner than to an event showing the same end-state, and the reverse was true for children who saw end-state gestures when the verbs were taught. Thus, iconic gestures helped children to learn and generalize verbs. Importantly, the findings show that children's semantic representation of the verbs included the aspects of the action events that were depicted in iconic gestures.

2.7 Development of Spontaneous Gesture Production

This section discusses the developmental stages at which spontaneous deictic gestures, iconic gestures and interactive gestures emerge, which are often seen as important milestones for communicative development. This is evident from the fact that many standardized communicative development measures now include checklists of the gestures that children understand and produce, besides checklists of words that children understand and produce (e.g., Alcock, Meints, & Rowland, 2017; E. Bates et al., 1986; Fenson et al., 2006).

One of children's first milestones is to indicate objects in their direct visual environment by means of pointing gestures. Children typically start producing such pointing gestures between 9 and 12 months of age (E. Bates, 1976; E. Bates, Benigni, Bretherton, Camaioni, & Volterra, 1979), although it has been reported that children of 8-month-old point too (e.g., Butterworth & Morisette, 1996). Thus, children typically produce pointing gestures before they produce their first words. Between 9 to 12 months of age, children mainly produce imperative pointing gestures ("I want that") and declarative pointing gestures ("Look at that!") to indicate interesting entities in their direct visual environment (E. Bates, 1976). These pointing gestures are an important developmental milestone because they show that young children have communicative intentions. Children produce imperative pointing gestures to request an object from the addressee and declarative pointing gestures to share attention to an object with the addressee (E. Bates, 1976). Children's pointing gestures thus show that they have the specific intention to convey a message to an addressee.

Spontaneous pointing gestures have also been related to early language development. Butterworth (2003) referred to pointing as the "royal road to language" (p. 9), because when a child looks in the direction of a point while an adult speaker utters the label of the indicated entity, the link between the word and its referent is established through a learned association between what is heard and seen. The next developmental milestone is children's production of pointing gestures in combination with speech to indicate entities in their direct visual environment. Iverson and Goldin-Meadow (2005) investigated whether pointing gestures merely precede language development or are indeed fundamentally related to it. In a longitudinal study, they followed children's language development from 10 to 24 months of age. Children were videotaped monthly during play sessions and meal times with their primary caregivers. Instances in which children referred to unique objects per session were coded using three categories: speech only (e.g., "bird"), gesture only (e.g., points at bird), and gesture plus speech (e.g., points at bird and says "bird"). Objects that children first indicated with gesture only were objects that they subsequently produced words for. On average, children pointed at an object three months before they first produced the label for this object. Moreover, children's use of supplementary gesture and speech combinations to indicate referents (e.g., point at a bird while saying "eat") predicted their onset of two-word utterances (e.g., "bird eat"), approximately two months later. This study thus shows that pointing gestures play a facilitating role in early language development.

The next milestone we discuss here is children's spontaneous production of iconic gestures. The earliest record of children's spontaneous iconic gestures is at 14 months of age (Özçalışkan, Gentner, & Goldin-Meadow, 2014), but in this observational study with English-speaking children, gestures children produced at 14 months of age did not accompany speech. In fact, in this particular study, iconic co-speech gestures did not occur frequently until the age of 26 months. The earliest record of children's spontaneous iconic *co-speech* gesture production is at 19 months of age. In a longitudinal study by Furman, Küntay, and Özyürek (2014), spontaneous speech and co-speech gestures of eight Turkish-speaking children aged one to three were recorded. From 19 months of age onward, children produced iconic gestures that encoded the manner or path of motion events children expressed in their speech. It is important to interpret these findings in the context of the language the children were learning. Evidence that English-speaking children also produce iconic gestures at 19 months of age is scarce. Studies on English-speaking children often do not report any use of spontaneous iconic co-speech gestures until age 2. For example, Marentette, Pettenati, Bello, and Volterra (2016) coded the spontaneous iconic co-speech gesture use of 2-year-old Italian- and English-speaking children during a picture naming task. Both groups of children produced iconic gestures that depicted how an object was used (e.g., acting on the object) or what it looked like (e.g., its shape or size), but the gesture rate of English-speaking children was low compared to Italian-speaking children. Özyürek et al. (2008) investigated whether differences in the language spoken affects the spontaneous iconic co-speech gestures of Turkish and English, which are typologically different languages. Children at ages 3, 5, and 9 were asked to produce short narratives of videos that showed motion events (e.g., a tomato rolled down the hill). The spontaneous gestures that children produced when narrating these events were coded for manner (e.g., rolling) and path (e.g., down). English- and Turkish-speaking children's gestures looked similar at ages 3 and 5 (i.e., they encoded manner and path in separate gestures), differing from each other only at age 9 and in adulthood (i.e., English speakers encoded manner and path in one conflated gesture, but Turkish speakers encoded manner and path in separate gestures). Taken together, these studies suggest that spontaneous iconic co-speech gestures emerge at varying ages depending on the culture and the language spoken, but generally these gestures emerge between the ages of 18-24 months.

Previous research suggest that spontaneous production of interactive gestures emerges before age 5, but studies on spontaneous interactive gesture production in children are sparse. In a study by Coletta et al. (2014), the spontaneous production of interactive gestures, among other co-speech gestures, was investigated in 5- and 10-year-old native English-speaking children. Participants were asked to watch a short video of a wordless cartoon (Tom and Jerry) and to retell the story it depicted in front of a camera. The task was set up in a monologue format, but an experimenter was there when children narrated the story. Recordings of the children's narrations were coded for speech and gesture and the gesture rate per clause was calculated for each child. It was found that 5-year-old children produced more interactive gestures per clause than 10-year-old children. Coletta et al. (2014) suggest that 5-year-old children find it difficult to narrate a story in a monologue format. The gestures children produced reflected these difficulties and often prompted the experimenter to scaffold their narrative production. As such, 5-year-old children's interactive gestures in this study mainly reflected their constant move back towards a dialogue format. This study thus shows that by age 5, children adequately produce interactive gestures to prompt adults to scaffold their speech production. As such, children's use of interactive gestures may scaffold their story-telling, speech fluency, and perhaps even language learning. More research is needed, however, to investigate if children also spontaneously produce interactive gestures at this rate in different dialogue formats, if these gestures can be observed in younger age groups, and if children's interactive gesture production is related to language learning.

2.8 Gesture as a "Window" into the Mind

The gestures that children produce are not just important milestones for the development of their communicative abilities, but can also provide insight into how their cognitive abilities are developing. One of the leading views in gesture studies is that gestures can reveal the mental process of the speaker-gesturer (McNeill, 2000). This view is based on the idea that somehow gesture, speech, and thought are inherently related to each other. This section presents four lines of empirical evidence that support this view and explore the nature of the relationship between gesture, speech, and thought.

First, studies in which gesture was elicited without speech provide a window into children's level of symbolic understanding. For instance, 3- and 4-year-old children produce iconic gestures that are less symbolic than iconic gestures produced by 5-year-old children. In a study by Boyatzis and Watson (1993), 3-, 4-, and 5-year-old children were asked to pretend to use eight common objects (e.g., "Use your hands to pretend you are brushing your teeth with a toothbrush"), without those objects being present. There was a developmental shift in the symbolic quality of the produced gestures, where 3- and 4-year-olds used their body parts to represent objects (e.g., using an extended index finger to represent the toothbrush) and 5-year-olds used imaginary gestures (e.g., using a hand shape as if holding a toothbrush). Body part gestures represent salient aspects of the form of an object and were therefore considered less symbolic than imaginary gestures, which represent the use of an imaginary object but not the form of the object itself. In a follow-up experiment, the same 3-year-old children were asked to imitate the experimenter's gestures, who produced iconic gestures different from the gestures the children produced themselves in the first part of the task. That is, if the child produced a body part gesture before, then the experimenter would produce an imaginary gesture, and vice versa. Three-year-olds found it particularly difficult to imitate an imaginary gesture if they had produced a body part gesture themselves, which suggests that they struggle with gestural representations that exceed their own symbolic level.

Similarly, in a study by Overton and Jackson (1973), children under the age of 6 years pretended to use objects (e.g., a comb) by producing gestures in which their hands took on the form of the objects itself (e.g., fingers shaped as the teeth of the comb) and children over the age of 6 years produced gestures in which they pretended to hold and use objects (e.g., fingers shaped as if holding an imaginary comb). Taken together, these studies show that gesture offers researchers a window into children's level of symbolic understanding, by allowing them to "see" what children choose to represent with their hands.

Second, naturalistic observations of co-speech gestures reveal how people internally organize concepts such as left and right. Kita and Eggesbey (2001) observed naturalistic situations of giving route directions in Ghana. In the Ghanaian culture, pointing with the left hand is considered to be a taboo. Especially when talking to strangers or the elderly, there is a politeness convention to hide the left hand from the interlocutor by placing it on the lower back. The study showed that Ghanaian speaker-gesturers often engaged in an anatomically straining position to indicate directions towards the left-hand side with right hand pointing gestures. The key finding was that during verbalization of the concept left in combination with a right-hand finger point, left-hand pointing gestures were not fully suppressed. Ghanaian speaker-gesturers often produced a small gesture with their left hand, which was positioned on their lower back outside the addressee's view, before they gestured to the left with their right hand. This use of gesture thus shows that concepts such as left and right are largely grounded in sensorimotor experiences. In other words, gesture production is influenced by how people think of using the body to interact with the physical environment (Kita, 2000).

Third, spontaneous co-speech gestures elicited in tasks in which people were asked to describe concepts of time reflect how people internally organize information about time. A study by Gu, Mol, Hoetjes, and Swerts (2017) demonstrated that Chinese-English bilinguals produce more gestures along the vertical axes when talking about Chinese time references with vertical spatial metaphors than when talking about time conceptions in the English translations, and when talking about Chinese time references with no spatial metaphors. These findings thus show that when speaking Chinese, Chinese-English bilinguals organize their thoughts about time vertically, corresponding with the ancient Chinese measures of time such as the hour glass in which the sand runs from top to bottom, as well as incense, which burns from top to bottom. Thus, this study shows that gesture revealed how time is conceptualized in the mind during the moment of speaking.

Moreover, spontaneous co-speech gestures in tasks in which people were asked to narrate a story reveal how people organize verb clauses. Kita and Özyürek (2003) asked English speakers, Japanese speakers, and Turkish speakers to describe events shown in a short video clip of the cartoon Tweety and Sylvester. In the stimulus video, Sylvester swallowed a bowling ball and rolled down the hill into a bowling alley. Many of the speakers mentioned this event when they narrated the story to an experimenter. Turkish speakers and Japanese speakers described the event of Sylvester rolling down the hill using two verb clauses: one to express the "spinning/rolling" motion and another to express the "descending/downward" direction or path. English speakers used only one verb clause to express the same event: "rolling down". The differences in the number of verb clauses that the speakers used was also reflected in the number of gestures they used. Turkish and Japanese speakers produced one gesture per verb clause: one spinning hand or finger movement for the motion and another downward hand movement. English speakers, however, produced one conflated gesture that depicted the spinning motion and downward path simultaneously. Thus, the gestures of Turkish speakers, Japanese speakers, and English speakers reflect how language is organized in their mind during the moment of speaking. These findings are consistent with the idea that gestures, together with speech, can constitute thought (McNeill, 1992). Even if gestures and speech are revealing the same thoughts (i.e., they show equivalent information), it is important to analyze both modalities because gesture and speech convey information in different representational formats. Gestures present information in an analog, imagistic format, whereas speech presents information in a discrete, segmented format (McNeill, 1992).

Fourth, gestures can also convey unique ideas that are different from the ideas presented in accompanying speech. In a study by Church and Goldin-Meadow (1986), 5-8-year-old children were asked to judge Piagetian conservation problems. In one of the trials children were asked whether two rows of coins had the same or a different number of coins. This question was asked once after the experimenter had perfectly aligned the two rows of coins in front of the children and once after the experimenter had spread out the coins in one row to make it wider than the other row in front of the children. Most children answered that the number of coins in each row was the same when the rows were aligned, but different when one row of coins was spread out to be wider than the other. When children were asked to explain their second answer they said "because one row is wider", and they spontaneously produced gestures congruent with their speech (e.g., indicating the width of rows) or incongruent with their speech (e.g, spreading out the fingers). Children who spread out their fingers followed the alignment between the two rows of coins, which should have led them to the answer that the rows contained the same number of coins. When there was a mismatch between gesture and speech, two beliefs about the same problem were simultaneously expressed–one in gesture (i.e., alignment is key to the answer) and another in speech (i.e., width is key to the answer). Thus, this study shows that gesture can also offer a unique window into a child's mind, importantly, one that is different from speech. This is important because the simultaneous activation of multiple beliefs suggests that a child is in a transitional state and ready to learn (Goldin-Meadow et al., 1993). In Chapter 7 we present a suggestion for future research which relates this "readiness to learn" conveyed by gestures that children produce themselves to the research presented in this dissertation.

The empirical chapters that follow now are based on self-contained manuscripts which are either published, accepted for publication, under review at a scientific journal, or being prepared for publication. Therefore, each chapter has its own abstract, introduction, and discussion. As most references overlap between the chapters, all references are included in a reference list at the end of this dissertation.

Chapter 3

GestuRe and ACtion Exemplar (GRACE) Video Database

Human locomotion is a fundamental class of events, and manners of locomotion (e.g., how the limbs are used to achieve a change of location) are commonly encoded in language and gesture. To our knowledge, there is no openly accessible database containing normed human locomotion stimuli. Therefore, we introduce the GestuRe and ACtion Exemplar (GRACE) video database, which contains 676 videos of actors performing novel manners of human locomotion (i.e., moving from one location to another in an unusual manner) and videos of a female actor producing iconic gestures that represent these actions. The usefulness of the database was demonstrated across four norming experiments. First, our database contains clear matches and mismatches between iconic gesture videos and action videos. Second, the male actors and female actors whose action videos matched the gestures in the best possible way, perform the same actions in very similar manners and different actions in highly distinct manners. Third, all the actions in the database are distinct from each other. Fourth, adult native English speakers were unable to describe the 26 different actions concisely, indicating that the actions are unusual. This normed stimulus set is useful for experimental psychologists working in the language, gesture, visual perception, categorization, memory, and other related domains.

3.1 Introduction

Human locomotion (e.g., movement of the human limbs to change location) is a topic widely studied in the field of experimental psychology. For instance, expressions of human locomotion have been studied in spoken language (e.g., Malt, Gennari, Ameel, Tsuda, & Majid, 2008; Slobin, Ibarretxe-Antuñano, Kopecka, & Majid, 2014; Malt et al., 2014), written language (e.g., Slobin, 2004, 2006), sign language (e.g., Supalla, 2009; Slobin & Hoiting, 1994), and gesture (e.g., Özyürek & Kita, 1999; Kita & Özyürek, 2003; Özçalışkan, Lucero, & Goldin-Meadow, 2016). Also, in many word learning experiments, researchers teach children verbs for novel manners of human locomotion (e.g., Mumford, 2014; Mumford & Kita, 2014; Imai, Kita, et al., 2008; Scott & Fisher, 2012). In memory experiments, locomotion stimuli are often used to study visual memory of agents and their actions (e.g., J. Wood, 2008). In categorization experiments, human locomotion is used to study, inter alia, how children perceptually categorize manners of locomotion (e.g., Salkind, 2003; Salkind, Golinkoff, & Brandone, 2005; Pulverman, Hirsh-Pasek, Golinkoff, Pruden, & Salkind, 2006).

Particularly in studies on verb learning, human locomotion stimuli are often used along with iconic gestures. Iconic gestures (McNeill, 1992) represent actions, motions or attributes associated with people, animals, or objects (e.g., wiggling the index and middle fingers to represent a person walking; tracing a shape). Researchers have investigated whether novel verb meanings are shaped by iconic gestures that are shown when the verb is taught (e.g., Spencer, McDevitt, & Esch, 2009; Goodrich & Hudson Kam, 2009; Mumford, 2014; Mumford & Kita, 2014).

Developing human locomotion stimuli can be very laborious. Nevertheless, most researchers develop such stimuli solely for the purpose of their own research. As a consequence, there is no openly accessible video database containing manners of human locomotion and iconic gestures that represent these manners.

3.2 The Current Research

3.2.1 Contents of the GRACE Video Database

We developed and normed the GestuRe and Action Exemplar (GRACE) video database, which includes 676 videos of 26 actors (13 males, 13 females) performing 26 *novel* manners of human locomotion (i.e. moving from one location to another in an unusual manner), and 26 videos of a female actor who produces iconic gestures that represent these manners. Figure 3.1 presents three examples of the gestures and the corresponding manners of locomotion (in the upper right corner of each panel). The gesturing hands represent the actor's feet (panel A), the actor's legs (panel B), and the actor's whole body (panel C).

The GRACE video database is openly available from the Warwick Research Archive Portal at `http://wrap.warwick.ac.uk/78493`. Along with 676 video files,

Figure 3.1: Three panels (A, B, and C) with cropped stills of videos in which a female actor gestures iconically to represent the manners of human locomotion performed by actors in the upper right corners of the panels. In the actual norming study, the action video and the gesture video had the same size and were presented side-by-side. Gestures and actions are included in separate video files in the database. From left to right the panels show the following gesture videos: "00F_scurrying.mp4", "00F_mermaiding.mp4", and "00F_twisting.mp4", and action videos: "01F_scurrying.mp4", "09F_mermaiding.mp4", "01M_twisting.mp4".

we have made the raw data from our norming studies available and the Python scripts that we used to process the data. We also included a manual that contains guidelines on how to use the GRACE video database.

3.2.2 Norming the GRACE Video Database

In this section, we identify and motivate four essential requirements for the type of stimuli in the GRACE video database. These requirements guided the design of our norming studies to assure its usefulness for experimental psychologists. The GRACE video database is particularly useful for researchers who need unusual human locomotion stimuli to study language and gesture, memory, and categorization. Below, we discuss the implications of each norming study in the context of these research areas.

First, the GRACE video database includes videos that were normed for the degree of match between action pairs and matching and mismatching iconic gestures. Many experiments in developmental psychology use two-way forced choice tasks. In such tasks, pairing actions that would appear as two distinct choices is important. The design of our first norming experiment is motivated by this future use. Also, pairing actions made data collection for this study more manageable; if we did not pair, participants would have to rate a large number of action-gesture combinations that make "mismatches". Action pairs with matching and mismatching gestures could be used in experiments with a two-way forced choice task in which one of the actions is congruent with gesture, but the other is incongruent. This is useful for research on word learning with the help of iconic gestures (e.g., Mumford & Kita, 2014; Mumford, 2014; Özçalışkan et al., 2014; Goodrich & Hudson Kam, 2009), the intake of information conveyed by gesture and speech (e.g., McNeill, Cassell, & McCullough, 1994; Cassell, McNeill, & McCullough, 1999; Özyürek et al., 2007), and memory recall for sentences with the help of gesture (e.g., Feyereisen, 2006; Madan & Singhal, 2012). Furthermore, these stimuli are useful for studies on processing gesture-speech combinations, in which researchers often manipulate the semantic relations between the two channels (i.e., gesture and speech match, mismatch, or complement each other) (e.g., McNeill et al., 1994; Cassell et al., 1999; Özyürek et al., 2007; Spencer et al., 2009). Thus, the first norming study tested matches and mismatches between iconic gestures and manners of human locomotion in all the 676 action videos. We then ran an algorithm over the norming scores to identify the best possible matches between iconic gestures and actions performed by male actors and female actors, separately. This led to a one-to-one assignment of male actors and female actors to action pairs. Action videos of the selected actors were used in the next norming study.

Second, GRACE contains videos that were normed for the similarity of the same actions within action pairs performed by male actors and female actors and the (dis)similarity of the different actions within action pairs performed by male actors and female actors. Researchers who introduce an actor-change in their experimental task (e.g., to test actor memory or verb generalization) often do this by changing between male actors and female actors, as they have naturally distinct appearances (e.g., Mumford, 2014). For instance, word learning studies that take an exemplar-based approach could use videos that show different actors performing the same actions and the same actors performing different actions (e.g., Maguire et al., 2002; Maguire, Hirsh-Pasek, Golinkoff, & Brandone, 2008; Scott & Fisher, 2012). Videos that show different actors moving in the same manner could also be useful for creating generalization tasks to test people's understanding of locomotion verbs (e.g., Imai, Kita, et al., 2008), and recognition tasks and change-detection tasks to test their memory of actors (e.g., Imai et al., 2005; J. Wood, 2008). In all these tasks it is important that the manner of human locomotion is similar across the actor-change. Thus, the second norming study tested how similar male actors and female actors perform the same actions within action pairs, and how distinct each male actor and female actor performs the two different actions within action pairs. All actions that are included in the database were normed in this study, but participants rated only the videos of male actors and female actors who were assigned to an action pair because their performance matched corresponding gestures very well in the first norming study.

Third, GRACE includes 26 actions which were normed for how distinct they are compared to every other action in the database. In this norming study, we let go of the notion of action pairs to obtain a measure of distinctiveness for all the actions in the database. There are three advantages of using this approach. First, norming the distinctiveness between all 26 actions is useful for studies on the ways in which people can categorize various semantic components of motion verbs such as figure (e.g., the man, the woman, Pulverman et al., 2006) and manner (e.g., Salkind, 2003; Salkind et al., 2005). Second, such norms are useful for studies on infants' ability to discriminate manners of motion (e.g., Pulverman, Golinkoff, Hirsh-Pasek, & Sootsman Buresh, 2008; Pulverman, Song, Golinkoff, & Hirsh-Pasek, 2013), which use change-detection tasks with more than two options (e.g., four actions presented to participants on each quadrant). Third, the manners of locomotion that are shown to one participant need to be highly distinctive from each other to avoid confusion in any given task. For example, if a participant is taught a novel label for a locomotion manner in a word learning task, then this manner should be distinct from all manners that are subsequently labeled to avoid a bias in test performance. Therefore, the third norming study tested the similarity between all combinations of actions to obtain a measure of distinctiveness for each action in the database. In this norming study, human raters were presented with a subset of the videos from the database, in which each video showed one of the 26 actions performed by either a male or female actor.

Finally, the 26 actions in the GRACE video database were normed for how accurately and concisely they can be described by adult native English speakers. We asked whether the English language contains existing single-word or multi-word labels for the actions, which we used as a measure of how unusual the actions are. It is important that the stimuli are unusual to ensure that a given task performance occurs as a function of an experimental manipulation and not as a consequence of participants being familiar with the stimuli prior to the task. This is important for language research: if a participant already knows a label for an action action that is labeled in a word learning task, then this may cause a bias in test performance. It is also important for memory research: if people commonly perform these actions in real life, then this may cause a bias in test performance. Therefore, the fourth experiment assessed how accurately and concisely each action can be described by adult native speakers of English. Participants described the 26 actions in the database based on the same set of videos as in the third norming study.

3.2.3 General Methods for Developing the GRACE Video Database

The GRACE video database originated in work by Mumford and Kita (2014) and Mumford (2014), who developed 14 unusual manners of human locomotion and iconic gestures representing these manners. GRACE includes these 14 manners and 12 additional manners of human locomotion and corresponding iconic gestures, resulting in a total of 26 manners and gestures.

Action videos

We recruited 13 male actors between 22–40 years old (M = 27.00, SD = 4.98) and 13 female actors between 20–42 years old (M = 27.08, SD = 6.36). The national origin of the actors varied from British, Czech, Japanese, Polish, Dutch, Indian, Irish, German, Canadian, Nigerian, Mauritian, Bulgarian, Pakistani, Singaporean, Malaysian to Chinese. All actors were educated to the university degree level.

Actors participated in individual recording sessions. They were instructed to keep their arms and hands by their side when performing the actions, because we needed the hand gestures of the female actor to unambiguously represent the actors' feet, leg, and body movements. Actors were also required to carry out each action as an ongoing motion without any breaks.

Prior to recording each action, actors watched an example video of a model. The videos of the model were not included in the database so that all actors shared the same reference point when performing the actions. Subsequently, the actors were required to move across the length of a scene in the same manner as the model. The starting point and the ending point were marked on the floor just outside the camera view. Each action was recorded at least twice from a distance of approximately 4.5 meters. If actors struggled with one of the actions, the researcher showed them their last recorded video and practised the movement with them repeatedly until they were ready to record again. Every recording session lasted approximately 1 hour. Informed written consent was obtained at the end of each recording session.

Gesture videos

Hand gestures of a female actor were recorded from a distance of approximately 1.5 meters. This actor watched the video recordings of the model performing an action prior to recording the gesture that was designed to match this action. Gestures were designed by the researchers based on the definition of iconic gestures by McNeill (1992) so that the form of gesture resembled the referent action.

Specifically, all gestures iconically represented the body part that was most prominent for each movement (i.e., feet, legs, or whole body), its dynamic shape, and the rate at which the movement was carried out. Gestures representing the whole body were performed with the right hand. Gestures representing the legs were performed by both hands, where the right hand represented the right leg and the left hand represented the left leg. Gestures representing the feet were performed with the fingers, where the right hand fingers matched the right foot and the left hand fingers matched the left foot.

3.2.4 Apparatus

Videos were recorded using a Canon Legria HFR56 camera with autofocus in a room with controlled light settings. Recordings were muted, cut, optimized for HTML, and converted to MP4 files of 640 x 480 pixels using *avconv* on Linux. The total size of the GRACE video database is 185 mega-bytes.

3.3 Experiment 1

The first experiment tested the degree of match (and mismatch) between iconic gestures and manners of human locomotion. During the development of the database, 26 iconic gestures were created that matched each action. A mismatch between iconic gestures and actions was set up in the following way. Every action was paired up with another action from the set to create 13 action pairs (see Table 3.1). We then showed participants each action with a matching iconic gesture, but also with the iconic gesture that was created for the other action in the action pair as a mismatching iconic gesture. Participants rated these matches and mismatches on a seven-point scale.

We predicted that match ratings for matching iconic gestures and actions would be higher than match ratings for mismatching iconic gestures and actions. Additionally, we predicted that matches would be rated higher than the neutral score on a seven-point scale and that mismatches would be rated lower than the neutral score.

3.3.1 Method

Participants

We recruited 301 individuals (183 males, 117 females) from the university's online participant pool. Eight participants were excluded from further analyses because they indicated that the videos did not display, or run smoothly. The final participant sample included 293 individuals (179 females, 113 males) between 18–67 years old ($M = 22.19, SD = 6.66$). The majority of participants reported English as their native language (58.7%), followed by Asian languages (23.2%), and other Indo-European languages (18.1%). Participants automatically entered a lottery for an Amazon voucher upon completing the task.

Materials

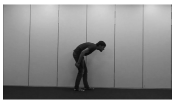
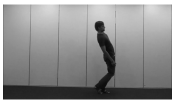
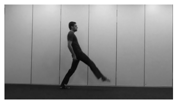
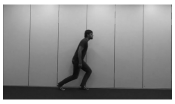
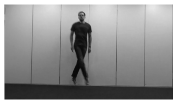
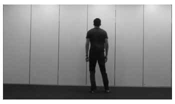
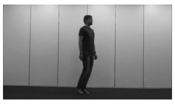
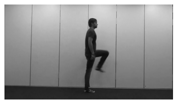
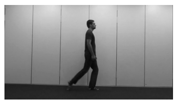
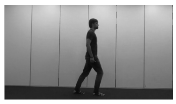
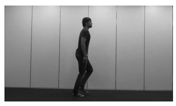
We used videos of 26 manners of locomotion carried out by 26 actors (676 videos in total), and 26 videos of a female actor producing iconic gestures. Actions were organized in pairs (see Table 3.1) so that matches and mismatches between iconic gestures and actions could be created. Figure 3.2 shows the matches and mismatches between iconic gestures and actions for action pair 1. For instance, participants were shown *bowing* with a *bowing gesture* (Panel A), *bowing* with a *skating gesture* (Panel B), *skating* with a *skating gesture* (Panel C), and *skating* with a *bowing gesture* (Panel D).

We created 26 batches of videos to keep the length of the experiment reasonable. Each video batch contained videos of the 26 actions, but performed by different actors to ensure that all 676 action videos appeared in one of the batches. Each action video was combined with a matching and mismatching gesture video within a batch, which resulted in 52 trials. Each action video–gesture video combination was rated by on average 23 participants (range=18 to 28).

Procedure

The experiment was set up in a web-based environment. Participants signed a digital consent form and were asked for demographic information. The instruction page showed participants a still frame of a gesture video and a still frame of an action video from the model as an example of a very good match. Participants were then shown two videos side-by-side, which started playing on loop automatically when a trial started. Participants were instructed to rate the match between the hand

Table 3.1: Twenty-Six Manners of Human Locomotion Organized in Action Pairs

Pair	Action *a*	Still frame	Action *b*	Still frame
1.	bowing		skating	
2.	wobbling		marching	
3.	mermaiding		overstepping	
4.	creeping		crisscrossing	
5.	turning		hopscotching	
6.	swinging		skipping	
7.	jumping		crossing	
8.	dropping		folding	
9.	twisting		stomping	
10.	trotting		hopping	
11.	flicking		dragging	
12.	grapevining		shuffling	
13.	groining		scurrying	

Note. Still frames are taken from the videos of the male actor whose videos file names start with "08M_". Short-hand action labels are used to refer to the manners of locomotion and follow after the underscore in the file names of the database (e.g., "08M_bowing.mp4", "08M_skating.mp4").

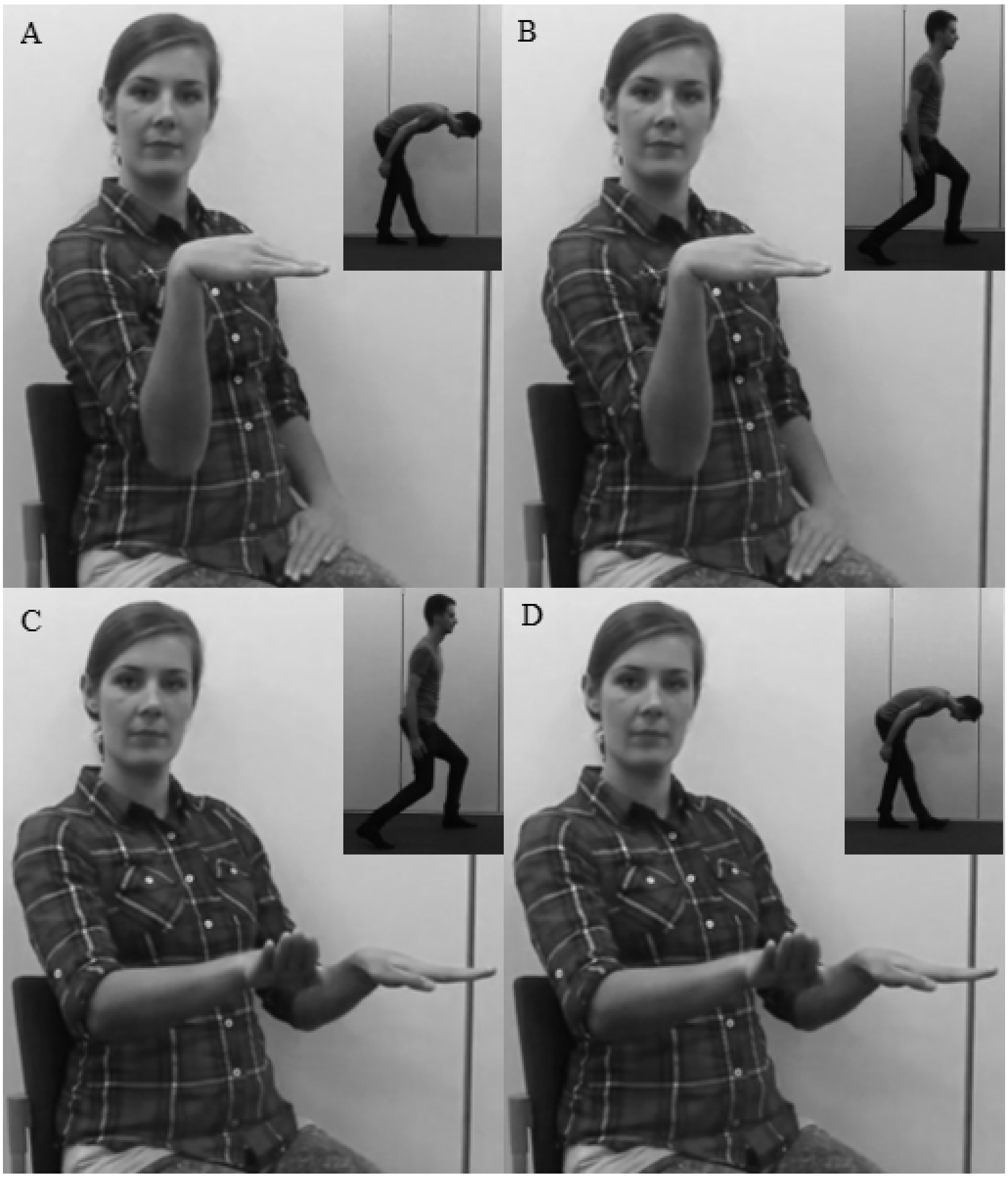


Figure 3.2: Four panels (A, B, C, and D) with cropped stills of videos in which a female actor gestures iconically to represent the actions of pair 1, as performed by a male actor in the upper right corners of the panels. Panels A shows a bowing gesture with a bowing movement (match), Panel B shows a bowing gesture with a skating movement (mismatch), Panel C shows a skating gesture with a skating movement (match), and Panel D shows a skating gesture with a bowing movement (mismatch). Gesture videos are “00F_bowing.mp4” (Panel A and B) and “00F_skating.mp4” (Panel C and D). Action videos are “06M_bowing” (Panel A and D) and “06M_skating” (Panel B and C).

gesture of the female actor (left video) and the manner in which an actor moved (right video) on a seven-point scale, where 1 indicated a very bad match, 4 indicated neither a good nor a bad match, and 7 indicated a very good match. Participants were randomly assigned to one of the 26 batches and trials were randomly displayed for each participant. After they had seen all the trials, they were asked if all the videos ran smoothly, and if not, what type of problems had occurred.

Data Analysis

Using the *irr* package in *R* software for statistical analysis (R Development Core Team, 2011), we computed Kendall's W (also known as Kendall's coefficient of concordance) to assess agreement between participants who rated the same video batch. Kendall's W is a non-parametric test statistic that takes into account the number of raters and the fact that the videos were rated on an ordinal scale. Its coefficient ranges from 0 (no agreement) to 1 (complete agreement).

We used non-parametric tests to analyze the ratings for matches and mismatches between iconic gestures and actions, because these ratings were not normally distributed. The *R* code containing the basic code for all analyses reported in this paper is uploaded as supplementary material.

The Hungarian Algorithm

We split the data based on the gender of the actors, because our aim is to identify the best possible match between iconic gestures and action pairs carried out by male actors and by female actors. The matrix containing average ratings for female actors was subjected to the Hungarian algorithm (Kuhn & Yaw, 1955; Kuhn, 1956) to find the most profitable (here best overall *match* between gestures and actions) assignment of 13 female actors to 13 action pairs (each actor can be assigned to only one action pair). In order to achieve a one-to-one assignment the matrix has to have the same number of rows and columns. The same procedure was carried out for the matrix containing average ratings for 13 male actors.

The Hungarian method (Kuhn & Yaw, 1955; Kuhn, 1956) finds an optimal assignment for a given $n \cdot n$ matrix in the following way. Suppose we have n action pairs to which we want to assign n actors on a one-to-one basis. The average ratings are the profit of assigning each actor to each action pair. We wish to find an optimal assignment which maximizes the total profit.

Let $P_{i,j}$ be the profit of assigning an ith actor to the jth action pair. We define the profit matrix to be the $n \cdot n$ matrix:

$$P = \begin{bmatrix} P_{1,1} & P_{1,2} & \cdots & P_{1,n} \\ P_{2,1} & P_{2,2} & \cdots & P_{2,n} \\ \vdots & \vdots & & \vdots \\ P_{n,1} & P_{n,2} & \cdots & P_{n,n} \end{bmatrix}. \tag{3.1}$$

An *assignment* is a set of n entry positions in the matrix, none of which lie in the same column or row. The sum of the n entries of an assignment is its profit. An assignment with the highest profit is called an optimal assignment. We implemented this algorithm in Python using the *Munkres* package. Our Python scripts are available from the Warwick Research Archive Portal at `http://wrap.warwick.ac.uk/78493`.

3.3.2 Results and Discussion

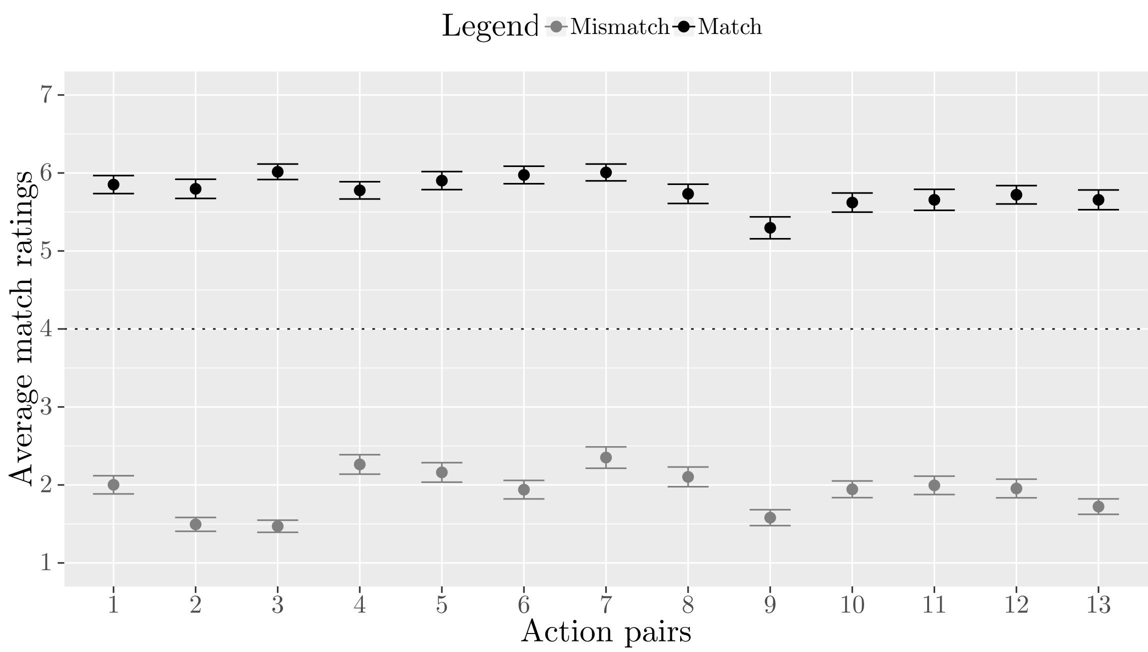


Figure 3.3: Average ratings for the degree of match between matching and mismatching iconic gestures and actions, organized by action pair. Error bars represent 95% confidence intervals of the means. Rating scores are averaged across all actors and represent the degree of match between iconic gestures and actions on a scale of 1 (“very bad match”) to 7 (“very good match”). The dotted line indicates the neutral score of 4 on the seven-point scale.

Inter-Rater Reliability

Kendall's W averaged over all 26 video batches was .72 ($SD = 0.07$) and ranged between .54 and .81. This coefficient was statistically significant for all batches ($p < .001$), indicating that participants were applying the same standards when rating the stimuli.

General Findings

Figure 3.3 displays the average ratings for the degree of match between iconic gestures and actions. Black dots represent average ratings for matches between iconic gestures and grey dots represent average ratings for mismatches between iconic gestures and actions. The 95% confidence intervals for both match and mismatch ratings are generally very narrow, indicating strong agreement among the participants.

We asked whether ratings differed between match and mismatch combinations of iconic gestures and actions. Ratings for matches and mismatches between iconic gestures and actions were averaged across all action pairs for each participant. A Wilcoxon rank sum test demonstrated that the median of average match ratings ($Mdn = 5.92$) was significantly higher than the median of average mismatch ratings ($Mdn = 1.77$), $W = 316.5, p < .001$, 95% CI of the difference [-4.12, -3.88].

Furthermore, we compared the averaged ratings for matches and mismatches across action pairs against the neutral score on our seven-point scale. A Wilcoxon signed rank test indicated that the median of average match ratings was significantly higher than a neutral score of 4, $W = 42638, p < .001$, 95% CI of the median [5.77, 5.92]. In contrast, the median of average mismatch ratings was significantly lower than a neutral score of 4, $W = 137, p < .001$, 95% CI of the median [1.75, 1.92]. Thus, matching iconic gestures and actions were rated as good matches and mismatching iconic gestures and actions were rated as bad matches.

The 95% confidence intervals of the means in Figure 3.3 clearly demonstrate that there is some variability between action pairs. When we compared the median of the averaged match and mismatch ratings for every action pair against a neutral score of 4, Wilcoxon signed rank tests revealed that matches and mismatches for all action pairs differed significantly from the neutral score ($p < .001$ for all comparisons).

Assigning Actors to Action Pairs

The Hungarian Algorithm optimally assigned 13 female actors to 13 action pairs, and did the same for 13 male actors. The Algorithm used "profit" matrices for actors and action pairs, created in the following way (one matrix for female actors, and another one for male actors). For each action performed by each actor, 10–14 participants rated the match between each action and a matching gesture. The ratings were averaged across participants, and then the two average ratings for actions that comprise an action pair were averaged again to create a "profit" for the action pair and actor.

For females, the algorithm selected the female actor with the highest match rating for an action pair eight times, the female with the second highest match rating for an action pair four times, and the female with the fourth highest match rating for an action pair one time. As the 13 females were assigned to 13 action pairs, the highest possible profit that could have been achieved was 91 (13 x 7). The algorithm assigned female actors to action pairs with a total profit of 80.63 (88.6% of 91), with the lowest average match rating for an assigned actor being 5.56 out of 7 (see Figure A.1 in Appendix).

For males, the algorithm selected the male actor with the highest match rating for an action pair six times, the male with the second highest match rating for an action pair two times, the male with the third highest match rating two times, the male with the fourth highest match rating two times, and the male with the fifth highest match rating one time. The algorithm assigned male actors to action pairs with a total profit of 81.02 (89.0% of 91), with the lowest average match rating for an assigned actor being 5.64 out of 7 (see Figure B.1 in Appendix).

Experiment 1 provided norming scores for all the videos in the GRACE videos database. With these ratings we evaluated the match and mismatch between iconic gestures and actions within action pairs. Moreover, the Hungarian algorithm over these ratings optimally assigned male actors and female actors to action pairs, to maximize the overall degree of match between gestures and action pairs. These assignments will be used in subsequent experiments.

3.4 Experiment 2

The second experiment tested whether the male actors and female actors who were assigned to an action pair based on Experiment 1 perform the same actions in similar manners and the two different actions in distinct manners. Participants rated the similarity between two action videos on a seven-point scale. These videos showed either the same actor performing two different actions, or two different actors (male vs. female) performing the same action.

We predicted that two actors performing the same action would be rated more similar than the same actor performing two different actions. Additionally, we predicted that two actors performing the same action would be rated more similar than the neutral score on a seven-point scale and the same actor performing a different action would be rated less similar than the neutral score.

3.4.1 Method

Participants

We recruited 42 individuals (19 males, 22 females, and 1 would rather not say) from the university's online participant pool. Two participants were excluded from further analyses because they indicated that the videos did not display, or run smoothly. The final participant sample included 40 individuals (20 females, 19 males, and 1 would rather not say) between 18–57 years old ($M = 24.30, SD = 8.25$). The majority of participants reported English as their native language (67.5%), followed by other Indo-European languages (22.5%), and Asian languages (10.0%). Participants automatically entered a lottery for an Amazon voucher upon completing the task.

Materials

We used videos of male actors and female actors, who were assigned to the action pairs based on Experiment 1. Trials included either two videos of the same actor (male or female) performing the two different actions in a pair, or two videos of two different actors performing the same actions in a pair (action a or action b). Thus, for each action pair we created four trials, resulting in a total of 52 trials (13 action pairs x 2 actor gender x 2 same or different action).

Counterbalancing

The left–right position of the action videos on each trial was counterbalanced across participants using two different versions of the experiment.

Procedure

The procedure of this online experiment was similar to Experiment 1. The instruction page showed two videos of the same action performed by a male actor and a female actor (who were not included in the database) as a "very similar" example. The instructions stated that participants should not proceed if they were unable to view the videos properly.

During the main task, participants saw two videos side-by-side and rated the similarity between two movements on a seven-point scale, where 1 indicated very dissimilar, 4 indicated neither similar nor dissimilar, and 7 indicated very similar. Both videos started playing on loop automatically when a trial commenced. Participants were randomly assigned to an experiment version and trials were displayed in a random order for each participant. After they had seen all the trials, they were asked if all the videos ran smoothly, and if not, what type of problems had occurred.

Data Analysis

The data were analyzed in the same way as in Experiment 1.

3.4.2 Results and Discussion

Inter-Rater Reliability

A statistically significant Kendall's W of .77 ($p < .001$) was computed for the similarity ratings, indicating that participants reached agreement when rating the stimuli.

General Findings

Figure 3.4 displays the average similarity ratings for the same and different actions within each action pair, carried out by the male actors and female actors who were assigned to these action pairs based on Experiment 1. The 95% confidence intervals of the means for both the same and different actions are generally very narrow, indicating that participants reached agreement.

We asked whether ratings differ between different actors performing the same action and the same actors performing a different action. Ratings for the same actors performing two different actions and two different actors performing the same actions were averaged across action pairs for each participant. A Wilcoxon rank sum test demonstrated that the median of average ratings was significantly higher for two different actors performing the same action ($Mdn = 6.62$) than for the same actors performing a different action ($Mdn = 1.48$), $W = 1.5, p < .001$, 95% CI of the difference [-5.19, -4.73].

We also predicted that two different actors performing the same action would be rated more similar than a neutral score of 4 and that the same actors performing

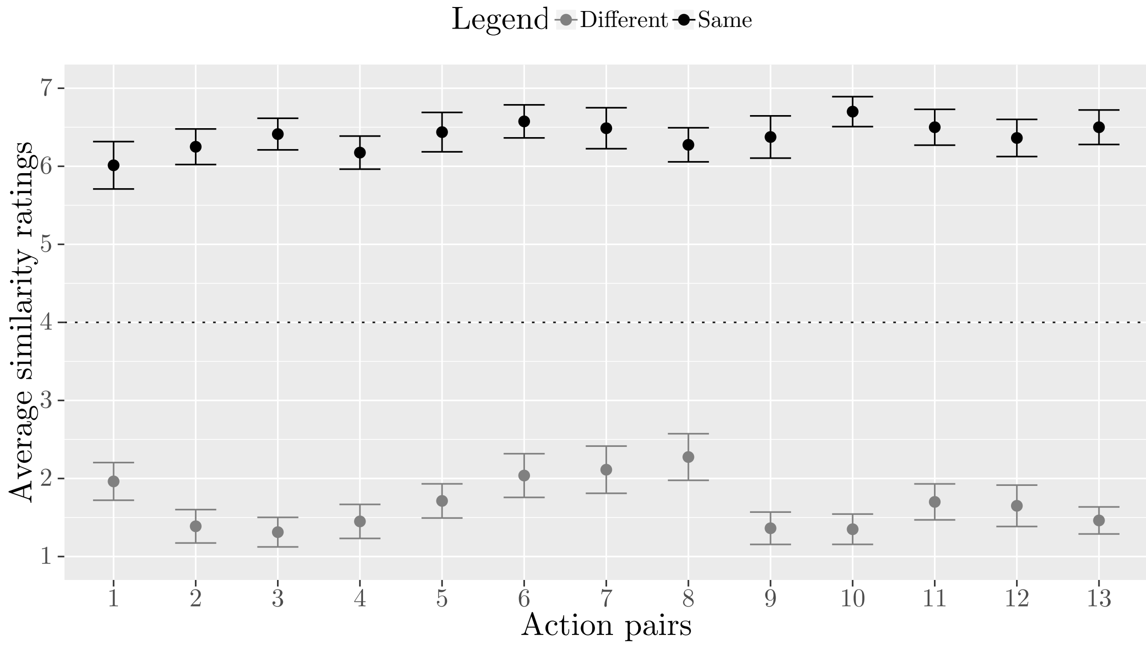


Figure 3.4: Average similarity ratings for actions within each action pair. Error bars represent 95% confidence intervals of the means. For each participant, ratings were averaged across the male actor and the female actor who were assigned to an action pair based on Experiment 1, separately for the same and different actions within each action pair. Rating scores represent the similarity between two actions, on a scale of 1 ("very dissimilar") to 7 ("very similar"). The dotted line indicates the neutral score of 4 on the seven-point scale.

a different action would be rated less similar than a neutral score of 4. Wilcoxon signed rank tests confirmed these predictions (different actors performing the same action $W = 817, p < .001$, 95% CI of the median [6.38, 6.65]; the same actors performing a different action, $W = 820, p < .001$, 95% CI of the median [1.40, 1.77]).

The 95% confidence intervals of the means in Figure 3.4 evidently show that there appears to be some variability between action pairs. When we compared the median of averaged ratings for every action pair (for the same actor performing two different actions and two different actors performing the same actions) against a neutral score of 4, Wilcoxon signed rank tests revealed that ratings for all action pairs differed significantly from the neutral score ($p < .001$ for all comparisons). Overall, Experiment 2 thus shows that male actors and female actors, who were assigned to an action pair based on Experiment 1, perform the same actions in similar manners and different actions in distinct manners.

3.5 Experiment 3

The third experiment tested how distinct the 26 actions are from every other action in the set. We used a subset of the video database, which included videos of the 26 actions carried out by the male or female actors who were assigned to an action pair based on Experiment 1. Participants rated the similarity between every combination of two action videos on a seven-point scale.

3.5.1 Method

Participants

We recruited 225 individuals (88 males, 137 females) through the university's online participant pool. Three participants were excluded from further analyses because they indicated that the videos did not display, or run smoothly. The final sample included 222 individuals (87 males, 135 females) between 18–73 years old ($M = 24.04, SD = 8.69$). The majority of participants reported English as their native language (55.9%), followed by other Indo-European languages (22.5%), and Asian languages (21.6%). Participants automatically entered a lottery for an Amazon voucher upon completing the task.

Materials

We used a set of 26 videos showing the 13 action pairs. For each action pair, we randomly determined whether each action was performed by the male or female actor that was assigned to that pair based on Experiment 1. If the male actor was selected for one action of the action pair, then the female actor was automatically selected for the other action of the action pair, and vice versa. Thus, 13 videos showed a male actor and 13 videos showed a female actor.

All possible combinations of two *different* action videos (26 x 25) were then divided over 26 video batches to keep the length of the experiment reasonable. We made sure that every action video appeared in each batch. Across batches each action video thus appeared with every other action video.

Procedure

The same procedure as Experiment 2 was used. Participants were presented with two action videos side-by-side, and rated the similarity between the actions on a seven-point scale, where 1 indicated very dissimilar, 4 indicated neither similar nor dissimilar, and 7 indicated very similar. Participants were randomly assigned to a video batch and trials were randomly displayed for each participant. After they had seen all the trials, they were asked if all the videos ran smoothly, and if not, what type of problems had occurred.

Participants were allowed to rate multiple video batches, because each batch presented participants with new combinations of action videos. We recorded 260 responses from 222 individuals. Every combination of two action videos was rated by on average 20 participants (range=19 to 22).

Data Analysis

Inter-rater reliability was calculated in the same way as in Experiment 1-2. A similarity matrix was created by averaging the ratings over every combination of two actions.

3.5.2 Results and Discussion

Table 3.2: Similarity Rating Matrix with Averages (above the Diagonal) and Standard Deviations (below the Diagonal) for Every Combination of Two Action Videos

	1a	1b	2a	2b	3a	3b	4a	4b	5a	5b	6a	6b	7a	7b	8a	8b	9a	9b	10a	10b	11a	11b	12a	12b	13a	13b
1a	■	2.38	4.23	1.43	1.57	2.00	2.68	1.48	1.10	1.95	1.38	2.23	1.95	1.86	3.67	1.74	2.23	1.89	1.43	1.62	1.48	2.19	1.26	3.95	2.19	2.24
1b	1.28	■	2.29	1.95	2.09	3.33	4.86	1.81	2.16	3.24	2.59	2.10	3.90	3.00	4.89	3.05	2.11	3.18	2.52	1.57	2.19	5.53	1.76	3.81	4.10	1.95
2a	1.57	1.35	■	1.24	1.71	1.38	1.95	1.38	1.48	2.33	1.57	2.58	1.62	1.58	2.38	1.52	4.11	1.90	1.43	1.48	1.62	1.57	2.36	2.43	1.84	2.90
2b	0.68	1.07	0.44	■	1.38	5.14	3.33	1.62	1.38	2.05	4.05	4.00	2.58	3.43	1.81	4.14	1.24	3.67	2.86	1.57	4.84	2.76	1.29	2.32	3.62	1.68
3a	0.87	1.27	1.27	0.67	■	1.19	1.43	3.33	4.27	2.84	1.16	2.14	2.14	1.43	2.00	1.57	3.38	2.16	1.90	5.42	1.67	1.29	3.19	1.62	1.67	2.29
3b	1.38	1.91	0.74	1.78	0.40	■	4.43	1.53	1.42	2.00	4.52	3.48	2.90	4.05	2.05	6.10	1.24	4.81	3.32	1.57	4.57	3.48	1.38	2.63	4.05	1.95
4a	1.80	1.74	1.09	1.85	0.68	1.33	■	2.05	1.81	2.57	2.48	3.41	3.33	4.41	5.43	4.00	1.19	3.90	3.10	1.67	3.16	4.95	1.68	3.57	3.95	2.67
4b	1.03	1.21	0.80	1.24	1.80	1.02	1.40	■	3.32	2.68	1.57	2.29	4.14	2.48	1.58	1.48	1.67	1.52	1.95	3.91	1.67	2.14	5.81	1.37	2.00	2.05
5a	0.30	1.21	0.75	1.12	1.32	0.84	1.21	1.83	■	2.82	1.57	1.62	2.19	1.43	1.79	1.38	3.76	1.29	1.81	4.05	1.43	1.71	2.14	1.24	1.76	1.62
5b	1.20	1.45	1.46	1.18	1.64	1.23	1.57	1.67	1.53	■	1.81	4.24	3.71	2.43	2.67	2.33	1.43	2.67	3.90	2.32	3.10	1.86	1.52	2.19	1.71	3.26
6a	0.74	1.50	1.03	1.40	0.50	1.50	1.44	0.75	1.40	1.03	■	2.24	1.90	3.76	1.73	3.84	1.38	2.58	1.90	1.42	2.76	3.27	1.67	1.81	4.33	1.62
6b	1.34	1.45	1.46	1.82	1.46	1.60	1.79	1.42	0.97	1.73	1.34	■	4.29	2.47	2.67	3.05	1.52	2.81	3.33	2.14	5.29	2.89	1.33	2.95	1.90	2.19
7a	1.32	1.84	0.86	1.54	1.15	1.61	1.32	1.28	1.03	1.78	1.22	1.62	■	3.10	2.71	2.19	1.33	3.71	3.91	2.05	4.16	3.29	2.42	3.23	2.73	3.16
7b	1.15	1.41	1.17	1.29	0.75	1.78	1.76	1.86	0.93	1.50	1.87	1.74	1.55	■	3.29	4.33	1.48	4.95	3.58	1.48	2.68	3.29	2.36	2.52	6.63	1.86
8a	1.98	1.79	1.24	0.98	1.20	1.28	1.54	0.69	1.40	1.59	0.98	1.71	1.62	1.79	■	2.62	1.33	3.05	2.38	1.91	1.90	3.84	1.90	3.43	2.90	2.71
8b	1.37	1.53	0.75	1.71	1.21	1.22	1.63	0.81	0.67	1.20	1.86	1.68	1.50	1.49	1.53	■	1.68	4.53	3.62	1.48	3.71	3.77	1.52	2.33	4.19	2.19
9a	1.69	1.33	1.94	0.54	1.80	0.70	0.68	1.28	1.97	0.87	0.80	0.90	0.73	0.81	0.66	0.89	■	1.59	1.71	2.10	1.76	1.32	2.43	1.76	1.62	3.43
9b	1.45	1.79	1.04	1.46	1.46	1.72	1.37	0.75	0.72	1.53	1.71	1.47	1.45	1.69	1.72	1.74	0.91	■	5.38	1.76	2.62	3.23	1.71	2.86	5.00	2.81
10a	0.68	1.40	1.16	1.61	1.04	1.38	1.41	1.40	1.17	1.81	1.22	1.91	1.93	1.68	1.16	1.72	0.90	0.86	■	1.38	4.86	2.10	1.90	1.95	3.42	4.76
10b	1.16	1.21	1.03	1.47	1.43	1.36	0.97	1.66	1.89	1.63	0.69	1.35	1.28	1.03	1.27	0.93	1.14	1.49	0.59	■	1.48	2.00	2.71	1.62	1.29	1.89
11a	0.98	1.29	1.28	1.50	0.97	1.12	1.50	1.15	0.81	1.73	1.45	1.15	2.11	1.52	1.04	1.82	1.41	1.36	1.64	0.81	■	2.43	1.53	2.81	2.50	2.33
11b	1.33	1.68	1.03	1.64	0.56	1.54	1.53	1.39	1.19	0.85	1.75	1.49	1.65	1.62	1.84	1.74	0.95	1.41	1.34	1.45	1.63	■	1.38	3.90	2.95	1.86
12a	0.65	1.22	1.73	0.64	1.54	0.59	1.06	1.17	1.31	0.93	1.06	0.58	1.46	1.68	1.14	1.03	1.29	1.10	1.34	1.49	1.02	0.59	■	2.52	2.36	2.19
12b	1.63	1.57	1.50	1.04	0.74	1.61	1.47	0.68	0.54	1.40	1.17	1.43	1.80	1.17	1.89	1.53	0.89	1.71	1.31	0.97	1.63	1.64	1.17	■	2.47	2.77
13a	1.08	1.61	1.21	1.66	1.32	1.60	1.86	1.10	1.30	0.96	1.53	1.14	1.42	0.60	1.76	1.94	1.02	1.64	1.71	0.78	1.26	1.60	1.68	1.02	■	1.52
13b	1.37	1.20	1.73	0.95	1.35	1.05	1.32	1.43	1.32	1.79	1.02	1.21	1.80	0.91	1.27	1.40	1.63	1.54	1.58	1.20	1.39	1.20	1.44	1.34	0.75	■

Note. Columns and rows are labeled with numbers of the compared actions. N varies between 19 to 22 per cell. Average ratings are given above the diagonal line of black squares in grey cells. Ratings represent the similarity between two actions, on a scale of 1 ("very dissimilar") to 7 ("very similar"). Standard deviations are given below the diagonal line of black squares in white cells. See Table 3.1 for a more detailed description of what the action numbers (i.e., "1a", "1b", etc.) refer to.

Inter-Rater Reliability

Kendall's W averaged over all 25 video batches was .52 ($SD = 0.12$) and ranged between .27 and .68. This coefficient was statistically significant for all batches ($p < .001$), indicating that participants were applying the same standards when rating the stimuli.

General Findings

Table 3.2 shows similarity ratings for every combination of two actions. The average score was 2.56 ($SD = 1.71$) and ranged between 1.10 ($SD = 0.30$) for the combination of action 5a and 1a and 6.63 ($SD = 0.60$) for action 7b and 13a.

The distinctiveness of each action can be assessed by averaging similarity ratings between a given action and the other 25 actions: the smaller this average is, the more distinct the action is. According to this metric, action 9a ($M = 1.94, SD = 1.40$), 5a ($M = 2.02, SD = 1.48$), and 2a ($M = 2.03, SD = 1.43$) appear to be most distinct.

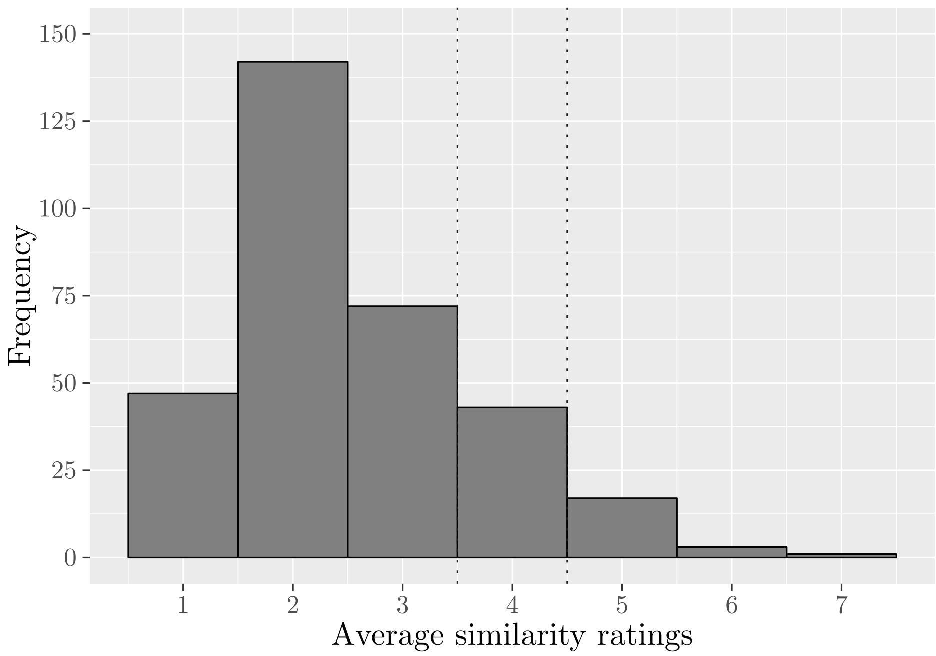


Figure 3.5: Frequency of average similarity ratings for all combinations of (different) actions in the database. $N = 325$ combinations of two different actions. Ratings represent the similarity between two actions, on a scale of 1 ("very dissimilar") to 7 ("very similar"). The dotted lines mark the neutral score of 4 on the seven-point scale.

Figure 3.5 shows that most combinations of actions (80.6%) were rated–on average–on the left side of the seven-point scale (i.e. the area left side of the first dotted line), indicating that most actions are distinct from each other. The area between the two dotted lines covers the combinations of actions that participants rated neutrally (12.9%). Very few combinations of actions (6.5%) were rated–on average–on the right end of the seven-point scale (i.e. the area on the right side of the second dotted line), indicating that only some of the actions are similar to each other.

3.6 Experiment 4

The fourth experiment assessed how accurately and concisely adult native English speakers can describe the actions in our database. This can also be used as the proxy measure for how unusual adult native English speakers find each action. If our participants find the actions unusual, then they should not converge on single-word or multi-word labels for the actions.

3.6.1 Method

Participants

We recruited 28 native English speakers (10 males, 18 females) from the university's online participant pool. One participant was excluded from further analyses because the videos did not display, or run smoothly. Three participants were excluded because they reported their first language to be something other than English. The final participant sample included 24 individuals (8 males, 16 males) between 18–48 years old ($M = 22.92, SD = 6.43$). Participants automatically entered a lottery for an Amazon voucher upon completing the task.

Materials

Experiment 4 used the same videos as Experiment 3.

Procedure

The experiment was set up in a web-based environment. Participants signed a digital consent form and were asked for demographic information. Prior to the main task, participants were shown a video of the model moving across the length of a scene. The instructions stated that every following video would also show an actor moving across the length of a scene, and that they had to describe the actor's manner of movement as concise and accurate as possible. Participants were instructed to type an "X" to skip a trial in case they could not come up with a description for the movement. Participants were also asked not to proceed if they were unable to view the video on the instruction page properly.

During the main task, a video started playing on loop automatically in the center of the screen on each trial. Participants were required to answer the question "Please describe the actor's manner of movement as *concise* and *accurate* as possible:" using a text box below the video. Participants also rated the difficulty of coming up with a description on a seven-point scale, where 1 indicated very difficult, 4 indicated neither difficult nor easy, and 7 indicated very easy. Trials were randomly displayed for each participant, until participants had seen all actions. After they had completed all trials, they were asked if all the videos ran smoothly, and if not, what type of problems had occurred.

Data Analysis

Verbatim responses were spell-checked and converted to lowercase letters. The length of the descriptions was measured by counting the number of words separated by a blank space. Any punctuation (e.g., hyphens) did not count towards the number of words in a description.

We then annotated the content words in the descriptions using a Cambridge English dictionary. Nouns, main verbs, adjectives and adverbs are content words, which usually refer to some object, action, or characteristic of an event. Verbs, adjectives, and nouns (i.e., *rotate*, *rotating*, and *rotation*) that have the same root were coded as the same responses using the root of the word (i.e. *rotate*). Annotations could contain the same root more than once, but only unique roots counted towards the total number of content words in a description. For instance, one participant described action 11a with "*jump* forward and alternate your legs with each *jump* like a scissor movement", using the word "jump" first as verb and then as a noun. These two words have the same root and therefore only added a count of one to the total count of content words per description. Auxiliary verbs, pronouns, articles, and prepositions are grammatical words and were therefore not coded. Annotations were checked by an independent researcher.

We used two key statistics to evaluate the conciseness of the descriptions: the average number of unique roots per description and the number of descriptions that contained a single root. We computed the percentage of participants that mentioned the same root for each action to measure agreement among participants. Subsequently, these roots were ranked based on how many of the participants used them in their description and the three most used roots were reported for each action. Difficulty ratings were averaged over each action.

3.6.2 Results and Discussion

Table 3.3 shows that participants provided quite lengthy descriptions for the actions (mean number of words per description: $M = 6.99, SD = 5.71$), ranging between 4.50 words for action 9b and 9.80 words for action 8b. On average, 4.68 ($SD = 3.02$) roots were annotated for the descriptions, ranging between 3.50 for action 9b and 10b and 6.15 for action 8b.

Participants generally approached the task by describing the actions using main verbs and modified these verbs using adjectives, adverbs, directional phrases, and nouns that specified the part of the body that was most involved in the movement. For instance, one participant described action 10b as "turn sideways and simply jump sideways whilst keeping your feet together" and another participant described action 4a as "walking forward crouching slightly with knees bent".

For the next analyses, we excluded 30 responses that stated an "X", because this indicated that participants could not come up with a description. Participants who attempted to describe the actions used only a single content word in 7.4% of the cases. In most responses, participants thus used more than one content word to describe the actions.

Participants rated the difficulty of coming up with a description on average with 4.22 ($SD = 1.74$), suggesting that the task was neither difficult nor easy. Participants found action 6b and 10a ($M = 2.67$) most difficult to describe and action 10b ($M = 6.08$) easiest to describe.

Finally, we correlated the length of the descriptions (i.e. the number of words) and the number of roots per description with the difficulty ratings for the actions. Participants who found it more difficult to describe the actions provided longer descriptions $r(595) = .10, p = .017$, and used more content words, $r(595) = .12, p = .003$.

Table 3.3: Descriptive Statistics of Written Descriptions for All 26 Actions in the GRACE Database

No.	Action	X (%)	No. words *M* (*SD*)	No. roots *M* (*SD*)	Root 1 (%)	Root 2 (%)	Root 3 (%)	No. single roots (%)	Difficulty *M* (*SD*)
1a.	bowing	0 (0.0)	8.04 (6.22)	5.29 (3.51)	bend (66.7)	walk (58.3)	forward (54.2)	1 (4.2)	4.58 (1.38)
1b.	skating	2 (8.3)	6.32 (4.24)	4.18 (2.30)	zigzag (25.0)	slide (25.0)	forward (25.0)	1 (4.2)	4.08 (1.79)
2a.	wobbling	1 (4.2)	8.43 (5.18)	5.48 (2.69)	body (62.5)	rotate (41.7)	upper (37.5)	1 (4.2)	3.67 (1.71)
2b.	marching	0 (0.0)	5.92 (3.60)	4.29 (2.18)	leg (66.7)	forward (41.7)	march (29.7)	3 (12.5)	5.29 (1.46)
3a.	mermaiding	0 (0.0)	7.71 (6.80)	4.96 (3.63)	side (79.2)	jump (58.3)	together (33.3)	0 (0.0)	5.04 (1.40)
3b.	overstepping	0 (0.0)	6.92 (4.60)	4.96 (2.72)	leg (54.2)	step (33.3)	forward (33.3)	1 (4.2)	4.33 (1.37)
4a.	creeping	0 (0.0)	5.29 (6.83)	3.58 (3.24)	walk (29.2)	forward (29.2)	slow (25.0)	5 (20.8)	4.50 (1.96)
4b.	crisscrossing	0 (0.0)	7.29 (5.22)	4.79 (2.64)	side (75.0)	cross (62.5)	leg (62.5)	1 (4.2)	3.92 (1.69)
5a.	turning	1 (4.2)	9.39 (8.54)	5.91 (3.99)	jump (79.2)	side (58.3)	turn (29.2)	0 (0.0)	3.96 (1.33)
5b.	hopscotching	2 (8.3)	7.27 (8.02)	4.64 (3.98)	hopscotch (50.0)	leg (33.3)	jump (29.2)	4 (16.7)	4.04 (1.78)
6a.	swinging	1 (4.2)	8.78 (5.38)	5.30 (2.67)	leg (70.8)	circle (50.0)	walk (33.3)	1 (4.2)	3.71 (1.52)
6b.	skipping	5 (20.8)	8.21 (8.49)	5.00 (4.18)	forward (45.8)	leg (37.5)	move (25.0)	2 (8.3)	2.67 (1.74)
7a.	jumping	0 (0.0)	7.04 (4.25)	4.71 (2.46)	jump (58.3)	forward (45.8)	leg (41.7)	1 (4.2)	3.96 (1.46)
7b.	crossing	2 (8.3)	8.09 (4.68)	5.41 (2.74)	cross (50.0)	leg (45.8)	walk (37.5)	0 (0.0)	3.54 (1.35)
8a.	dropping	1 (4.2)	6.61 (6.52)	4.83 (3.77)	walk (58.3)	squat (33.3)	crouch (29.2)	2 (8.3)	4.00 (1.47)
8b.	folding	4 (16.7)	9.80 (7.21)	6.15 (3.17)	leg (45.8)	walk (41.7)	step (33.3)	0 (0.0)	2.92 (1.79)
9a.	twisting	2 (8.3)	5.77 (6.64)	4.27 (2.81)	rotate (33.3)	degree (25.0)	side (20.8)	2 (8.3)	3.76 (1.69)
9b.	stomping	0 (0.0)	4.50 (3.20)	3.50 (2.02)	knee (50.0)	high (41.6)	stomp (33.3)	5 (20.8)	5.42 (1.32)
10a.	trotting	1 (4.2)	5.43 (3.64)	4.04 (1.89)	knee (62.5)	high (54.2)	forward (37.5)	2 (8.3)	2.67 (1.88)
10b.	hopping	0 (0.0)	4.79 (3.75)	3.50 (2.32)	side (87.5)	jump (58.3)	together (37.5)	1 (4.2)	6.08 (1.14)
11a.	flicking	2 (8.3)	5.73 (4.00)	4.00 (2.12)	leg (41.7)	forward (33.3)	quick (20.8)	1 (4.2)	3.50 (1.62)
11b.	dragging	1 (4.2)	7.17 (5.65)	5.17 (3.28)	forward (50.0)	leg (41.7)	step (37.5)	2 (8.3)	4.42 (1.74)
12a.	grapevining	1 (4.2)	6.48 (5.88)	4.48 (2.98)	cross (62.5)	side (58.3)	leg (50.0)	2 (8.3)	4.04 (1.23)
12b.	shuffling	0 (0.0)	5.83 (4.39)	3.71 (2.31)	walk (50.0)	forward (37.5)	lift (20.8)	3 (12.5)	5.42 (1.56)
13a.	groining	3 (12.5)	9.48 (6.30)	6.14 (3.45)	leg (54.2)	walk (29.2)	knee (29.2)	0 (0.0)	2.96 (1.68)
13b.	scurrying	0 (0.0)	5.38 (4.53)	3.92 (3.03)	tiptoe (58.3)	step (33.3)	fast (25.0)	3 (12.5)	5.36 (1.06)

Note. $N = 24$. Columns left to right describe the action number (No.), short-hand action label (Action), number (%) of participants who were not able to describe the action (X), length of a description (No. words), number of content words in a description (No. roots), three most used content words for each action (Root 1-3), number (%) of descriptions that contained a single content word (No. single roots), and the rated difficulty of coming up with a description (Difficulty, 1 = very difficult and 7 = very easy).

3.7 General Discussion

We developed the GestuRe and Action Exemplar (GRACE) video database, which is publicly available from the Warwick Research Archive Portal at `http://wrap.warwick.ac.uk/78493`. The GRACE video database contains 676 videos of 26 novel manners of human locomotion performed by 13 male actors and 13 female actors (i.e. actors moving from one location to another in an unusual manner), and videos of a female actor producing iconic gestures that represent these manners.

Our first norming study demonstrates that GRACE contains gesture and action videos that can be combined to create clear matches and mismatches between iconic gestures and manners of human locomotion. Based on the findings of this first norming study, we assigned two actors (one male and one female) to a pair of actions to maximize the match between the iconic gestures and actions. Our second norming study shows that male actors and female actors who were assigned to an action pair perform the same actions in very similar manners and the different actions in highly distinct manners. Our third norming study indicates that the majority of actions are, in fact, highly distinct from all other actions in the database. Our fourth norming study demonstrates that adult native English speakers do not converge on accurate and concise linguistic expressions for the actions in the database, indicating that these manners of human locomotion are unusual.

3.7.1 Conclusion

This database is useful for experimental psychologists working on action and gesture in areas such as language processing, vocabulary development, visual perception, categorization, and memory. By making our video database publicly available to the research community, we set out to inspire researchers to norm our videos for their own studies. We invite these researchers to share these norms with us and other researchers so that we can upload these along with the GRACE video database through the Warwick Research Archive Portal.

Chapter 4

Seeing Iconic Gestures while Encoding Events Facilitates Children's Memory of these Events

An experiment with 72 three-year-olds investigated whether encoding events while seeing iconic gestures boosts children's memory representation of these events. The events, shown in videos of actors moving in a novel manner, were presented with either iconic gestures depicting an actor performing this action, interactive gestures, or no gesture. In a recognition memory task, children in the iconic gesture condition remembered actors and their actions better than children in the control conditions. Iconic gestures were categorized based on how much of the actors was represented by the hands (feet, legs, or body). Only iconic hand-as-body gestures boosted actor memory. Thus, seeing iconic gestures while encoding events facilitates children's memory of those aspects of events that are schematically highlighted by gesture.

4.1 Introduction

Children spend a considerable proportion of their day watching what other people do. Accurate memory of *who* did *what* is crucial for their social-cognitive development because it lies at the core of social interactions (Vogelsang & Tomasello, 2016), cooperative activities (Milward, Kita, & Apperly, 2014), and learning how things "ought" to be done (Burdett et al., 2016; Schmidt, Rakoczy, & Tomasello, 2016). In order to make sense of the events they see, children must learn how to encode, process, and organize the various aspects of those events. Children can recognize people's actions in impoverished stimuli such as point-light displays at age three (e.g., Golinkoff et al., 2002), but rich action events with real people are much more complex to encode. This study focuses on children's memory of such events, more specifically, their memory representation of real-life actors and their actions.

4.1.1 The Challenge of Encoding Action Events

Action events are difficult to encode for young children (Imai et al., 2005). The challenge that they must overcome is understanding that an action event consists of both stable components such as people and objects, and transient components such as the things that people do (e.g., actions such as yawning or jumping). The transient nature of actions makes it difficult for children to remember them. In an event recognition task, children remember stable aspects of events (e.g., objects) better than actions (Imai et al., 2005). Because actions are transient and other aspects of an event are stable, it is also difficult for children to focus on an action as the sole referent of a verb. Word learning tasks in which children were taught a verb while watching an actor performing an action on an object (e.g., whipping the whisk) showed that children map a verb onto the combination of an object and action, rather than onto action alone (Imai et al., 2005; Imai, Li, et al., 2008). Verb learning studies have repeatedly demonstrated that 3-4-year-old children focus too much on stable aspects of action events (e.g., objects, instruments, actors) rather than on actions (Behrend, 1990; Forbes & Farrar, 1993; Kersten & Smith, 2002). This focus on stable aspects prevents children from generalizing action labels to new events in which the same actions are shown, but the objects, actors, or instruments have changed.

4.1.2 Encoding with the Help of Iconic Gestures

In the current study, we investigate whether seeing iconic gestures helps children to encode action events in a recognition memory task. People naturally produce iconic gestures when they speak (McNeill, 1992). Gestures are iconic when hand movements resemble a referent in shape and motion. For instance, a gesture can represent the things that people do (e.g., wiggling the index and middle fingers to represent a person walking).

Observing iconic gestures while encoding *verbally presented* information (e.g., words, explanations) can influence children's performance in a subsequent task, in which they use the encoded information (e.g., a recall task, a test of word meaning, problem solving). Three lines of evidence support this. First, iconic gestures facilitate memory for familiar and novel words. For instance, 4-5-year-old children recalled more familiar words (e.g., look, swim, brush) when they encoded these words while observing iconic gestures which were semantically related to those words than when encoding words alone (So, Sim, & Low, 2012). Additionally, 5-year-old French children who encoded common English words while observing iconic gestures representing the meaning of those words recalled more words than children who encoded the same words while observing pictures showing the word meanings (Tellier, 2008). Second, children use iconic gestures to disambiguate the meanings of novel verbs. In an experiment by Goodrich and Hudson Kam (2009), 3- and 4-year-olds were taught two novel verbs for actions performed by a puppet on unfamiliar toys (e.g., rolling down a ramp in a tube). The experimenter demonstrated two actions, one at a time. Subsequently, the experimenter taught the children one verb per action, while accompanying each verb with an iconic gesture that depicted the action (e.g., rolling down the ramp gesture; index finger tracing circles while moving downward at an angle). When the children were subsequently asked "Which toy lets the puppet go (novel verb)-ing?" (without a gesture), they picked the toy which operated in the way that corresponded to the verb. Third, iconic gesture facilitates learning of verbally explained strategies for problem solving. In a study by Ping and Goldin-Meadow (2008), 5-7-year-olds received verbal instructions on how to solve Piagetian conservation problems. For instance, in the case of liquid conservation with two differently shaped glasses that contained the same amount of water, an experimenter explained to children that one glass was tall and skinny and the other was short and wide. The explanation was either accompanied by iconic gestures indicating the height and width differences between the glasses, or no gesture. For half of the children, the glasses were not present during this instruction, and thus the gestures iconically depicted the dimensions of the glasses. The participants were then asked whether the amount of water in the two glasses was the same and if they could explain their answer. Children who saw iconic gestures during the instruction solved the quantity conservation problems more often than children who did not see gesture, even when the objects were not present during the instruction. The above studies indicate that seeing iconic gestures influences how children encode and subsequently use *verbally presented* information. However, much less is known about how seeing iconic gestures influences memory of *nonlinguistic* information, for instance, memory of events.

4.1.3 Seeing Iconic Gestures Influences How Children Recall Events

Previous research on the impact of seeing iconic gestures on children's event memory always presented children with iconic gestures at the recall stage, but never at the encoding stage. For instance, research on eyewitness testimony suggests that seeing iconic gestures at the recall stage can alter children's memory representation of an event long after they have witnessed this event. In a study by Broaders and Goldin-Meadow (2010), a musician visited 5-6-year-olds in their classroom and in the weeks after the visit, the children were asked questions about the appearance of the musician in scripted interviews. During the questioning, the interviewer conveyed misleading information in gesture (e.g., moving the hand towards the head as if putting on a hat), but not in speech (e.g., "What was the musician wearing?"). In their responses to the interviewer's questions, children often said that the musician wore a hat, which was false. This information corresponded to what was encoded in the interviewer's gestures, but not in their speech. When children narrated the event in a free recall task several weeks later, their stories included information gleaned from the interviewer's gestures during the scripted interviews. Children had thus incorporated misinformation from the interviewer's gestures in their memory representation of the event.

One might argue that children are prone to the influence of seeing iconic gestures at the recall stage, because they presume that the experimenter is signaling the correct answers to them. However, Kirk, Gurney, Edwards, and Dodimead (2015) showed that the influence of iconic gesture cannot solely be attributed to such a demand characteristic. In their study, 2-4-year-olds and 7-9-year-olds watched a video clip showing a series of events (e.g., a lady and a man roller-skating) and were required to narrate the events to the experimenter afterwards. The experimenter then questioned the children about the events (e.g., "What was the lady wearing?") under one of two conditions: accurate gesture (e.g., moving the hands towards the head as if putting on a hat) or misleading gesture (e.g., moving the left hand over the right hand as if putting on a glove). The lady in the video clip was in fact wearing a hat, but no gloves. After the interview, the children narrated the events again to the experimenter. In both gesture conditions and both age groups, children's post-interview narrations included information that was absent from their pre-interview narrations, but consistent with information the experimenter had encoded in gesture during the interview. Seeing iconic gestures at the recall stage thus changed children's memory of events. Importantly, when children retold their version of the witnessed events after the interview, they added more information gleaned from accurate gestures than from misleading gestures to their stories, which rules out the possibility that gestures represent a *demand characteristic* (i.e., children concur indiscriminately with any information conveyed by the experimenter's gestures under the presumption that the experimenter is signaling the correct answers to them).

To summarize, seeing iconic gestures influences children's memory of *verbally presented* information and *nonlinguistic* information. However, research on the influence of seeing iconic gestures on children's event memory is sparse, and in such studies, iconic gesture was always manipulated at the recall stage (Broaders & Goldin-Meadow, 2010; Kirk et al., 2015), but never at the encoding stage. From these studies, we can conclude that children's memory representation of an event is prone to the influence of (misleading) nonverbal cues at the recall stage. Yet, it remains unclear how seeing iconic gestures when children encode an event may influence their memory representation of this event. In the current study, children performed an event recognition task, in which iconic gestures were manipulated at the encoding stage. Specifically, we ask whether encoding action events with the help of iconic gestures leaves children with a stronger memory representation of these events.

4.1.4 Possible Mechanism

Observing iconic gestures can draw children's attention to certain aspects of an event and boost their memory of those aspects. Iconic gestures can encode information in an abstract, schematic manner (Kita, 2000; de Ruiter, 2000). For example, when depicting the hopping movement of a bunny going down a slope, an iconic gesture can capture this information simply by tracing the animal's trajectory (using the extended index finger to trace an arch for every hop, while generally going downward diagonally). Such a gesture focuses on the manner and path of the motion, stripping it from everything else (e.g., what the bunny looked like, any background objects and characteristics of the landscape).

The literature on gesture production suggests that the schematic nature of the gestural representation shapes the self-oriented function of gesture (Goldin-Meadow, 2015; Kita, Alibali, & Chu, 2017; Novack, Congdon, Hemani-Lopez, & Goldin-Meadow, 2014; Novack & Goldin-Meadow, 2016). More importantly for the current study, Kita et al. (2017) claim that the schematic nature of the gestural representation impacts *how the observer processes information* about complex events. Because schematic representations highlight only a certain aspect of a complex event, they may help observers to focus on certain information in the event. This idea is supported by a study which demonstrates how iconic gestures shape children's interpretations of novel verb meanings. Mumford and Kita (2014) taught

3-year-olds novel verbs that could be interpreted as manner verbs (e.g., "to pull") or change-of-state verbs (e.g., "to break"). In one of the trials, children saw a video of a hand creating a cloud shape by pushing pieces of paper together with the index finger. Children were then taught a novel verb by the experimenter; for some children, the experimenter produced a manner gesture (e.g., representing the manual action of pushing the pieces into place), and for others, the experimenter produced an end-state gesture (e.g., tracing the final cloud shape these pieces formed). Children's performance on a verb generalization task showed that they interpreted the novel verb meanings consistent with information encoded in gesture: as manner verbs when they saw manner gestures and as change-of-state verbs when they saw end-state gestures. Thus, the schematic representation in iconic gesture directed children's attention to a particular aspect of a complex event.

4.2 The Current Research

This study investigates whether seeing iconic gestures facilitates children's recognition memory of action events. The events, which included videos of real-life actors moving across a scene in a novel manner, were presented in three gesture conditions. The first condition showed iconic gestures depicting how the actors in the action events moved (i.e., their manner of locomotion). The second condition showed interactive gestures (Bavelas et al., 1992) which communicated excitement and surprise to the children, but were unrelated to the action events. The third condition did not include gestures at all. After a delay, children were asked to point out the video that they had seen before in a two-way forced choice task. Their memory of both actions and actors was tested. For action memory, children chose between the seen video and a video that included the same actor moving in a different manner, and for actor memory they chose between the seen video and an unseen video that included a different actor moving in the same manner (cf. Imai et al., 2005).

When children encode events while observing iconic gestures that represent how an actor moves, this may leave a stronger trace in children's memory because a schematic gestural representation focuses children's attention on the motion event itself as opposed to other information (e.g., details of the scene). Thus, we predict that seeing iconic gestures which highlight how an actor moves will boost children's action memory compared to seeing interactive gestures, or no gesture. As the iconic gestures also represent, to some extent, the actor who is carrying out the movement (e.g., both hands flicking upward to represent the actor's legs while marching), we predict that children's actor memory will also be boosted when they see iconic gesture, compared to seeing interactive gesture, or no gesture at all. However, since manner of motion is omnipresent in the gestures and in most cases only part of the actor's body is gesturally represented, we predict that action memory will be boosted more so than actor memory.

Furthermore, we predict that the extent to which the actor is represented in gesture could have an impact on actor memory. In our iconic gestures, the hands represent either the feet, legs, or body of the actors to depict how they moved (see Figure 4.1). While the gestures do not express any person-specific features of the feet, legs, or body (e.g., a particular actor with long legs), they should draw children's attention to particular aspects of the actor that they depict. This idea is supported by a previous finding that iconic gestures can make children focus on a particular aspect of a complex event in a verb learning task (Mumford & Kita, 2014). We inferred that when children focus on the actors' bodies, they have more opportunities to pick up person-specific information, as compared to when they focus on their feet or legs. Thus, observing a body gesture should lead to better actor memory; that is, actor memory performance increases as more of the actor is represented in gesture (e.g., hand-as-body gestures > hand-as-leg gestures > hand-as-foot gestures). If this prediction is borne out, then this is also in line with the idea that gesture schematizes information, which helps children to focus on particular aspects of events.

4.3 Experiment 5

4.3.1 Method

Design

The experiment had a mixed 3 x 2 x 3 design with gesture type as a between-subjects factor (iconic gesture vs. interactive gesture vs. no gesture) and memory type (action memory vs. actor memory) and semiotic type (feet vs. legs vs. body) as within-subjects factors. The dependent variable was children's performance in each of 12 trials of an event recognition memory task (binary: 1=correct, 0=incorrect).

Participants

The data were collected between the 23$^{\text{rd}}$ of March 2016 and the 27$^{\text{th}}$ of September 2016. Our sample size was determined a priori using G$*$Power version 3.1 (Faul, Erdfelder, Buchner, & Lang, 2009) based on pilot data that showed a small- to medium-sized effect (odds ratio= 2.30, α = 0.05, power= 0.80). We recruited

85 typically developing children from a database of families who showed interest in participating in child research and from six public and private nurseries in the West-Midlands and Warwickshire, England. A total of 13 children were excluded because they were outside the age range on the day of testing, or showed a strong bias towards the left or right side of the screen (i.e., responded exclusively to one side). The final sample included 72 children (35 girls) between the ages of 35–48 months old ($M = 41.11, SD = 3.67$). There were 24 children in each of the three gesture conditions. The gender distribution of the children was the same across conditions $\chi^2(2) = 0.11, p = .946$, as well as their average age in months, $F(2, 69) = 0.58, p = .561$. All children were exposed to the English language at home for >75% of the time (as indicated by their caregivers). Informed parental consent was obtained for all participants. Informed parental consent was obtained for all participants. In return for their participation, nurseries received a voucher for educational goods and children who were tested in the research lab received a certificate.

Materials

A set of 48 short video clips (4-14 seconds) was taken from the GRACE video database (Aussems, Kwok, & Kita, in press, 2017). The set included videos of 24 actors (12 males, 12 females), each performing two of 24 novel actions (see Table C.1 in Appendix C). Actors always moved from the left side to the right side of a scene such that the *path* of motion was the same for each action, but the *manner* of motion differed. To create a two-way forced choice task, we organized the actions in twelve pairs of distinctive actions. Actions within each pair were depicted by *one* male actor and *one* female actor in separate videos. Trials on which we tested for children's memory of actors always showed a distractor video of an actor of the opposite gender performing the same action as the actor in the target video, because males and females have naturally distinct appearances. We chose this set-up because we did not want to make the task too difficult for the children. Each action pair was performed by different pairs of male actors and female actors, whose videos were normed for how similar the same actions by different actors were and how dissimilar the different actions by the same actor were. The videos are available from Warwick Research Archive Portal (WRAP) at `http://wrap.warwick.ac.uk/78493`.

Video clips of action events were shown using slide presentation software Microsoft Office PowerPoint 2016 on a 14" touchscreen laptop. Using on-screen buttons and Visual Basic for Applications (VBA), the data were automatically saved.

Procedure

Children were tested individually in a quiet area of their nursery or in the research lab at the university. A female experimenter sat down with the children at a children's table with small chairs, always positioning herself on the left side of the child. The memory task consisted of two phases: an encoding and recognition phase. In the encoding phase, children were told that they were going to watch videos with the experimenter on a touch screen computer. A big button with a smiling star appeared on the screen and children were instructed to press the star with their index finger to start a video. Children were presented with 12 videos, which showed 12 unusual actions performed by different actors (six males, six females). Each video was shown twice in the following way. When the video played the first time the experimenter said: "Wow! Look at what he (or she) is doing!" and when the video played the second time the experimenter said: "Oh! Look, he (or she) is doing it again!". Depending on the condition, the experimenter produced iconic gestures, interactive gestures, or no gesture.

Figure 4.1 shows the three gesture conditions used in the experiment. Iconic gestures depicted the manners in which actors moved across a scene in hand shape and in motion. Interactive gestures indicated excitement and surprise, but were unrelated to the events (see Figure D.1 in Appendix D for more details). In the no gesture condition, the experimenter kept her hands in her lap when children viewed the action events.

Each iconic gesture matched one manner of motion. Iconic gestures were categorized into three semiotic types (see Figure 4.2). First, hand-as-foot gestures depicted the actors' manners of motion by representing the actors' feet with both hands (the left panel in Figure 4.2; the hand shape and the alternating circular hand movements resemble the actor's creeping feet movements). Second, hand-as-leg gestures depicted the actors' manners of motion by representing the actors' legs with both hands (the mid panel in Figure 4.2; the hand shape and the alternating lifting movements resemble the actor's trotting leg movements). Third, hand-as-body gestures depicted the actors' manners of motion by representing the body with one hand (the right panel in Figure 4.2; bending the hand at the wrist resembles the actor's body bending at the torso).

All gestures were performed for the entire duration of a video and the experimenter alternated her gaze between the child and the video (and did not look at her own gestures). Gestures were produced in the left part of the children's field of vision, in front of their left shoulder and at eye height, so that they would not have to turn their heads to look at the gestures while watching the videos. Note that the

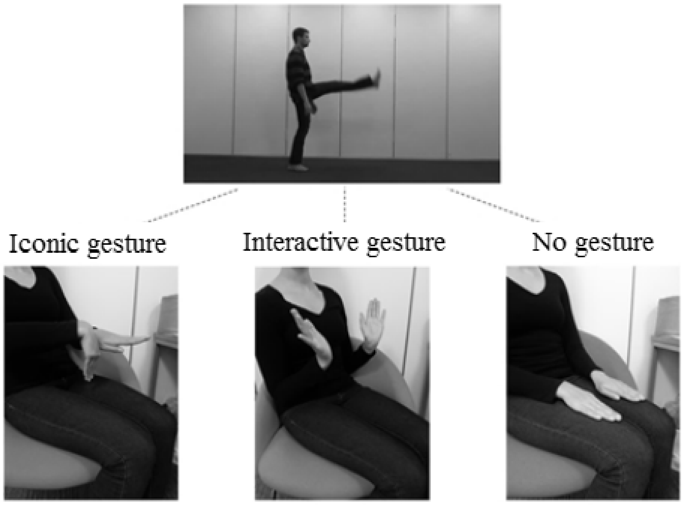


Figure 4.1: Top panel shows an actor performing a marching movement with stretched legs. The bottom panels show the type of gesture that children saw with this action event. From left to right the panels show an iconic gesture representing the actor's manner of motion, an interactive gesture unrelated to the event, or no gesture.

children were not instructed to look at the experimenter's gestures or remember the actors or actions. Children were not told about the upcoming test trials either.

After the encoding phase, children spent approximately five minutes decorating a wristband with colorful stickers. The experimenter asked children to count the stickers and name the colors during this distraction task.

The recognition phase consisted of two practice trials and 12 experimental trials. During practice trials participants saw a picture of a cat and a dog on the left and right sides of the screen. Children were asked to point (without touching the screen) at the cat and the dog to familiarize them with pointing at both sides of the screen. The experimental trials each showed two videos playing simultaneously side-by-side. Half of the test trials tested action memory and the other half actor memory. During action memory trials, six of the videos that children had seen during the encoding phase were paired up with videos of the same actor performing a different action. During actor memory trials, the other six videos from the encoding phase were paired up with videos of a different actor performing the same action (see Figure 4.3). On each trial, the experimenter asked the child "Which one did you see before?". The experimenter looked at the child when making this request and did not look at the screen. The videos played automatically on loop until the child pointed at one of them. If the child did not respond or asked whether a video

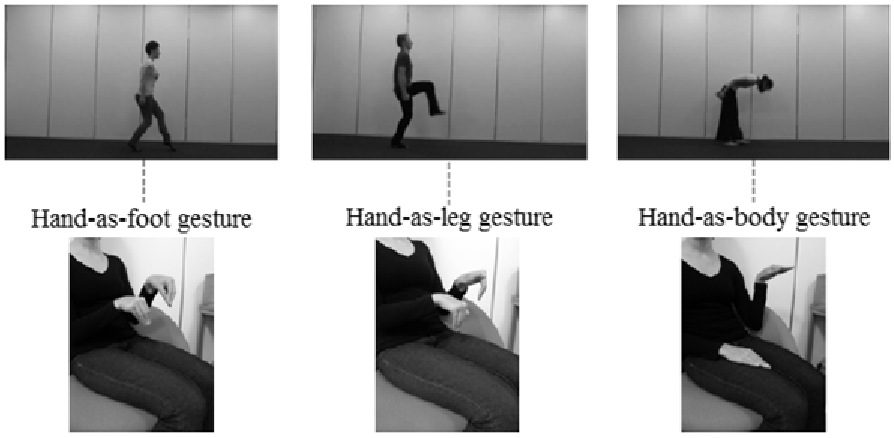


Figure 4.2: The semiotic types of iconic gestures used in the experiment. From left to right the experimenter's hands depict the manners in which actors move by representing the actors' feet (hand-as-foot gesture), legs (hand-as-leg gesture), and body (hand-as-body gesture).

was shown before, the question was repeated until a video was chosen. If the child pointed at both videos, the experimenter asked the child to pick one.

Randomization and Counterbalancing

We created 24 versions of the experiment in which every stimulus video appeared as a target and distractor on action memory trials and actor memory trials. We counterbalanced the gender of the actors in the videos, the left-right position of the videos on the screen, and we randomized the order of trials in each experiment version. Children were randomly assigned to conditions using the nursery registers, which were either ordered alphabetically by the children's surname or by their date of birth. One nursery did not provide a register and the experimenter used the order in which the consent forms were received. The conditions were rotated across participants within each testing site, and the experimenter continued the order of conditions when a participant was tested in the research lab in between nursery visits.

Data Analysis

Our binary dependent variable (correct vs. incorrect responses in the recognition memory task) was analyzed using mixed-effects logistic regression analyses. We used a maximal random-effects structure in all models (cf. Barr, Levy, Scheepers, & Tily, 2013), by including random slope variation, random intercept variation, and

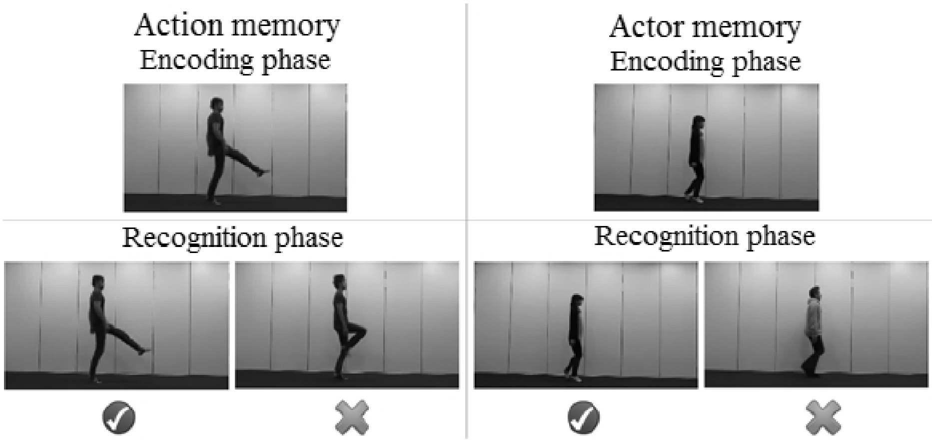


Figure 4.3: The procedure of Experiment 5. The overview shows stills of videos shown in the Encoding phase and Recognition phase of the memory task, which show examples of action memory trials (left panels) and actor memory trials (right panels). Check marks indicate correct answers and crosses indicate incorrect answers in the Recognition phase. After the Encoding phase, children had a five-minute break in which they completed a filler task before moving to the Recognition phase.

the covariance between the two for effect variation across participants and items. All analyses were carried out in R software for statistical analyses (R Development Core Team, 2011), using the *lme4* package (D. Bates, Mächeler, Bolker, & Walker, 2013). We compared each model with updated versions of the model that systematically excluded each main effect and interaction term using likelihood ratio tests (chi-squared). Using likelihood ratio tests (χ^2), we compared each model with updated versions of the model that systematically excluded the main effect and interaction terms of interest. Both marginal and conditional R^2 values were calculated using the *piecewiseSEM* package (Nakagawa & Schielzeth, 2013). Marginal R^2 reflects variance explained by fixed factors, and conditional R^2 reflects variance explained by both fixed and random factors. The raw data and the *R* script for the analyses are available from the Open Science Framework at `https://osf.io/tqk34/`.

4.3.2 Results and Discussion

Gesture Type and Memory Type

Figure 4.4 shows children's recognition memory performance organized by gesture type and memory type. Children's recognition memory performance (binary: 1=correct, 0=incorrect) was entered into a mixed-effects logistic regression analysis with gesture type as a between-subjects factor and memory type as within-subjects factor. The main effect of gesture type was significant, $\chi^2(2) = 13.18, p = .001$, but not the main effect of memory type, $\chi^2(1) = 0.25, p = .617$, or the interaction, $\chi^2(2) = 2.89, p = .236$. The model explained approximately 10% of the variance in children's recognition memory performance (marginal $R^2 = .03$, conditional $R^2 = .10$).

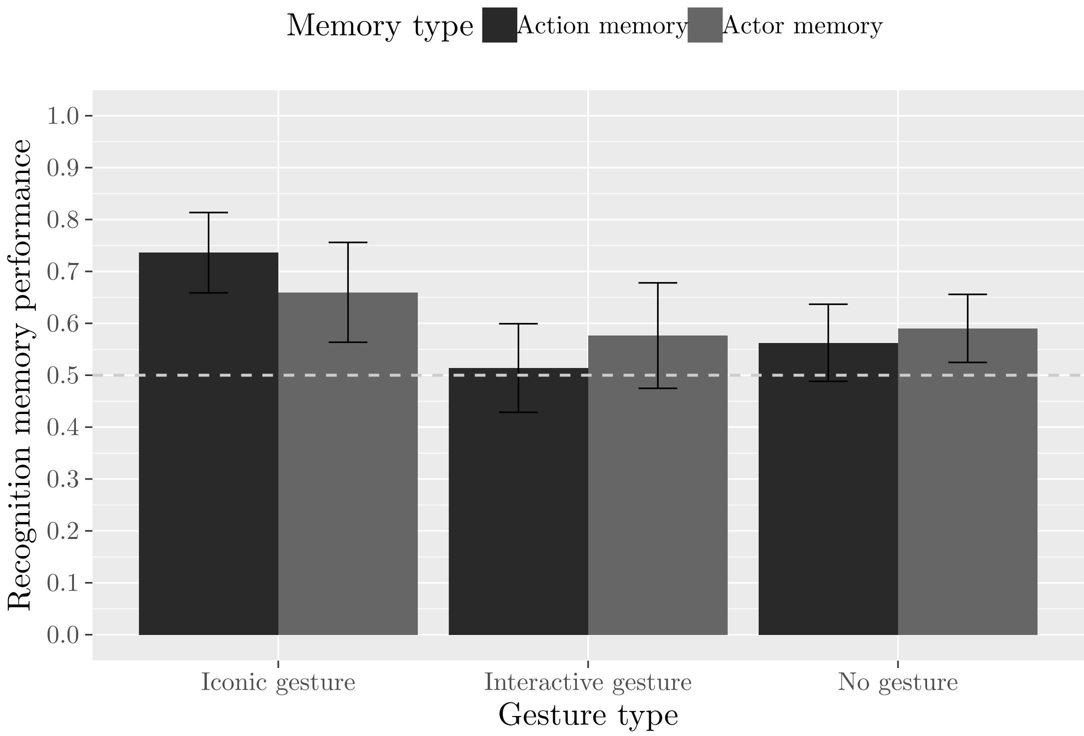


Figure 4.4: Children's recognition memory performance (y-axis shows proportion of correct responses) for actions (dark grey) and actors (light grey), organized by gesture type (x-axis). Error bars represent 95% confidence intervals of the means. Dotted line represents chance level.

To further explore the nature of the main effect of gesture type, we examined the beta estimates between the three gesture conditions. Children's recognition memory performance was significantly better in the iconic gesture condition than in the interactive gesture condition ($\beta = -0.99, SE = 0.27, p < .001$) and the no gesture condition ($\beta = -0.79, SE = 0.27, p = .003$). Then, we compared recognition memory performance between the two control conditions by releveling gesture type with the interactive condition as the reference point. The performance did not differ significantly between the interactive gesture condition and the no gesture condition

($\beta = 0.20, SE = 0.25, p = .426$).

To investigate whether children showed a better recognition memory performance than chance for each of these gesture types, we compared the proportion of correct trials in each condition against a test value of 0.5 (chance level of 50%). The proportion of correct trials was analyzed with one-sample t-tests. In the iconic gesture condition, children's memory performance was significantly above chance level ($M = 0.70, SD = 0.21, t(23) = 5.71, p < .001$, 95% CI $[0.63, 0.77]$), as well as in the no gesture condition ($M = 0.58, SD = 0.16, t(23) = 3.05, p = .006$, 95% CI $[0.52, 0.63]$), but not in the interactive gesture condition ($M = 0.55, SD = 0.22, t(23) = 1.26, p = .220$, 95% CI $[0.47, 0.62]$).

Thus far, our findings demonstrate that children who saw action events accompanied by iconic gestures related to these events, recognized these events more often than children who saw them accompanied by interactive gestures which were semantically unrelated to the events, and children who saw no gesture at all. Our prediction that action memory would be boosted more strongly than actor memory was not borne out statistically. Instead, a main effect of gesture type showed that iconic gestures boosted both action memory and actor memory, but descriptively they boosted actor memory less than action memory.

Gesture Type, Memory Type, and Semiotic Type

We conducted a more in-depth analysis of the effect of seeing different semiotic types of iconic gestures on children's actor recognition memory. We categorized our iconic gestures as representing the feet, legs, or body of the actors (see Figure 4.2 and Table C.1 in Appendix C). We reasoned that gestures which represent actors differently may influence children's memory for actors differently. In our analysis, we compared the iconic gesture condition with the no gesture condition, because the experimenter's hands in the no gesture condition certainly did not represent aspects of the action events (the analyses show the same results when the iconic gesture condition is compared with the interactive gesture condition).

Figure 4.5 shows children's recognition memory performance organized by gesture type, memory type and semiotic type. Children's recognition memory performance was entered in a mixed-effects logistic regression analysis with gesture type as a between-subjects factor and memory type and semiotic type as within-subject factors. Our analysis revealed a significant interaction effect between gesture type, memory type, and semiotic type on children's recognition memory performance, $\chi^2(2) = 6.51, p = .039$. The model explained approximately 23% of the variance in recognition memory performance (marginal $R^2 = .08$, conditional $R^2 = .23$).

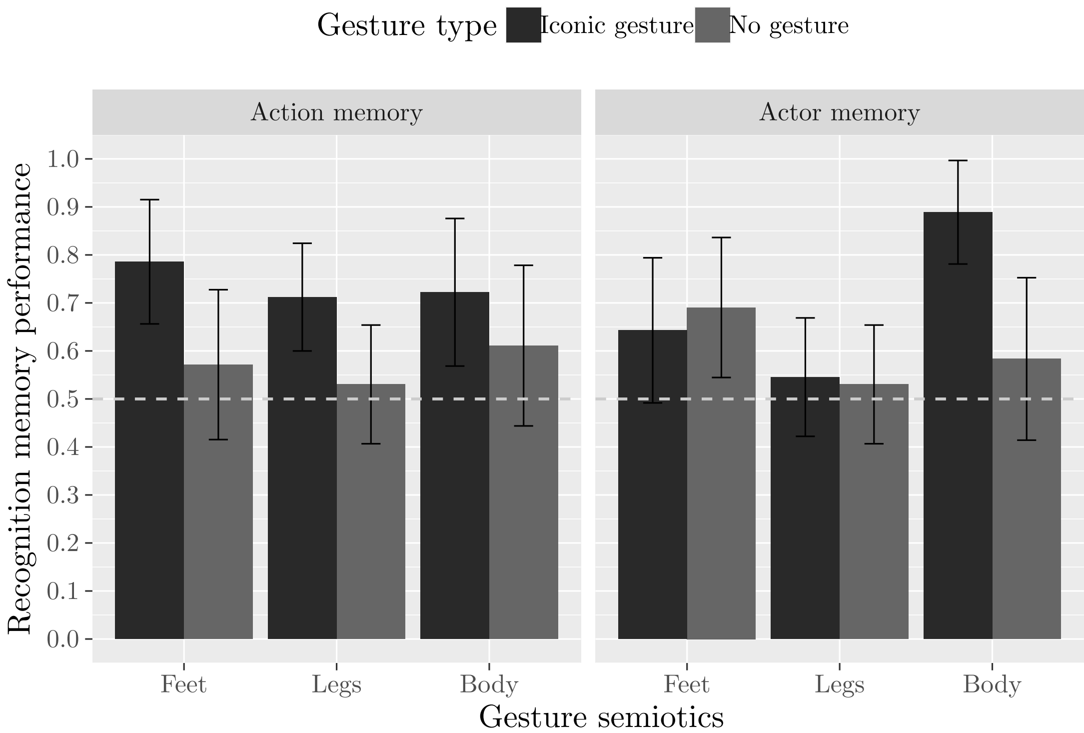


Figure 4.5: Children's recognition memory performance (y-axis shows proportion of correct responses) in the iconic gesture condition (dark grey) and no gesture condition (light grey), organized by memory type (panels) and semiotic type (x-axis). Error bars represent 95% confidence intervals of the means. Dotted line represents chance level.

To further investigate the three-way interaction, we split our data based on memory type (see panels of Figure 4.5). When action memory performance was entered into the analysis (left panel) with gesture type and semiotic type as predictors, the main effect of gesture type was significant, $\chi^2(3) = 10.40, p = .015$, but not the main effect of semiotic type, $\chi^2(4) = 1.42, p = .842$, or the interaction, $\chi^2(2) = 0.50, p = .777$. We predicted that actor memory would increase when more of the actor is represented in gesture (e.g., hand-as-body gestures > hand-as-leg gestures > hand-as-foot gestures). The right panel of Figure 4.5 shows that descriptively, the benefit of iconic gestures (compared to no gesture) on actor memory increases as more of the actor is depicted in gesture. When actor memory performance was entered into the analysis, we found a significant interaction effect between gesture type and semiotic type on actor memory performance, $\chi^2(2) = 6.81, p = .032$. The three-way interaction is thus driven by the interaction effect on actor memory. We compared the size of the iconic gesture vs. no gesture benefit by examining the beta estimates for the interaction effect of each semiotic type (with iconic gesture as a reference point for gesture type and hand-as-body gestures as a reference point for semiotic type, followed by hand-as-leg gestures). Hand-as-body gestures boosted actor memory in comparison to the no gesture condition more strongly than hand-as-leg gestures ($\beta = -1.88, SE = 0.92, p = .040$) and hand-as-foot gestures ($\beta = -2.15, SE = 1.03, p = .038$), which themselves did not differ ($\beta = 0.26, SE = 0.79, p = .739$). Though there was no statistically significant difference between hand-as-foot and hand-as-leg gestures, the overall descriptive trend was as predicted and the two conditions that were predicted to be most different from each other were significantly different from each other. This model explained approximately 27% of the variance in actor memory performance (marginal $R^2 = .11$, conditional $R^2 = .27$). We split the data based on semiotic type to test whether children performed better in the iconic gesture condition than in the no gesture condition. The iconic gesture benefit was significant when the experimenter's hands represented the body of the actors ($\beta = -1.76, SE = 0.65, p = .006$), but not their feet ($\beta = 0.17, SE = 0.68, p = .803$), or legs ($\beta = -0.07, SE = 0.36, p = .857$) (see the right panel of Figure 4.5).

4.4 General Discussion

In order to examine whether seeing iconic gestures can help children to encode *non-linguistic* information, we conducted an experiment in which we tested children's recognition memory of action events. This study has two key findings. First, 3-year-old children who saw action events (videos of actors moving in a novel manner) accompanied by iconic gestures depicting those events, remembered actions (manners) *and* actors better than children who saw the same events accompanied by interactive gestures unrelated to the events, or no gesture at all. Thus, seeing iconic gestures while encoding events facilitates children's memory of these events. Second, the benefit of iconic gesture on actor recognition memory increases as more of the actor is represented in gesture. Thus, iconic gestures boost event memory by schematically highlighting particular aspects of events. More specifically, we argue that iconic gestures facilitate action memory because they encode distinctive features of actions in a schematic manner, thereby drawing children's attention to the actions in a complex event. Hand-as-body iconic gestures facilitated actor memory because they drew children's attention to the actors' whole body, which in turn created more opportunities to pick up person-specific information about the actors.

Our findings go beyond the previous demonstration of the effect of seeing iconic gestures on cognitive processes in the following way. Previous research has shown that observing gesture while encoding *verbally presented* information (e.g., words and explanations) influences the way in which children remember and use this information (e.g., A. E. Booth, McGregor, & Rohlfing, 2008; Goodrich & Hudson Kam, 2009; Mumford & Kita, 2014; So et al., 2012). Fewer studies have investigated the impact of iconic gestures on memory of *nonlinguistic* information, such as event memory. Some of the existing studies have shown that seeing iconic gestures at the recall stage influences children's memory of past events (e.g., Broaders & Goldin-Meadow, 2010; Kirk et al., 2015). However, our study is the first to show that seeing iconic gestures when *encoding* events influences children's memory of those events.

The finding that seeing iconic gestures facilitates action event memory is in line with studies showing that producing iconic gestures facilitates action event memory (Cook, Yip, & Goldin-Meadow, 2010). In their study, Cook et al. (2010) instructed participants to gesture or not to gesture while encoding events shown in short video clips (e.g., a man spinning a bucket). It was found that when participants produced gestures that encoded aspects of the events (e.g., circular movement with the hand shaped as a fist as if holding a bucket), they mentioned more aspects of the events in a free recall memory task than participants who were instructed not to gesture. Producing gestures thus facilitates a stronger memory representation of witnessed events than not producing gesture.

Children's event memory performance in the no gesture condition is consistent with the experiment by Imai et al. (2005) on recognition memory for action events, in which the authors also found that 3-year-old children recognize action events above chance level. However, in our no gesture condition, and in neither of the gesture conditions, did children reach the 84.5% memory accuracy of children in the study by Imai et al. (2005) and this discrepancy needs further explanation. There are three possible reasons for the worse performance of children in the current study. First, children in our study had to encode twice as many events as in the study by Imai et al. (2005). Second, remembering an action and the actor who performs the action (the current study) may be more difficult than remembering an action and the object acted upon (Imai et al.'s study). Encoding an actor (e.g., a person), based on appearance, is more complex than encoding an object, which could be based simply on its shape. Furthermore, the actor cannot be physically separated from the action, whereas the object acted upon can. Third, it is possible that children in the gesture conditions of the current study divided their attention between the gestures and the stimulus videos, which were both in children's field of vision. However, this cannot be the sole reason for the discrepancies with Imai et al.'s results because the children in the no gesture condition, who had no reason to divide their attention, also performed worse than children in Imai et al.'s study.

Can gestures that do not carry meaning relevant to the task boost event memory in children? We argue that interactive gestures in our study did not improve memory performance because they do not encode any information useful for the task. Consistent with this interpretation, Goodrich and Hudson Kam (2009) found that interactive gestures did not help 2-, 3-, and 4-year-old children in a verb learning task. Similarly, So et al. (2012) showed that beat gestures did not help 4-5-year-old children to recall a list of words. However, Lüke and Ritterfeld (2014) showed that children remembered novel character names better when they encoded these names while seeing iconic gestures and "arbitrary gestures" than without gesture. In their study, 3-5-year-olds were introduced to cartoon characters that had distinctive visual features (e.g., a large nose). While the children heard the novel names of the characters, the experimenter produced iconic gestures that encoded the characters' distinctive visual features (e.g., extending the nose with the hand), arbitrary gestures that did not depict such features (e.g., producing a circular motion with an open palm facing inward in front of the face), and no gesture. Children then performed a picture selection task, in which they were presented with pictures of the characters, and required to point at the character that one of the novel names referred to. Children selected more characters correctly when they had seen iconic and arbitrary gestures than when they had seen no gesture. Their result on iconic gestures is compatible with the current finding; the iconic gestures helped children to zero-in on the distinctive features of the cartoon characters. What about their result on arbitrary gestures? We argue that the arbitrary gestures benefited the children because they also helped children focus on the information relevant to the task. The arbitrary gestures were hand movements produced around the face and neck areas, and the characters could be distinguished from each other by features visible in those areas (e.g., a large nose or a long beard). Thus, the arbitrary gestures may have focused children's attention on the (body) parts of the characters where the distinctive features could be seen, and this helped children to map different novel names to the characters. This explanation is similar to our explanation as to why the hand-as-body gestures, which did not encode any actor-specific information, improved actor memory in the current study. Gestures that do not iconically encode the specific relevant information for a task can still improve task performance if they draw children's attention to the part of the event where the useful information can be seen. The arbitrary gestures in the study by Lüke and Ritterfeld (2014) did so via *deixis* (spatio-temporal contiguity) and the hand-as-body gestures in the current study did so via *iconicity* (similarity).

Thus, the current study suggests that iconic gestures can boost memory in two different ways. Iconic gestures helped children focus on key parts of action events, namely the action and some parts of the actor's body, and this focusing had two consequences. For action memory, gestures directly encoded task-relevant information, namely, distinctive features of actions, which left children with a stronger memory trace. This lead to better action recognition memory. For actor memory, gestures did not directly encode task-relevant information, but highlighted particular parts of the event (e.g., the actor's body) that may include task-relevant information, and guided children's attention to these parts. This, in turn, helped children find and encode actor-specific features, leading to better actor recognition memory. It is well-documented that pointing gestures (e.g., Langton, O'Malley, & Bruce, 1996) and the deictic component of iconic gestures (i.e., the location in gesture space at which iconic gestures are produced) (e.g., Sekine & Kita, 2015), can direct the recipient's attention to particular areas of the interactional space. The current study demonstrates for the first time that iconicity in iconic gestures can also direct the recipient's attention to a particular part of an event that includes the referent of the gestures.

The two mechanisms proposed above are based on the fact that iconic gestures convey semantic information (Broaders & Goldin-Meadow, 2010; Goldin-Meadow, 2003; Goldin-Meadow et al., 1993; Hostetter, 2011; Kirk et al., 2015; Mumford & Kita, 2014) by depicting a referent in a schematic manner (Chu & Kita, 2008; de Ruiter, 2000; Goldin-Meadow, 2015; Kita et al., 2017; Novack et al., 2014; Novack & Goldin-Meadow, 2016). Specifically, such schematic representations are efficient in that they help children to focus on a subset of the information useful for the task at hand, which is crucial to how observing gestures promotes cognitive processing (Kita et al., 2017). In the current study, the iconic gestures schematically highlighted the relevant parts of the events, which helped children focus on the information relevant for the action memory and actor memory trials. This is in line with Mumford and Kita's (2014) word learning study, which showed that schematization of events by iconic gesture influences children's interpretation of novel verb meanings. Children interpreted novel verbs as manner verbs when manner was highlighted in iconic gesture, but as change-of-state verbs when end-state was highlighted in iconic gesture. Thus, observing iconic gestures can boost children's event memory and word learning by schematically highlighting the relevant part of complex events. That is, schematization helps children focus on the key information.

The current result may also provide an alternative explanation for the putative finding that gesture *production* influences solving math problems via schematization (Novack et al., 2014). In the study by Novack et al. (2014), children learned how to solve mathematical-equivalence problems on a white board (e.g., $6 + 3 + 8 = ... + 8$). During the training phase, children produced a pre-trained strategy in speech (e.g., “I want to make this side equal to the other side. Six plus three plus eight is seventeen, ...”) with one of three pre-trained hand movements: actions (e.g., moving magnetic numbers to the other side of the equal sign), concrete gestures (e.g., mimicking the movements of the practical actions), or abstract gestures (e.g., grouping the magnetic numbers important for solving the equations). Children in the gesture conditions solved more problems in the paper-and-pencil posttest with new, similar equations than children in the action condition. The authors concluded that gesture production, which is based on schematic representations, leads to deeper and more flexible understanding of mathematical-equivalence problems than actions. However, the gesture production manipulation was confounded with what children *observed* in the pre-instruction phase. In this phase, the experimenter produced the hand movements three times to show children how to move their hands, which children repeated in the subsequent training phase. Thus, it is not clear whether observing or producing gestures influenced the children’s posttest performance. Given the results of the current study, the most parsimonious explanation may be to attribute this effect to seeing gestures. However, there are two caveats. First, in Novack et al.’s study, children saw hand movements only three times during pre-instruction, but produced hand movements 15 times (three times during pre-instruction and 12 times during subsequent training). Thus, it is difficult to distinguish between a potential effect of gesture observation and gesture production in this study. But, an important note here is that gesture observation alone can indeed benefit children’s understanding of mathematical equivalence (Cook, Duffy, & Fenn, 2013). Second, the current study is about recognition memory, but Novack et al.’s study is about learning how to solve problems, thus the mechanisms involved may differ. More research is needed to investigate the beneficial effects of gesture observation and gesture production on children’s memory and learning.

For future research, it may also be interesting to investigate whether iconic gestures can help adults to teach children about fundamental movement skills such as stability (e.g., balancing, twisting), physical fitness (e.g., stretching, bending), locomotor skills (e.g., running, jumping), object manipulation and control (e.g., throwing, catching), and the way the human limbs work. If confirmed, iconic gestures would become an even more useful tool for teaching as the mastery of fundamental movement skills is widely believed to facilitate children's physical, cognitive, and social development and provides the foundation for an active, healthy lifestyle (Lubans, Morgan, Cliff, Barnett, & Okely, 2010).

4.4.1 Conclusion

To conclude, our study demonstrates that iconic gestures at the stage of encoding are meaningful social cues, which can facilitate action event memory in 3-year-old children. The mechanisms that underlie this effect are based on the information that gesture conveys. Gestures schematize particular aspects of complex events, and boost the recognition memory of information that they selectively highlight. This is important as action event memory helps children to construct knowledge of *who* does *what*, which is a key aspect of early social-cognitive development (Burdett et al., 2016; Milward et al., 2014; Schmidt et al., 2016; Vogelsang & Tomasello, 2016).

Chapter 5

Seeing Iconic Gestures Helps Children Use Prior Action Knowledge for Verb Learning

This study investigated what type of prior nonlinguistic action knowledge facilitates subsequent verb learning in 3-year-old children. Experiment 6 showed that children ($N = 96$) learned novel verbs better when pre-exposed to unlabeled exemplars of the referent actions ("retrospective exemplars"), but only if the referent actions were highlighted by iconic gestures. Experiment 7 showed that children ($N = 48$) learned novel verbs better when pre-exposed to retrospective exemplars with iconic gestures than when pre-exposed to two different retrospective exemplars of the same action (i.e., different actors) presented side by side without gesture. Thus, iconic gestures do not merely serve as extra retrospective exemplars, but help children to form action concepts based on schematic representations, which facilitates subsequent verb learning.

5.1 Introduction

Figuring out the meaning of a novel word is challenging for young children. Quine (1960) notes that even in ostensive situations, a novel word could refer to an infinite number of referents. However, this referential ambiguity may be reduced if children have encountered the referent before (even without hearing its label). This study investigates what type of prior action knowledge promotes subsequent verb learning in 3-year-old children.

5.1.1 Verb Learning is Challenging

Verbs typically refer to actions which are hard for children to individuate in complex events (Gentner, 1982). This is evident from studies showing that 3-year-old children struggle to generalize verbs to events that show the referent actions, but novel actors (e.g., Imai, Li, et al., 2008; Kersten & Smith, 2002), novel objects (e.g., Imai et al., 2005), or novel instruments (e.g., Behrend, 1990). This indicates that children's semantic representations of verbs include components of action events that are irrelevant to their meaning (i.e., actors, objects, instruments). Thus, if we can help children to individuate actions, then this should help children learn verbs with adult-like semantic representations that focus on action. This study investigates three ways in which this could be achieved.

5.1.2 Multiple Exemplars of the Same Action May Facilitate Verb Learning

One way to help children individuate actions in complex events is to present them with multiple exemplars that consistently show the components relevant to verb meaning while changing the components irrelevant to verb meaning (Childers, 2011; Forbes & Farrar, 1993, 1995; Haryu, Imai, & Okada, 2011). For instance, in a study by Childers (2011), 2.5-year-olds were taught novel verbs while seeing an experimenter perform target action events (e.g., rolling a ball down a ramp into an opaque box) followed by either the repetition of these exemplars (e.g., rolling a ball down a ramp into an opaque box to make it disappear), similar action exemplars repeating only the actions (e.g., rolling a ball down a curved tube), or similar action exemplars repeating only the results (e.g., covering the ball with a piece of cloth to make it disappear). In the test phase, children were then required to enact the novel verb meanings with a set of objects that included the objects used in the target action events (e.g., a ramp), novel objects that could be used to enact the actions (e.g., a curved pipe), and novel objects that could be used to enact the results (e.g., an opaque bag). Children who saw similar exemplars that repeated the actions were more likely to generalize the verbs to novel objects with which the same actions could be performed, and children who saw similar exemplars that repeated the results were more likely to generalize the verbs to novel objects which led to the same results. In contrast, children who saw the repetition of one exemplar were conservative in generalizing the verbs as they just enacted the actions on the same objects as in the target action events. Thus, this study shows that children's ability to compare exposures to multiple action exemplars may help them to individuate shared components between those exemplars that are important for verb meaning (i.e., manners, results), which facilitates verb learning.

5.1.3 Iconic Gestures that Encode Actions May Facilitate Verb Learning

Another way to help children individuate actions in complex events is to highlight verb meanings with iconic gestures (e.g., Goodrich & Hudson Kam, 2009; Mumford & Kita, 2014). Gestures are iconic when hand movements depict features of objects (i.e., shape) or actions (i.e., manner of motion) (McNeill, 1985, 1992). The form of an iconic gesture and its meaning are linked through similarity (e.g., wiggling the index and middle fingers to depict walking). As such, iconic gestures can focus children's attention on components of action events that are important for verb meaning. For example, in a study by Mumford and Kita (2014), 3-year-old children were taught novel verbs that could be interpreted as manner verbs (e.g., "to push") or result verbs (e.g., "to break"). Children saw videos of an actor manipulating objects (e.g., sprinkling sand into a square shape) with either iconic manner gestures (e.g., depicting the manual action of sprinkling) or iconic end-state gestures (e.g., tracing the square shape) while the experimenter labeled each action event with a novel verb. Subsequently, children were asked to generalize the novel verb to one of two videos in a forced-choice task: one video showed a manner verb interpretation (e.g., sprinkling powder into a triangle shape) and the other a result verb interpretation (e.g., a square made of pieces of paper). Children who saw iconic manner gestures when the verbs were taught interpreted the verbs as manner verbs and children who saw iconic end-state gestures as result verbs. Thus, iconic gestures can focus children's attention on components of an action event that are important for verb meaning (i.e., manners and results), which facilitates verb learning.

These studies on multiple exemplars and iconic gestures, however, leave it open whether prior (non-linguistic) action knowledge can promote subsequent verb learning. In previous verb learning experiments, novel actions were always labeled on each encounter (e.g., Childers, 2011; Haryu et al., 2011; Imai et al., 2005; Maguire et al., 2008; Mumford & Kita, 2014). This set-up does not allow children to integrate prior non-linguistic action knowledge in their semantic representation of a verb when this action is subsequently labeled. The current study experimentally investigates this step of the verb learning process for the first time. In doing so, our study addresses a fundamentally different question than verb learning studies in which children were exposed to multiple labeled action exemplars (e.g., Childers, 2011; Haryu et al., 2011; Maguire et al., 2008). Such studies examined how initial linguistic representations of verbs influence how children process subsequent linguistic representations of these verbs. However, our study investigates how prior (non-linguistic) representations of actions can influence how children process initial linguistic representations of those actions when they are labeled with a novel verb. Thus, our study investigates a step in the verb learning process that precedes fast mapping (Carey & Bartlett, 1978) in the initial labeling occasion.

5.1.4 Pre-Exposure to Referents May Facilitate Word Learning

Studies on noun learning provide evidence that Quine's ambiguity problem (1960) may be reduced when children have seen the referent objects repeatedly (Clerkin, Hart, Rehg, Yu, & Smith, 2017; Graham, Turner, & Henderson, 2005; Kucker & Samuelson, 2011). For example, in a study by Clerkin et al. (2017), a head-mounted camera recorded what 8-10-month-old infants saw during mealtimes. Though most of the time, the child's view was highly cluttered with objects, a frame-by-frame analysis showed that only a small set of objects was repeatedly present in the child's view. Receptive vocabulary norms revealed that these high-frequency objects are possible referents of the first nouns that children learn. This study demonstrates that not all components of a scene that children see are equal contenders for the novel words they hear. Thus, in the case of noun learning, pre-exposure to objects seems to significantly reduce referential ambiguity, though it is not clear whether pre-exposures were accompanied by linguistic labels, because Clerkin et al. (2017) did not analyze audio recordings.

Furthermore, it is not clear whether the suggestive evidence for nouns (e.g., Clerkin et al., 2017; Graham et al., 2005; Kucker & Samuelson, 2011) extends to verbs, because the nature of pre-exposure to verb referents and noun referents is fundamentally different (Childers & Tomasello, 2002; Gentner, 1982; Hirsh-Pasek & Golinkoff, 2006; Imai et al., 2005; Imai, Li, et al., 2008; Kersten & Smith, 2002; Maguire et al., 2002). Children observe the same small set of objects every day (e.g., objects in and around their house). That is, they are often exposed to the same object exemplar multiple times. When children hear a novel noun that refers to an object in such situations, it is easy for them to link the current exposure and prior exposures to the same object, which helps them to form a concept for the object. But, children are not exposed to the same action exemplar every day, unless they are recorded and shown to them again (e.g., Maguire et al., 2008). Thus, we assume that under natural circumstances, children see each action exemplar only once. When children hear a novel verb that refers to an action, it is hard for them to link multiple exemplars of the same action, especially when there is no external cue to link exemplars (e.g., a verb heard with each exemplar). Furthermore, as each action exemplar is encountered only once, children need to extract the invariance of action across exemplars to form an action concept. So, we do not know under which conditions children can take advantage of pre-exposure to unlabeled action exemplars when learning verbs, and this study investigates this question.

5.1.5 Possible Mechanism

Iconic gestures that depict actions may help children to extract the invariance of action between exemplars of the same action. Kita et al. (2017) suggest that iconic gestures schematize a subset of information that is potentially relevant to the task at hand (see also Chu & Kita, 2008; de Ruiter, 2000; Goldin-Meadow, 2015; Novack & Goldin-Meadow, 2016; Novack et al., 2014). Schematization by gesture is a form of abstraction–in the case of verb learning, iconic gestures that depict actions strip action events from components that are irrelevant to verb meaning (e.g., actors), while maintaining the components that are relevant to verb meaning (e.g., manners of motion). Schematic representations of action events depicted in iconic gestures can thus help children to individuate actions. Such schematized information is "light-weight" and efficient for memory (Kita et al., 2017). Indeed, seeing iconic gestures promotes children's memory of action events (e.g., Aussems & Kita, in press) and route directions (e.g., Austin & Sweller, 2017). Schematic representations of actions depicted in iconic gestures can thus help children to recognize actions in future exemplars of the same actions they may encounter. To summarize, iconic gestures help children to individuate actions in a schematic manner, and children may thus recognize these actions in exemplars of the same action and this helps them to form action concepts, which facilitates subsequent learning of linguistic labels for these actions.

5.2 The Current Research

The current study investigates if pre-exposure to unlabeled actions promotes subsequent verb learning and if seeing iconic gestures with the unlabeled actions influences this process. In Experiment 6, we manipulated pre-exposure to actions and the gesture type that children saw with these exposures. We presented children with a novel verb learning task that had three phases. In the *initial phase* of the task, children were exposed to unlabeled action exemplars which they either saw with iconic gestures that highlighted the actions in the exemplars or interactive gestures (Bavelas et al., 1992) that did not depict any aspect of the exemplars. In the following *label* *phase*, children were either taught verbs with exemplars that showed novel actors performing the actions they were pre-exposed to in the initial phase (retrospective exemplar condition) or with exemplars that showed novel actors performing novel actions that were not pre-exposed to in the initial phase (irrelevant exemplar condition). In the following *test phase*, children's knowledge of the novel verb meanings was tested in a two-way forced choice generalization task. The choice was between an exemplar of a novel actor performing the referent action and an exemplar of the same actor as in the label phase performing a novel action.

We have three predictions. First, children who see retrospective exemplars with interactive gestures will be more successful in the verb generalization task than children who see irrelevant exemplars with interactive gestures. This is because children's have a better chance of forming an action concept (i.e., extracting the invariance of action) when they see two exemplars of the same action (i.e., two different actors performing the same action) than children who see two exemplars of different actions (i.e., two different actors performing different actions). Second, children who see retrospective exemplars with iconic gestures will be more successful in the verb generalization task than children who see retrospective exemplars with interactive gestures. This is because iconic gestures depict how the actors move, which helps children to individuate actions and extract the invariance of action between two different exemplars. Third, children who see retrospective exemplars with iconic gestures will be more successful in the verb generalization task than children who see irrelevant exemplars with iconic gestures. This outcome would support the idea that iconic gestures schematize a subset of information relevant to the task at hand, and importantly, do not benefit verb learning through teaching children a general strategy (i.e., the task is about looking at actions). Any advantage of the retrospective exemplar conditions over the irrelevant exemplar conditions would support the idea that prior non-linguistic action knowledge promotes subsequent verb learning.

Experiment 7 examined whether seeing retrospective exemplars with iconic gestures facilitates subsequent verb learning because gestures schematize a subset of information relevant to the task at hand, or because iconic gestures function as extra retrospective exemplars. We manipulated the type of pre-exposure to actions. We showed a new group of children retrospective exemplars with iconic gestures to replicate Experiment 6. A second group of children were taught verbs after seeing two retrospective exemplars (i.e., two different actors performing the same action), which were presented side by side without gesture. We predict that children who see retrospective exemplars with iconic gestures will generalize verbs to novel events showing the referent actions more successfully than children who see two different retrospective exemplars. This outcome would support the idea that iconic gestures help children to form action concepts based on schematic information, and importantly, that they go beyond serving as extra retrospective exemplars (i.e., extra opportunity to extract the invariance of action).

5.3 Experiment 6

5.3.1 Method

Design

The experiment had a 2 x 2 between-subjects design with pre-exposure to actions (retrospective exemplars vs. irrelevant exemplars) and gesture type (iconic gesture vs. interactive gesture) as independent variables. The dependent variable was children's verb generalization performance in each of six trials (binary: 1=correct, and 0=incorrect).

Participants

The data were collected between the 23rd of March 2016 and the 27th of September 2016. Our sample size was determined a priori using G$*$Power version 3.1 (odds ratio= 2.30, $\alpha = 0.05$, power= 0.80) (Faul et al., 2009). The final sample included 96 typically developing children (49 girls) between 36–48 months old ($M = 41.14$, $SD = 3.70$). An additional nine children were tested, but excluded from the analysis because they were too old on the day of testing ($N = 6$) or pointed exclusively to the right or left side of the screen in the test procedure ($N = 3$). Participants were recruited via 11 nurseries in the West-Midlands and Warwickshire areas (United Kingdom), and via a database of families who expressed interest taking part in language development research. Children's age in months did not differ between conditions, $F(1, 94) = 0.00, p = .999$, nor did their gender, $\chi^2(3) = 0.13, p = .989$. Three groups counted 12 boys and 12 girls and one group 11 boys and 13 girls. All children were exposed to the English language for >75% of the time (as indicated by their caregivers). Informed consent was obtained for all participants. Nurseries received a voucher for their participation and children who participated in the research lab received a certificate and a toy. All studies reported in this paper were approved by the Humanities and Social Sciences Ethics Committee of the University of Warwick.

Materials

A set of 96 videos (4–15 seconds) depicting 24 novel actions was taken from the GRACE database (Aussems et al., in press, 2017). Stimulus videos showed 24 actors (12 males, 12 females) moving across the length of a scene in an unusual manner using their feet, legs, or body. The actors always kept their arms by their side, fingers pointing downward, parallel to their torso. The unusual actions were normed based on the match between iconic gestures and actions, the similarity between different actors performing the same actions and the same actors performing different actions, and how unusual they appeared to be to adult native English speakers (for more detail, see Aussems et al., in press, 2017). Each action was depicted by two male actors and two female actors in separate videos. The file names of these videos and details about when each video was shown in the experiment task are available on the Open Science Framework at `https://osf.io/t52cn/?view_only=4f4ffb66c68e460ca58bdcc928b35144`. The novel words that were used to label the actions were *daxing*, *blicking*, *larping*, *stumming*, *pilking*, and *krading*. These words follow the phonotactical rules of English and are commonly used in word learning paradigms (e.g., Childers, 2011; Maguire et al., 2008; Mumford & Kita, 2014; Roseberry, Hirsh-Pasek, Parish-Morris, & Golinkoff, 2009).

Randomization and Counterbalancing

We created 24 versions of the experiment in which every stimulus video appeared equally in the initial phase, the label phase, and in the test phase (as target and distractor stimulus). Furthermore, male actors and female actors were equally represented in the stimuli videos, target videos appeared equally on the left and ride sides of the screen, and the order of trials was randomized. We administered each experiment version to one child in each condition before moving on to the next version. Participants recruited from different nurseries and the research lab are therefore equally represented in each condition. Participants were pseudo-randomly assigned to conditions based on their gender and age in months, before the experimenter met the children.

Procedure

In the initial phase of the experiment task, children watched six videos with an experimenter in a quiet area of the nursery or in the research lab. When a video played for the first time in the initial phase the experimenter said “Wow! Look at what he (or she) is doing!”, and when the video played again the experimenter said

“Oh! He (or she) is doing it again!”. Depending on the condition, the experimenter accompanied these utterances with either an iconic gesture or an interactive gesture (Bavelas et al., 1992). Note that the experimenter did not label the action events in the initial phase.

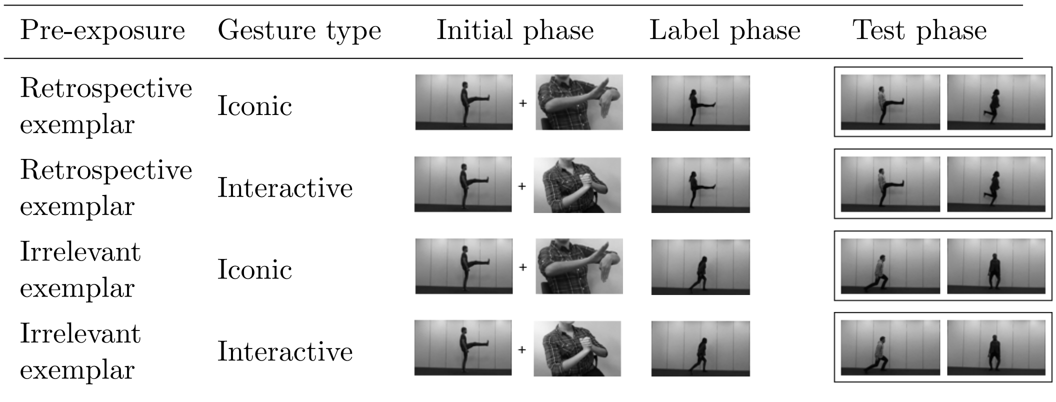


Figure 5.1: The procedure of Experiment 6. From left to right, the overview shows Pre-exposure to actions (Column 1) and Gesture type (Column 2), followed by examples of what children saw in the Initial phase (Column 3), the Label phase (Column 4), and the Test phase (Column 5). Note that actors always changed between the phases of the experiment task, but that distractor videos in the Test phase (most right in the overview) showed actors from the Label phase. Actions were never labeled in the Initial phase. After the Initial phase, there was a five-minute break, during which children performed an irrelevant filler task.

Figure 5.1 shows an example of an iconic gesture and an interactive gesture. There were 24 iconic gestures (for video examples see Aussems et al., in press, 2017) and three interactive gestures. Iconic gestures (McNeill, 1985, 1992) depicted how the actors in the videos moved. Interactive gestures (Bavelas et al., 1992) were not related to any aspect of the action events, but communicated excitement and surprise. Video examples of the three interactive gestures can be viewed at `https://osf.io/t52cn/?view_only=4f4ffb66c68e460ca58bdcc928b35144`, but still frames of these videos can also be seen in Appendix D. A female experimenter rotated the three interactive gestures for the six trials in the same way between conditions. The experimenter was always seated on children’s left side at a low table and produced all gestures live during the task. Gestures were produced to the left side children’s field of vision, so that they could see the stimulus videos and gestures simultaneously.

After the initial phase, children spent five minutes decorating a paper wristband with stickers. The experimenter asked the children to name the colors of the stickers and to count the number of stickers on the wristband during this distraction task.

In the following label phase, children watched six videos of novel actors. We alternated actor gender between exemplars of the same action to make it clear that the actors changed (see Figure 5.1). The experimenter now labeled the way these actors moved with a novel verb: "Look! He (or she) is [*daxing*]!". The video automatically played a second time and the experimenter labeled the action again: "Wow! He (or she) is [*daxing*] again!". Depending on the condition, children were either taught labels for actions that they had seen during the initial phase (retrospective exemplars) or novel actions that they had not seen before (irrelevant exemplars). Note that the experimenter gestured during the initial phase, but not during the label phase. After hearing the action label twice, children's knowledge of the verb's meaning was tested immediately in a test phase that included two videos side by side. One video showed a novel actor performing the target action, and the other video showed the actor from the label phase performing a novel action (see Figure 5.1). The videos started playing automatically at the start of each trial and the experimenter asked the child: "Which one is [*daxing*]?". The experimenter looked at the child while making this request and did not look at the screen until the child pointed at an answer. The videos played continuously on loop until the child picked one. If the child did not respond or asked the experimenter whether a particular video showed [*daxing*], the question was repeated until one video was chosen. If the child pointed at both videos, the experimenter asked the child to pick one.

Stimuli were displayed using slide presentation software Microsoft Office PowerPoint 2016 on a 14" touch screen laptop. Using on-screen buttons and Visual Basic for Applications (VBA), the data were automatically saved.

Data Analysis

Verb generalization performance (binary: 1=correct, 0=incorrect) was entered into a mixed-effects logistic regression analysis using the *lme4* package (D. Bates et al., 2013) in the R software for statistical analyses (R Development Core Team, 2011). Fixed factors included pre-exposure to actions (retrospective exemplars vs. irrelevant exemplars) and gesture type (iconic gesture vs. interactive gesture). The model included a maximal random-effects structure (cf. Barr et al., 2013), i.e. random slope and intercept variation, and the co-variance between the two, for participants and items (e.g. the stimulus videos that were labeled with a novel verb). Likelihood ratio tests (χ^2) were used to compare the full model with updated versions of the model that systematically excluded the main effects and interaction terms of interest. Comparisons between groups were made by running the main analysis for subsets of the data that included the groups of interest. One-sample t-tests in which equal variance was assumed were calculated with the built-in *t.test()* function. The raw data and R Markdown files for all plots and analyses are available from the Open Science Framework at `https://osf.io/t52cn/?view_only=4f4ffb66c68e460ca58bdcc928b35144`.

5.3.2 Results and Discussion

Figure 5.2: Verb generalization performance (y-axis shows the proportion of correct responses) organized by pre-exposure to actions (x-axis) and gesture type (shades of grey). Error bars represent 95% confidence intervals of the means. Dotted line represents chance level.

Figure 5.2 shows children's verb generalization performance (in proportion) by pre-exposure to actions (retrospective exemplars vs. irrelevant exemplars) and gesture type (iconic gesture vs. interactive gesture). Children's verb generalization performance (binary: 1=correct, 0=incorrect) was entered into a logistic regression analysis with pre-exposure to actions and gesture type as predictors. The interaction effect between pre-exposure to actions and gesture type on children's verb generalization performance was significant, $\chi^2(1) = 7.50, p = .006$.

Next, we carried out three comparisons to investigate our predictions. First, children who saw retrospective exemplars with interactive gestures did not generalize significantly more verbs to novel events showing the referent actions than children who saw irrelevant exemplars with interactive gestures, $\beta = -0.06, p = .830$, 95% CI $[-0.61, 0.49]$. Second, children who saw retrospective exemplars with iconic gestures generalized significantly more verbs to novel events showing the referent actions than children who saw retrospective exemplars with interactive gestures, $\beta = -1.18, p < .001$, 95% CI $[-1.92, -0.53]$. Third, children who saw retrospective exemplars with iconic gestures generalized significantly more verbs to novel events showing the referent actions than children who saw irrelevant exemplars with iconic gestures, $\beta = -1.12, p < .001$, 95% CI $[-1.74, -0.57]$.

Finally, we investigated whether children in each group performed better than chance (see the dotted line in Figure 5.2). Children's verb generalization performance was averaged across six trials and entered into separate one-sample t-tests (two-tailed) with a test value of 0.5 (i.e., the proportion of correct answers at chance level). Children who saw retrospective exemplars with iconic gestures performed significantly better than chance, $t(23) = 5.59, p < .001$, 95% CI $[0.66, 0.85]$, but not children who saw irrelevant exemplars with iconic gestures, $t(23) = 0.35, p = .732$, 95% CI $[0.43, 0.60]$, neither children who saw retrospective exemplars with interactive gestures, $t(23) = 0.41, p = .684$, 95% CI $[0.42, 0.63]$, nor children who saw irrelevant exemplars with interactive gestures, $t(23) = 0.16, p = .877$, 95% CI $[0.42, 0.60]$.

Children used prior non-linguistic action knowledge for subsequent verb learning, but only if the referent actions (and not irrelevant actions) were highlighted by iconic gestures. Thus, iconic gestures do not promote a general strategy (i.e., the task is about looking for actions). Rather, iconic gestures help children to extract the necessary information for verb learning (i.e., the invariance of action) from unlabeled and labeled exemplars of the same action.

5.4 Experiment 7

Experiment 7 examined whether seeing retrospective exemplars with iconic gestures helps children to form action concepts with schematic information or whether iconic gestures simply give children extra opportunity to form action concepts (i.e., iconic gesture as an extra retrospective exemplar).

5.4.1 Method

Design

The experiment had a between-subjects design with type of pre-exposure to actions as the independent variable. One group of children was presented with retrospective exemplars and iconic gestures in the same way as in Experiment 6 and a second group of children was presented with two different retrospective exemplars (i.e., two different actors), which were presented side by side without gesture. The dependent variable was the children's verb generalization performance in six trials (binary, 1=correct, and 0=incorrect).

Participants

The data were collected between the 22nd of October 2016 and the 14th of February 2017. Our sample size was determined a priori using G∗Power version 3.1 (odds ratio= 2.30, $\alpha = 0.05$, power= 0.80) (Faul et al., 2009). The final sample included 48 typically developing children (22 girls) between 36-47 months old ($M = 39.84$, $SD = 3.30$). An additional four children were tested, but were excluded from the analysis because they were too old on the day of testing ($N = 1$), pointed exclusively to answers on the left side or right side of the screen ($N = 2$), or were diagnosed with a language disorder ($N = 1$). Participants were recruited via the same nurseries as in Experiment 6, and via a database of families who expressed interest taking part in language development research. Children's age in months did not differ between the four conditions, $F(1, 46) = 0.03, p = .861$, nor did their gender, $\chi^2(1) = 0.08, p = .772$. There were 12 boys and 12 girls in the iconic-gesture group and 14 boys and 10 girls in the control group. All children were exposed to the English language for >75% of the time (as indicated by their caregivers). Informed consent was obtained for all participants. Nurseries received a voucher for their participation and children who participated in the research lab received a certificate.

Materials

The materials were the same as in Experiment 6.

Randomization and Counterbalancing

Randomization and counterbalancing were the same as in Experiment 6.

Procedure

In the iconic-gesture group, children were presented with retrospective exemplars and iconic gestures in the initial phase in the same way as in Experiment 6. In the control group, children were shown two different retrospective exemplars (i.e., two different actors) in the initial phase, which were presented side by side and without gesture (see Figure 5.3). When the videos played for the first time the experimenter said "Wow! Look at what they are doing!", and when the videos played again the experimenter said "Oh! They are doing it again!". The label phase and test phase followed the same procedure as in the iconic-gesture condition.

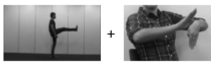
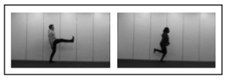
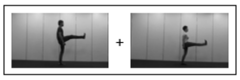
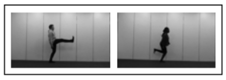
Type of prior experience	Initial phase	Label phase	Test phase
Retrospective exemplar + Iconic gesture			
Two different retrospective exemplars			

Figure 5.3: The procedure of Experiment 7. The overview shows Type of pre-exposure to actions (Column 1), followed by examples of what children saw in the Initial phase (Column 2), the Label phase (Column 3), and the Test phase (Column 4). Exemplars of the same action always showed a different actor, but distractor videos (most right in the overview) in the Test phase showed actors from the Label phase performing a novel action. In the condition with two different retrospective exemplars, two videos were presented simultaneously side by side in the Initial phase.

Data Analysis

The data were analyzed in the same way as in Experiment 6.

5.4.2 Results and Discussion

Figure 5.4 shows children's verb generalization performance (in proportion) by type of pre-exposure to actions. Verb generalization performance (binary: 1=correct, 0=incorrect) was entered into a logistic regression analysis with type of pre-exposure to actions as a predictor. The main effect of type of pre-exposure to actions on children's verb generalization performance was significant, $\chi^2(1) = 6.63, p = .014$. Children who saw retrospective exemplars with iconic gestures generalized significantly more verbs to novel events showing the referent actions than children who saw two different retrospective exemplars, $\beta = -0.67, p = .013$, 95% CI $[-1.23, -0.14]$.

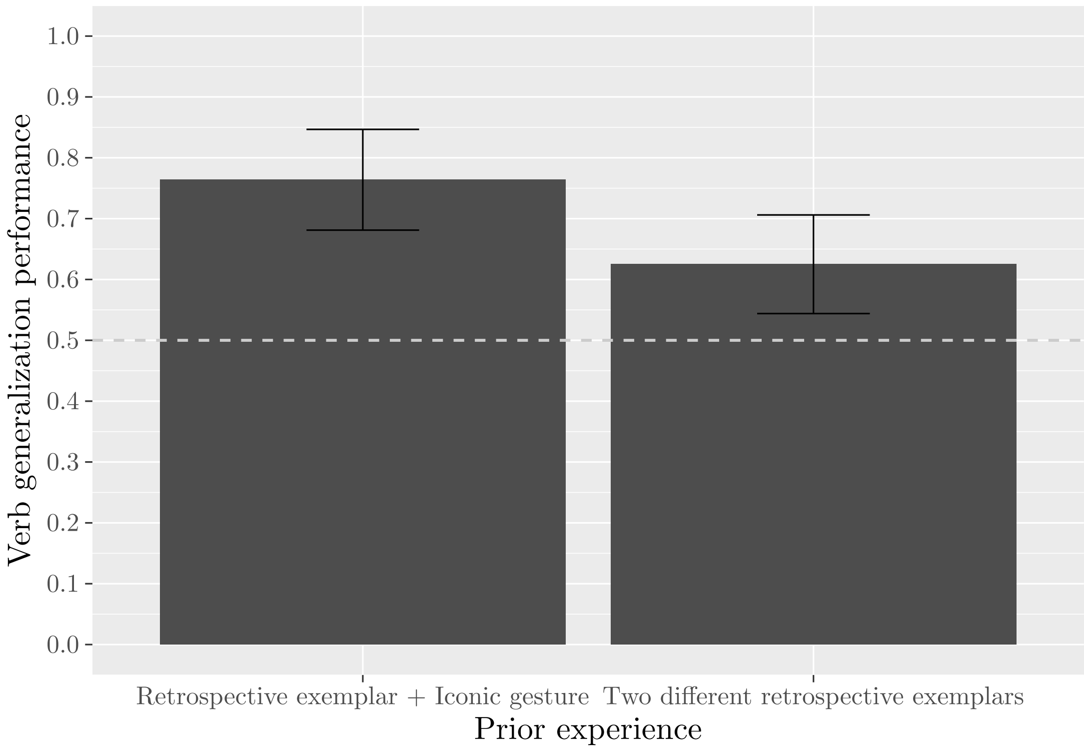


Figure 5.4: Verb generalization performance (y-axis shows proportion of correct responses) organized by type of pre-exposure to actions (x-axis). Error bars represent 95% confidence intervals of the means. The dotted line represents performance at chance level.

Next, we investigated whether children in each condition performed better than chance (see the dotted line in Figure 5.4). Children's verb generalization performance was averaged across six trials and entered into separate one-sample t-tests (two-tailed) with a test value of 0.5 (i.e., the proportion of correct answers at chance). Children who saw retrospective exemplars with iconic gestures performed significantly better than chance, $t(23) = 6.59, p < .001$, 95% CI $[0.68, 0.85]$, and so did children who saw two different retrospective exemplars, $t(23) = 3.19, p = .004$, 95% CI $[0.54, 0.71]$.

Children who saw retrospective exemplars with iconic gestures outperformed children who saw two different retrospective exemplars (i.e., two actors performing the same action) side by side without gesture in a subsequent novel verb learning task. Thus, iconic gestures help children to extract the invariance of actions between unlabeled and labeled exemplars. Importantly, iconic gestures do not serve merely as extra retrospective exemplars (i.e., an extra opportunity to extract the invariance of action).

5.5 General Discussion

This study examined if pre-exposure to unlabeled actions promotes subsequent verb learning in 3-year-old children and if seeing iconic gestures influences this process. We measured how successful children are in generalizing a newly learned verb to a novel event showing the referent action (the same action as in the training video) performed by a novel actor. There are four key findings. First, children who saw retrospective exemplars with interactive gestures were not more successful than children who saw irrelevant exemplars with interactive gestures (Experiment 6). Thus, a retrospective exemplar shown with an interactive gesture that did not depict any action information did not facilitate verb learning. Second, children who saw retrospective exemplars with iconic gestures were more successful than children who saw retrospective exemplars with interactive gestures (Experiment 6). Taken together, pre-exposure to unlabeled referent actions (retrospective exemplars) promoted subsequent verb learning in 3-year-old children, but only when the referent actions were highlighted with iconic gestures. Third, children who saw retrospective exemplars with iconic gestures were more successful than children who saw irrelevant exemplars with iconic gestures (Experiment 6). Thus, iconic gestures did not teach children a general strategy (i.e., that the task was about looking for actions), but schematized a subset of information relevant to the task at hand. Fourth, children who saw retrospective exemplars with iconic gestures were more successful than children who saw two different retrospective exemplars (i.e., two actors performing the same action) presented side by side without gesture (Experiment 7). Thus, iconic gestures do not merely serve as extra retrospective exemplars, but help children to form action concepts based on schematic information, which facilitates subsequent verb learning.

5.5.1 Prior Nonlinguistic Action Knowledge Facilitates Verb Learning

The findings from the current study contribute to the existing literature in three important ways. First, the idea that children use retrospective exemplars for word learning is novel. Retrospective exemplars are exposures to word referents that occur before these referents receive a linguistic label. Previous research has suggested that repeated exposure to objects reduces referential ambiguity and facilitates noun learning in young children (e.g., Clerkin et al., 2017; Graham et al., 2005; Kucker

& Samuelson, 2011), but these are cases of pre-exposure to the same exemplar. In contrast, a retrospective exemplar is a different exemplar (in our study, action performed by a different actor) from a subsequently encountered one that was labeled. Furthermore, as the analysis by Clerkin and colleagues only included what children saw and not what children heard it was unclear whether the referents were labeled. Our study is the first to show that children can use unlabeled retrospective exemplars of the referent action for subsequent verb learning. Second, retrospective exemplars help us to investigate a crucial step of the verb learning process; a step in which relevant conceptual knowledge is formed prior to linguistic labeling. As such, children's use of retrospective exemplars highlights an additional ability that children bring to word learning, beyond fast mapping (Carey & Bartlett, 1978) and cross-situational learning (Yu & Smith, 2008), both of which concern children's exposure to word-referent combinations. Third, iconic gestures may be an important cue for children to link multiple exemplars of the same action. Previous studies have shown that iconic gestures can help children learn verb meanings in one-shot learning tasks with a single exemplar (e.g., Mumford & Kita, 2014), but not in multiple exemplar learning tasks. Our findings show that children's ability to integrate information from multiple sources plays a role important in linking exemplars for word learning. Our study shows this is true even for exposures to referents that have not yet received a linguistic label.

5.5.2 Iconic Gestures Help Children to Link Multiple Exemplars of the Same Action

Iconic gestures helped children to extract the invariance of action between exemplars, which helped them to form action concepts and facilitated subsequent verb learning. This is especially evident in Experiment 7, which shows that children benefit more from seeing retrospective exemplars with iconic gestures in a novel verb learning task than from two different retrospective exemplars of the same action (i.e., two different actors) shown side by side without gesture. The mechanism that underlies this effect is that iconic gestures provide "top-down" guidance in forming action concepts based on schematic information. Children used those action concepts to identify the same actions in subsequently labeled exemplars, which facilitated mapping a label to these actions and extending this label to novel events showing the referent actions. Children who saw two different retrospective exemplars of the same action performed by different actors needed to extract invariance through a "bottom-up" process, in which they needed to consider all elements in the two exemplars as possible invariant components. Thus, they may not have extracted action as the invariant component and possibly had traces of the actors' appearances lingering in their conceptual representation of the two retrospective exemplars. This may have prevented them from extending the newly learned verbs to events showing the referent actions performed by novel actors. To summarize, iconic gestures helped children to create an action concept from one retrospective exemplar, which helped them to form adult-like semantic representations of verbs that focus on action in the subsequent labeling phase.

5.5.3 Retrospective Exemplars: The First Step in Word Learning?

Should retrospective exemplars generally be considered as a first step in the word learning process? We argue that, under the right conditions, children can indeed benefit from unlabeled exposures to word referents in subsequent word learning contexts. We showed this for verb learning in the current study, but our findings could also explain the putative finding from Clerkin et al. (2017) that the objects children see frequently are possible referents for the first nouns they learn, although it was unclear whether parents labeled the objects in their child's view or not. Children could explore two possible routes for word learning, which are not mutually exclusive. First, children could form concepts of referents through exposures (that do not involve labeling) before they learn linguistic labels for those referents. Second, children could form concepts of referents and semantic representations of words in parallel. Indeed, the act of labeling itself promotes creation of object categories in infants (A. Booth & Waxman, 2002; Fulkerson & Haaf, 2003; Waxman & Markow, 1995). It could be interesting to investigate whether seeing iconic gestures with unlabeled exemplars of word referents also leads to categorical knowledge about these referents, and future referents that belong to the same category. Furthermore, when adults provided further conceptual knowledge (e.g., animate vs. artifacts) about objects after these objects were labeled with novel nouns, the type of conceptual knowledge they provided shaped infants' semantic representations, leading to different patterns in noun generalization (A. Booth, Waxman, & Huang, 2005). It may be the case that for words that children acquire early, children form concepts and semantic representations simultaneously, but for words that children acquire later, children form concepts before they form semantic representations. The route children explore may also depend on type of referents (i.e., are they referents for verbs or nouns). This is because exposures to noun referents and verb referents are fundamentally different. Children often see one object exemplar repeatedly, which makes it easy to form an object concept, but action exemplars naturally always look different, which makes it hard to form an action concept. Children may be more likely to form concepts of referents before learning the linguistic labels for these referents if the conditions allow them to link exemplars of the same referents. However, it is unclear whether the same conditions help children to link exemplars of word referents from different word classes (e.g., verbs, nouns, prepositions). Thus, more research is needed to investigate the conditions under which children make us of prior (nonlinguistic) knowledge for word learning.

5.5.4 The Saliency of Actors Prevents Verb Learning

Our finding from Experiment 6 that children who saw retrospective exemplars with interactive gestures did not outperform children who saw irrelevant exemplars with interactive gestures in a novel verb learning task did not support our prediction. One possible explanation for this outcome is that children were not able to extract the invariance of action between exemplars in the retrospective exemplar condition, because they only heard a novel verb with one of the exemplars so there was no "cue" to compare exemplars of the same action (Childers, 2011). However, this is unlikely because children who saw retrospective exemplars with iconic gestures could compare unlabeled and labeled exemplars of the same action without such a cue. Another possible explanation is that children were not able to extract the invariance of action between exemplars, because the saliency of seeing a novel actor in each exemplar prevented children from individuating action and mapping a novel verb to action alone. This interpretation is consistent with findings from a study by Maguire et al. (2008), who showed in a novel verb learning task that 3-year-old children were better at extracting the action component from multiple exemplars in which the same actor performed the same action than from multiple exemplars in which different, novel actors performed the same action. This interpretation may also explain why children who saw two different retrospective exemplars in Experiment 7 barely performed above chance level in the novel verb learning task. Even when exemplars are shown side by side, which is a structural cue for children to compare exemplars, it is difficult for them to extract the action component when those exemplars show novel actors.

5.5.5 Conclusion

We conclude that multiple sources of information about verb referents can be provided to children before these referents are labeled, and children use these sources of information to learn the linguistic labels for these referents, but only if they are structured in a particular way. Helping children to individuate actions is key to verb learning; pre-exposing children to action exemplars while the actions are highlighted in iconic gestures is the best way to do so, but pre-exposing them to two exemplars side by side without gesture helps too. Importantly, iconic gestures helped children to form action concepts with schematic representations of actions, which led to more adult-like semantic representations of individual verb meanings in a subsequent verb learning task than side by side presentation of exemplars. The way children interpret verbs (i.e., what is included in their semantic representations) is a product of how they form action concepts. Under the right circumstances, children can form action concepts before linguistic labels for these actions are introduced, which facilitates mapping those labels to their referents. Thus, prior nonlinguistic action knowledge may be an important first step of the verb learning process.

Chapter 6

Seeing Iconic Gestures Promotes Lasting Word-Category Knowledge about Verbs in Children

Preschool-aged children benefit from seeing iconic gestures during verb learning when these gestures depict the actions to which the verbs refer. This study investigates if this advantage is limited to word-specific knowledge of individual verbs, or extends to word-category knowledge about what verbs are. We found that in a novel verb learning task, children who saw iconic gestures depicting the referent actions, generalized more verbs to novel events showing the referent actions than children who saw interactive gestures. More importantly, the iconic-gesture group outperformed the interactive-gesture group even in subsequent trials that were administered immediately (Experiment 8) and after a one-week delay (Experiment 9), in which all children saw interactive gestures when a new set of different novel verbs was taught. We conclude that seeing iconic gestures promotes lasting word-category knowledge that verbs typically refer to actions. Thus, iconic gestures help children to figure out how to learn verbs.

6.1 Introduction

Verbs are an important part of speech. Recognizing a verb is often a key step in understanding the meaning of a sentence. As such, verbs play a vital role in children's acquisition of grammar and vocabulary. Furthermore, children's early language skills and vocabulary size predict later academic success (Anderson & Freebody, 1979; Morgan et al., 2015; Rowe et al., 2012). It is therefore crucial to understand how children learn verbs and this study focuses on this phenomenon in preschool-aged children.

6.1.1 Verb Learning is Challenging

Verb learning presents a challenge to children because verbs refer to relational meanings (Gentner, 1978). Verbs typically refer to relationships between people, objects, and instruments (e.g., "a woman is cutting paper with scissors"), which are hard for children to individuate in the world (Gentner, 1981, 1982; Gentner & Boroditsky, 2001). For example, children find it difficult to extend newly learned verbs to events that show the referent actions, but novel actors (Imai, Kita, et al., 2008; Kersten & Smith, 2002), objects (Imai et al., 2005), or instruments (Behrend, 1990; Forbes & Farrar, 1993). Children's semantic representation of verbs thus includes components of action events that are irrelevant to their relational meanings (i.e., actors, objects, and instruments). This may be because in ostensive learning situations a novel verb could, in theory, refer to any of the components in the scene (i.e., woman, cutting, paper, scissors). Children must solve this problem of indeterminacy of word meaning (Quine, 1960).

6.1.2 Iconic Gestures Facilitate Verb Learning

Iconic hand gestures (McNeill, 1985, 1992) that depict shapes, motions, and actions, promote learning of individual verbs (Goodrich & Hudson Kam, 2009; Mumford & Kita, 2014). Importantly, they do so by highlighting the component of a complex event that is important for verb meaning (e.g., wiggling the index and middle fingers to depict walking). For example, Mumford and Kita (2014) showed short video clips of action events to three-year-old children, while an experimenter produced novel verbs and iconic gestures that matched either the manner or end-state in the stimulus events. Children interpreted the verbs as manner verbs (e.g., "to push objects in a particular manner") when they saw manner gestures (e.g., pushing) and as change-of-state verbs (e.g., "to cause objects to move into a particular shape") when they saw end-state gestures (e.g., shape tracing). Thus, iconic gestures guided children's verb learning by focusing their attention on components of events important for verb meaning (i.e., manners, end-states).

6.1.3 Abstract Linguistic Knowledge May Facilitate Word Learning

Previous research is limited because it only shows that iconic gesture promotes word-specific knowledge of verb meanings. That is, children learn that a verb refers to the action depicted in iconic gesture. However, it remains unclear whether iconic gestures can also facilitate word-category knowledge about verbs (i.e., verbs typically refer to actions), which children can use in subsequent verb learning. This study investigates this question.

Any theory of word learning must address not only how children acquire word-specific knowledge for each word, but also word-category knowledge. Many studies on vocabulary development have focused on the acquisition of individual words (e.g., Behrend, 1990; Forbes & Farrar, 1993; Goodrich & Hudson Kam, 2009; Imai et al., 2005; Imai, Kita, et al., 2008; Kersten & Smith, 2002; Maguire et al., 2008; McGregor et al., 2009; Mumford & Kita, 2014). However, the vocabulary is more than a list of words. Each word belongs to a linguistic category (e.g., verb, noun, preposition), which has specific properties (e.g., verbs typically refer to actions). Such abstract linguistic knowledge of word categories should help children learn novel words rapidly.

Iconic Gestures Promote Abstract Nonlinguistic Knowledge

Seeing iconic gestures promotes *nonlinguistic* abstract knowledge in children. In a study by Cook et al. (2013), second-, third-, and fourth-grade children were instructed in mathematical equivalence (e.g., $8 + 6 = \ldots +2$) with either speech and gesture or speech alone. In both conditions, the instructor verbalized an equalizer strategy, stating the two sides of the equation must be equal. Only in the speech and gesture condition, the instructor swept the left hand back and forth under the numbers to the left of the equal sign while saying "one side", and the right hand under the numbers to the right of the equal sign while saying "the other side", expressing equivalence of the two sides. Children in the speech and gesture condition solved more similar mathematical equations than children in the speech alone condition in an immediate and delayed posttest, but also in a delayed transfer test of novel equations with multiplication. Thus, gestures that contributed to the expression of abstract knowledge about equations (i.e., the two sides of the equation must be equal) promoted *nonlinguistic* abstract knowledge that children can use later. However, it remains unclear whether seeing iconic gestures can promote *linguistic* abstract knowledge (i.e., word-category knowledge) that children can use later, and whether this is possible even when iconic gestures (and speech) do not express abstract knowledge, but word-specific knowledge.

6.2 The Current Research

The current study investigates if seeing iconic gestures facilitates word-specific knowledge of verbs and word-category knowledge about verbs in preschool-aged children. First, we replicate the effect of iconic gestures on acquiring individual verbs (Goodrich & Hudson Kam, 2009; Mumford & Kita, 2014), which would support the idea that iconic gesture promotes word-specific knowledge of individual verb meanings (Experiments 1 and 2). Second, we investigate whether iconic gestures promote word-category knowledge about verbs (i.e., verbs typically refer to actions) (Experiments 1 and 2). Third, we determine whether iconic gestures have a lasting effect, by testing whether they promote verb learning after a one-week delay (Experiment 9).

6.3 Experiment 8

We taught children novel verbs for unusual actions and tested their ability to generalize these verbs to novel actors across two blocks. In block 1, one group saw iconic gestures depicting the referent actions and the other group saw interactive gestures (Bavelas et al., 1992) that did not depict any aspect of the events. In block 2, all children saw interactive gestures. We predicted that the iconic-gesture group would outperform the interactive-gesture group in block 1 and, crucially, also in block 2.

6.3.1 Method

Design

The experiment had a 2 x 2 mixed design. Trials were grouped into a within-subjects variable block, which corresponded to either trial 1-6 (block 1) or trial 7-12 (block 2) of the experiment task. The between-subjects variable was the type of gestures presented in block 1: iconic gesture vs. interactive gesture. In block 2, children in both groups were presented with interactive gestures. The dependent variable was the children's verb generalization performance in two-way forced choice trials (binary: 1=correct, 0=incorrect).

Participants

The data were collected between the 25^{th} of January and the 3^{rd} of March 2017. Our sample size was determined a priori using G∗Power version 3.1 (odds ratio= 2.30, α = 0.05, power= 0.80) (Faul et al., 2009). Participants were recruited via two public nurseries in the West-Midlands (United Kingdom), and via a database of families who expressed interest taking part in language development research. The sample included 48 typically developing children (24 girls, 24 boys) between 35–48 months old (M = 41.26, SD = 4.01). One additional child was tested, but excluded from the analysis because she pointed exclusively to answers on the right-hand side of the screen. There were 12 boys and 12 girls in each condition. Children's age in months did not differ between the two groups, $F(1, 46) = 0.19, p = .666$. Twenty-five percent of the children had a racial background other than White (i.e., 8% Asian and 17% Black). Informed consent was obtained for all participants. All children were exposed to the English language for at least 75% of the time and English was their primary language (as indicated by their caregivers). Participating nurseries received a book voucher and children who were tested in the research lab received a certificate and a toy. All studies in this paper were approved by the Humanities and Social Sciences Ethics Committee of the University of Warwick.

Materials

A set of 36 video clips (4–15 seconds) depicting 24 unusual actions was taken from the GestuRe and ACtion Exemplar (GRACE) database (Aussems et al., in press, 2017). Stimulus videos showed 24 actors (12 males, 12 females) moving across the length of a scene in an unusual manner using their feet, legs, and body. Their arms and hands were always kept to the side of the body, fingers pointing down, parallel to their torso.

The 24 unusual actions were organized in pairs, which were used in two-way forced choice generalization trials in the following way. Each action had a corresponding an iconic gesture. For a given action pair, the gesture was congruent with one action and incongruent with the other action. The match between actions and iconic gestures was normed for each male actor and female actor in a rating study, reported elsewhere (for more detail see Aussems et al., in press, 2017). The male actors and female actors with the best match ratings were paired up and assigned to an action pair. For each action pair, the actor with the best match and mismatch ratings between actions and congruent and incongruent iconic gestures performed both the target action and the distractor action (24 video clips), and the other actor performed only the target action (12 video clips). There were as many female actors as male actors performing the target and distractor actions. In the test phase, the correct choice in the two-way forced choice trials was a video of a different actor (of the opposite sex) performing the same action as seen in the training phase. We introduced an actor of the opposite sex to make it clear that the actors changed.

The experimenter produced three different interactive gestures (Bavelas et al., 1992) during the experiment task, which expressed excitement and surprise. An example of an interactive gesture can be seen in Figure 6.1, but we have also made video recordings of the interactive gestures available online: `https://osf.io/3jx4b/?view_only=561e0282eef842d39e71b136b553349e`, and still frames of these videos can also be seen in Appendix D. Importantly, none of the interactive gestures depicted any aspects of the action events. The experimenter rotated the three interactive gestures across trials in the training phase, and did so in the same way for children in each group.

The following novel verbs were used to label the unusual actions: *daxing, blicking, larping, stumming, pilking, krading, poffing, wepping, howning, mipping, glabbing*, and *yoofing*. These words are widely used in word learning studies with English-speaking children (e.g., Maguire et al., 2008; Mumford & Kita, 2014; Roseberry et al., 2009).

Video clips of action events were shown using slide presentation software Microsoft Office PowerPoint 2016 on a 14" touch screen laptop. Using on-screen buttons and Visual Basic for Applications (VBA), the data were automatically saved.

Randomization and Counterbalancing

We created 12 versions of the experiment in which we rotated the order of the 12 target actions. Target videos appeared equally on the left and right sides of the screen in the test phase. We administered each experiment version to one child in each condition before moving on to the next version. Participants tested at different nurseries and the research lab are therefore equally represented across the two conditions. All experiment versions were completed by two children in each condition. Participants were pseudo-randomly assigned to a condition before the experimenter met them, based on their gender and age in months.

Procedure

Children were tested individually in a quiet area of their nursery or in the research lab. The experimenter and child participant sat next to each other at a low table, and the experimenter placed the laptop in front of the child. Children completed two warm-up trials followed by two blocks of six verb learning trials. The warm-up trials familiarized children with selecting answers on each side of the screen in the following way. Children were shown pictures of a cat and a dog on either side of the screen and asked to show the experimenter each animal (e.g., "Where is the dog?"). Verb learning trials followed immediately after and included a training phase and test phase (see Figure 6.1). During the training phase, children watched a video of an actor who performed an unusual movement. When the video played a first time the experimenter said "Look! He (or she) is *daxing*!", and when the video played a second time the experimenter said "Wow! He (or she) is *daxing* again!". Depending on the condition, the experimenter accompanied these utterances with either an iconic gesture representing the way the actor in the video moved or an interactive gesture that was not related to the video in any way. In the immediate test phase that followed, the experimenter asked the children to generalize the newly learned verb to one of two videos that played side by side on the screen ("Which one is *daxing*?"). The experimenter looked at the child and not at the screen when making this request. One video showed the same actor as in the training phase performing a distractor movement and the second video showed a novel actor (of the opposite gender) performing the target movement that was labeled with a verb in the training phase. The videos started playing simultaneously and played continuously on loop until the child picked one. If the child did not respond or asked the experimenter whether one of the videos showed *daxing*, the question was repeated until one video was chosen. If the child pointed at both videos, the experimenter asked the child to pick one. This procedure was repeated for six verbs in block 1, followed immediately by six verbs in block 2. Note that all children were taught verbs while seeing interactive gestures in block 2.

Data Analysis

Verb generalization performance (binary: 1=correct, 0=incorrect) was analyzed with a mixed-effects logistic regression analysis using the *lme4* package (D. Bates et al., 2013) in the R software for statistical analyses (R Development Core Team, 2011). Fixed factors included gesture type (iconic vs. interactive) and block (1=trial 1-6 vs. 2=trial 7-12). The model included a maximal random-effects structure (cf.

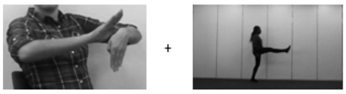
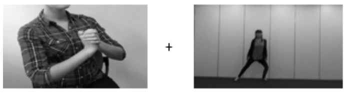
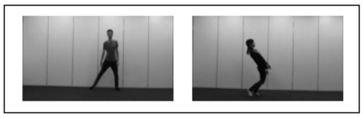
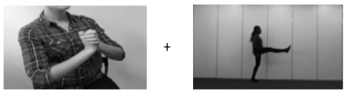
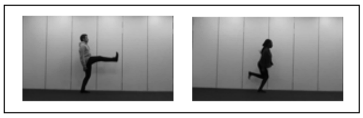
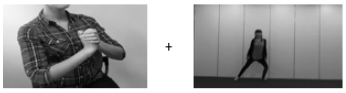
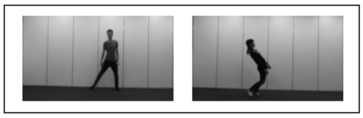
Block	Gesture type	Training phase	Test phase
1.	Iconic		
2.	Interactive		
1.	Interactive		
2.	Interactive		

Figure 6.1: The procedure of Experiments 8 and 9. Children were shown action events in the Training phase accompanied by either iconic gestures in Block 1 and interactive gestures in Block 2 or interactive gestures in both blocks (Columns 1-3). In the Test phase, children were required to generalize the heard verb to one of two videos, which were shown side by side (Column 4). Note that all children completed 12 trials, but that the overview presents examples of only one trial in each condition. In Experiment 8, Block 2 was administered immediately after Block 1, but in Experiment 9, Block 2 was administered one week after Block 1.

Barr et al., 2013), i.e., random slope and intercept variation and the co-variance between the two, for participants and items (i.e., the stimulus videos that were labeled with a novel verb). Likelihood ratio tests (χ^2) were used to compare the full model with updated versions of the model that systematically excluded the main effect and interaction terms of interest. Planned comparisons were carried out by running our analysis separately with children's performances in block 1 and block 2 as the dependent variables. The *confint()* function was used to compute 95% confidence intervals around the beta estimates of each effect. To test whether children's performance was above chance in a given condition, one-sample t-tests in which equal variance was assumed were calculated with the *t.test()* function. The raw data and R Markdown files for all plots and analyses are available from the Open Science Framework at `https://osf.io/3jx4b/?view_only=561e0282eef842d39e71b136b553349e`.

6.3.2 Results and Discussion

Figure 6.2 shows children's performance on the verb generalization trials (in proportion) by gesture type and block. Children's verb generalization performance

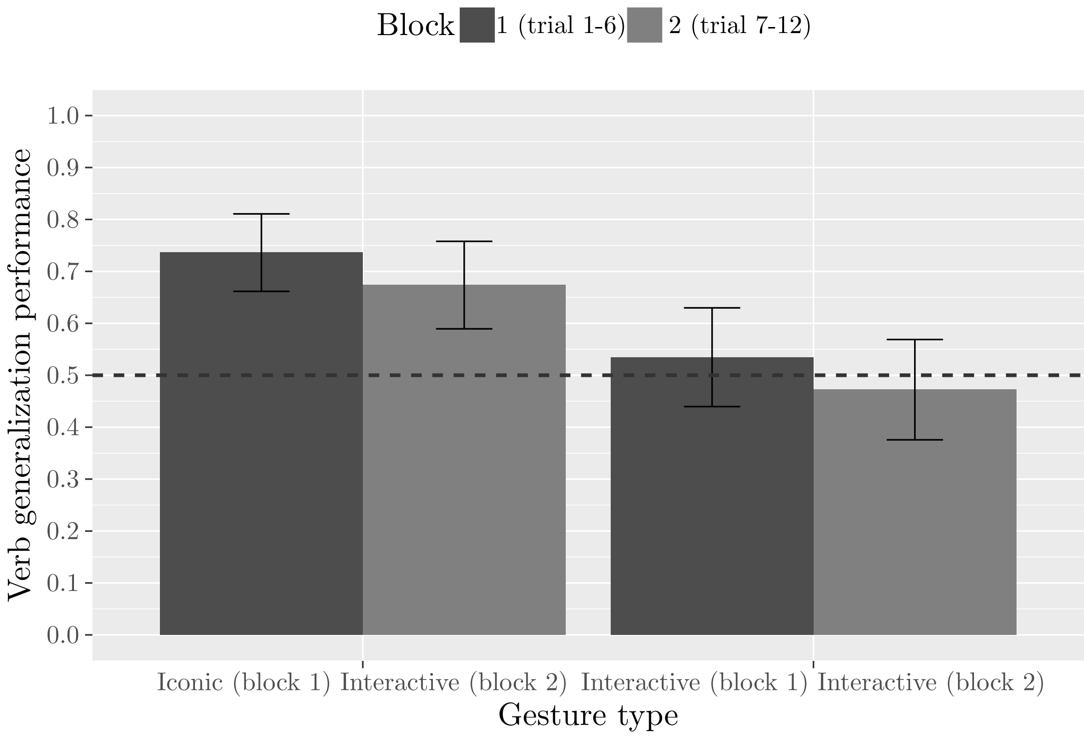


Figure 6.2: Verb generalization performance (y-axis shows proportion of correct responses) organized by gesture type (x-axis) and block (shades of grey). Error bars represent 95% confidence intervals of the means. The dotted line represents performance at chance.

(1=correct, 0=incorrect) was entered into a logistic regression analysis with gesture type and block as predictors. The main effect of gesture type on verb generalization performance was significant, $\chi^2(2) = 14.83, p < .001$, but not the main effect of block, $\chi^2(2) = 2.13, p = .344$, or the interaction, $\chi^2(1) = 0.01, p = .922$. Overall, children who saw iconic gestures in block 1 and interactive gestures in block 2 when the verbs were taught generalized more verbs to novel events that showed the referent actions than children who saw interactive gestures in both blocks ($\beta = -0.91, p < .001$, 95% CI $[-1.37, -0.47]$).

To investigate our predictions, we compared performances between the two groups in block 1 and 2. First, children who saw iconic gestures when the verbs were taught in block 1 generalized significantly more of those verbs to novel events that showed the referent actions than children who saw interactive gestures, ($\beta = -0.90, p < .001$, 95% CI $[-1.46, -0.39]$). Second, this difference in verb generalization performance between groups was also significant in block 2, when all children saw interactive gestures when different novel verbs were taught ($\beta = -0.87, p = .001$, 95% CI $[-1.42, -0.35]$).

Finally, we investigated whether each group performed better than chance. Children's proportions of correct answers on the verb generalization task were entered into one-sample t-tests (two-tailed) with a test value of 0.5. Children who saw iconic gestures in block 1 and interactive gestures in block 2 performed significantly better than chance in block 1, $t(23) = 6.55, p < .001$, 95% CI $[0.66, 0.81]$, and block 2, $t(23) = 4.26, p < .001$, 95% CI $[0.59, 0.76]$, but children who saw interactive gestures in both blocks did not perform significantly different from chance in block 1, $t(23) = 0.76, p = .458$, 95% CI $[0.44, 0.63]$, and block 2, $t(23) = -0.59, p = .558$, 95% CI $[0.38, 0.57]$.

The iconic gesture advantage in block 1 indicates that iconic gestures promote word-specific knowledge. This advantage was still visible in block 2, which demonstrates that iconic gestures promote word-category knowledge, because the verbs and referent actions in block 2 were completely novel and none of the gestures provided any information about the referent actions. Thus, children must have deduced from iconic gestures in block 1 that verbs typically refer to actions, and used this knowledge in block 2.

6.4 Experiment 9

Experiment 9 investigates whether the effect found in Experiment 8 can be observed when block 2 is administered one week after block 1.

6.4.1 Method

Design

The design was the same as in Experiment 8, but we introduced a one-week delay between the two blocks of the experiment task.

Participants

The data were collected between the 12^{th} of April and the 16^{th} of May 2017. Consistent with Experiment 8, the final sample included 48 typically developing children (26 girls, 22 boys) between 35–50 months old (M=42.98, SD=3.87). An additional two children were tested, but excluded from the analysis because they were unavailable for the second testing moment or diagnosed with a speech and language disorder. None of the children had participated in Experiment 8, and the nurseries differed between experiments. Participants were recruited via two public nurseries and one Early Years Teaching Center in the West-Midlands and Warwickshire (United Kingdom). A total of 45 children was tested seven days apart, two children eight days apart, and one child nine days apart. Six percent of the children had a racial background other than White (i.e., 4% Asian and 2% Black). There were 13 girls and 11 boys in each condition. Children's age in months did not differ between the two groups, $F(1, 46) = 0.38, p = .539$. Informed consent was obtained for all participants. All children were exposed to the English language for at least 70% of the time (as indicated by their caregivers). Participating nurseries received a book voucher.

Materials

The materials were the same as in Experiment 8.

Randomization and Counterbalancing

Randomization and counterbalancing were the same as in Experiment 8.

Procedure

The procedure was the same as in Experiment 8 (see Figure 6.1), apart from the following. We introduced a one-week delay between the two blocks of the experiment task. Children performed the warm-up trials again when they were presented with block 2 after the delay to familiarize them with pointing at each side of the screen.

Data Analysis

The data were analyzed in the same way as in Experiment 8.

6.4.2 Results and Discussion

Figure 6.3 shows children's performance on the verb generalization trials (in proportion) by gesture type and block. Children's verb generalization performance (1=correct, 0=incorrect) was entered into a logistic regression analysis with gesture type and block as predictors. The main effect of gesture type on verb generalization performance was significant, $\chi^2(2) = 11.18, p = .004$, but not the main effect of block, $\chi^2(2) = 1.24, p = .539$, or the interaction, $\chi^2(1) = 1.10, p = .294$.

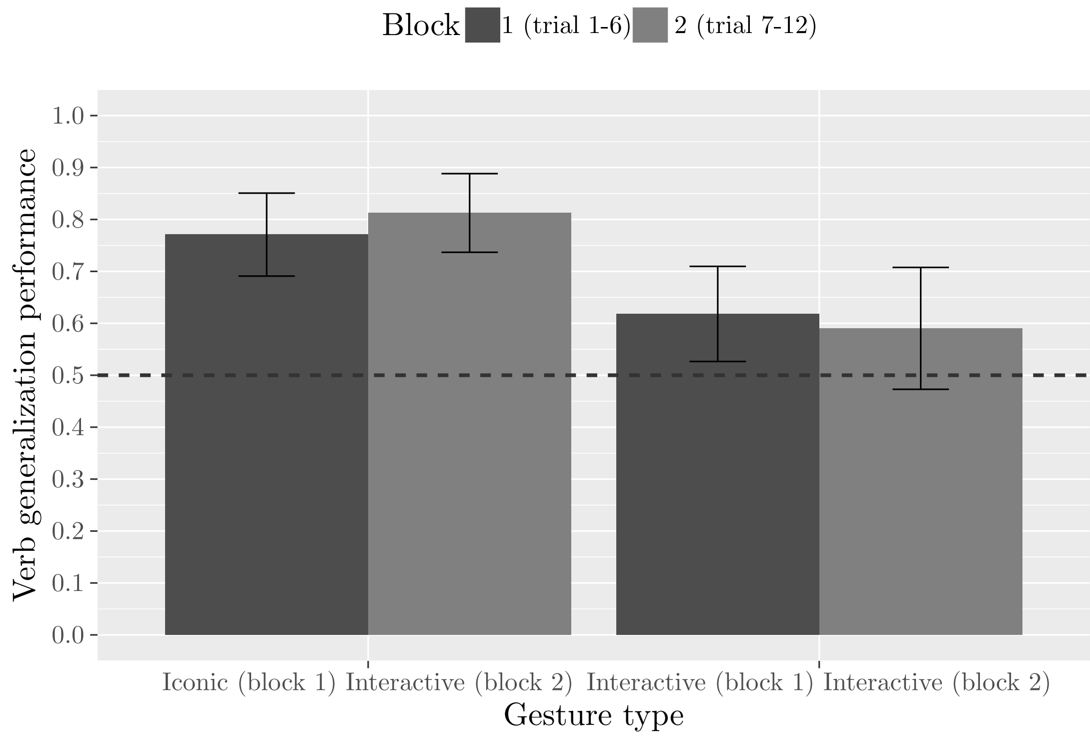


Figure 6.3: Verb generalization performance (y-axis shows proportion of correct responses) organized by gesture type (x-axis) and block (shades of grey). Error bars represent 95% confidence intervals of the means. The dotted line represents performance at chance.

Overall, children who saw iconic gestures in block 1 and interactive gestures in block 2 when the verbs were taught generalized more verbs to novel events that showed the referent actions than children who saw interactive gestures in both blocks ($\beta = -0.91, p < .001$, 95% CI $[-1.58, -0.40]$).

To investigate our predictions, we compared performances between the two groups in block 1 and 2. First, children who saw iconic gestures when the verbs were taught in block 1 generalized significantly more of those verbs to novel events showing the referent actions than children who saw interactive gestures ($\beta = -0.81, p = .009$, 95% CI $[-1.47, -0.21]$). Second, this difference in verb generalization performance between groups was also significant in block 2, when all children saw interactive gestures when different novel verbs were taught ($\beta = -0.87, p = .001$, 95% CI $[-2.08, -0.49]$).

We also investigated whether each group performed better than chance. Children who saw iconic gestures in block 1 and interactive gestures in block 2 performed significantly better than chance in block 1, $t(23) = 7.01, p < .001$, 95% CI [0.69, 0.85], and block 2, $t(23) = 8.54, p < .001$, 95% CI [0.74, 0.89], and children who saw interactive gestures in both blocks performed significantly above from chance in block 1, $t(23) = 2.67, p = .014$, 95% CI [0.53, 0.71], but not in block 2, $t(23) = 1.59, p = .125$, 95% CI [0.47, 0.71].

The results show that iconic gestures promote word-specific knowledge of verbs (block 1) and word-category knowledge about verbs that children could use even after a one-week delay (block 2).

6.5 General Discussion

This study has three key findings. First, preschool-aged children who were taught novel verbs while seeing iconic gestures that depicted the referent actions extended those verbs to novel events more often than children who saw interactive gestures. Thus, seeing iconic gestures promotes *word-specific* knowledge of verbs. Second, children who saw iconic gestures during verb learning outperformed children who saw interactive gestures, in a subsequent block in which all children saw interactive gestures and learned a new set of verbs. Thus, children deduced *word-category* knowledge about verbs from seeing iconic gestures that encoded word-specific knowledge. In other words, iconic gestures helped children to *create new abstract knowledge*. Third, children used this word-category knowledge even after one week, demonstrating a *lasting* advantage.

6.5.1 Iconic Gestures Promote Word-Specific Knowledge of Verbs

Our finding that iconic gestures help children gain word-specific knowledge is consistent with studies showing that iconicity scaffolds learning individual verbs (Goodrich & Hudson Kam, 2009; Mumford & Kita, 2014) and prepositions (McGregor et al., 2009). For instance, the study by Mumford and Kita (2014) showed that iconic gestures guided children's attention to components of action events (i.e., manners, end-states), which influenced children's interpretations of novel verb meanings. We extended their findings on transitive verbs to intransitive verbs in the current study.

Furthermore, this finding is consistent with research on sound symbolism (i.e., the non-arbitrary relationship between sound and meaning) showing that iconicity in spoken language helps children to map verbs to their referents. For instance, in a study by Imai, Kita, et al. (2008), 3-year-old Japanese-speaking children could extend the meaning of novel sound-symbolic verbs (e.g., *batobato* for slow heavy steps and *chokachoka* for quick small steps), but not non-sound-symbolic verbs, to events in which novel actors performed the target movements. Because sound-symbolic verbs sound like what they refer to, they focused children's attention on actions. The similarity between iconic gestures and actions fulfills the same function. However, future studies should investigate whether sound symbolism can also facilitate word-category knowledge about verbs. Reduplicated words (e.g., *chokachoka*) that sound-symbolically represent repeated manners of motion (e.g., fast walking with small steps), which have been shown to benefit word learning (e.g., Imai, Kita, et al., 2008; Kantartzis et al., 2011; Ota & Skarabela, 2016; Yoshida, 2012), could be useful for promoting children's abstract linguistic knowledge about manner verbs, because these verbs often refer to repeated actions (e.g., eating, walking, canoeing). A future study could experimentally investigate whether reduplication in the word form leads children to consistently map manner verbs to action events that show repetition.

6.5.2 Iconic Gestures Promote Word-Category Knowledge about Verbs

This study goes beyond the existing literature in three important ways. First, seeing gesture promotes abstract *linguistic* knowledge (i.e., verbs typically refer to actions) in children, just as seeing iconic gestures promotes abstract *nonlinguistic* knowledge about mathematical equivalence (Cook et al., 2013). Second, in Cook et al.'s study, gestures in the context of speech directly encoded the abstract knowledge that the two sides of equations must be equal, but, in our study, iconic gestures encoded word-specific knowledge, not word-category knowledge, and speech did not express word-category knowledge either. Nevertheless, the iconic gestures in our study helped children deduce abstract knowledge about the word category verbs. Thus, iconic gestures helped children generate new abstract knowledge. This is important because general knowledge of what verbs are is essential for vocabulary development. Third, seeing iconic gestures has a lasting effect on promoting *linguistic* abstract knowledge (i.e., word-category knowledge), even after one week. Thus, the lasting benefit is not confined to *nonlinguistic* abstract knowledge (Cook et al., 2013).

6.5.3 Possible Mechanism

As seeing iconic gestures facilitates both linguistic and nonlinguistic abstract knowledge, learning via gesture observation may rely on a domain-general cognitive process. We suggest that three properties of iconic gestures play a key role in this process. First, iconic gestures represent information in a schematic manner, which leads to efficient communication (Kita et al., 2017) and to generalizable knowledge (Goldin-Meadow, 2015). In teaching contexts, iconic gestures can convey only the most essential information children need for the task. For mathematical equivalence, for instance, gestures represent each side of an equation without focusing on specific values or the plus sign (Cook et al., 2013), and for verb learning in the current study, they strip the action component of a scene from irrelevant details (e.g., the actor's clothes, background). Second, iconic gestures are representational actions in the sense that they stand for something else (Kita et al., 2017; Novack & Goldin-Meadow, 2016). Iconic gesture as a symbol is "removed" from the referent, and this symbolic distancing (Werner & Kaplan, 1963) may promote abstract understanding. Third, iconic gestures are typically produced in ways semantically coordinated with speech (Kita & Özyürek, 2003; McNeill, 1985, 1992). Thus, when children receive information in speech-gesture combinations, they may naturally tend to ground information in speech to spatio-motoric information in gesture (Kelly, Özyürek, & Maris, 2010; Novack & Goldin-Meadow, 2016; Valenzeno, Alibali, & Klatzky, 2003). This may be especially helpful when speech directly encodes abstract knowledge (e.g., Cook et al., 2013; Valenzeno et al., 2003).

6.5.4 Differences between Experiments 8 & 9

Two reasons may explain why children in the interactive gesture condition performed significantly above chance in Experiment 9, but not in Experiment 8. First, Experiment 9 was partly conducted in an Early Years Teaching Center, a special type of nursery with a strong academic emphasis. Second, the education levels of parents were higher in Experiment 9 than in Experiment 8. According to 2011 census, the percentage of the residents (around the nurseries) with a higher education degree or similar was 41.7% for Experiment 9, but 25.1% for Experiment 8. Thus, the children in Experiment 9 likely came from families with a higher socioeconomic status (SES), which is often associated with better vocabulary skills (Hoff, 2006).

6.5.5 Future Research

Future studies could investigate whether iconic gestures also facilitate word-category knowledge of words from other categories. Iconic gestures may be useful for explaining spatial relationships as the key semantic property of prepositions (e.g., "above", "below", "off"). Children struggle to learn such prepositions (Washington & Naremore, 1978) and do not fully acquire them until 4-6 years of age (Owens, 2008). Iconic gestures could depict the relation between two or more entities (e.g., "the basement is below the house"), and thus promote the word-category knowledge that prepositions tend to refer to spatial relationships.

Another interesting direction would be to further investigate the effects of seeing iconic gesture on nonlinguistic abstract knowledge, such as children's understanding of balance (Pine, Lufkin, & Messer, 2004) or Piagetian conservation problems (Church & Goldin-Meadow, 1986). Iconic gestures could depict a balance between two different elements or dimensions such as width and height in explaining conservation problems.

6.5.6 Conclusion

The real-world benefit of seeing iconic gestures on verb learning is not always clear, because people do not produce an iconic gesture with every utterance. However, this study shows that iconic gestures do not merely facilitate word-specific knowledge of individual verb meanings, but also helped children generate word-category knowledge about verbs (i.e., verbs typically refer to actions) that children can use in subsequent verb learning, even in the absence of iconic gestures. Thus, seeing iconic gestures has a more far-reaching impact on verb learning than was previously thought. Iconic gestures help children to figure out *how* to learn verbs.

Chapter 7

Discussion and Conclusions

When children encounter a novel action label, they have to determine to which action in the world this word refers. Typically, a novel action label occurs in the context of rich action events that are complex to encode for children, which creates uncertainty about the correct mapping (Quine, 1960). Despite this referential ambiguity, even children as young as one year old break into word learning (Smith & Yu, 2008; Vouloumanos & Werker, 2009). This dissertation focuses on the multimodal nature of input that children receive from adults when they encounter complex action events and novel action labels. Specifically, the studies presented in this dissertation investigate the scaffolding effects of seeing iconic gestures on 3-year-old children's action event memory and verb learning.

In this final chapter, the main findings are summarized and integrated in answers to the four research questions in Section 7.1. Furthermore, the theoretical and practical implications of these findings are discussed in Section 7.2. We highlight some of the research output in Section 7.3 and provide a methodological note on the statistical analyses used in this dissertation in Section 7.4. Suggestions for future research are presented in Section 7.5. This chapter closes with a summary of the main conclusions in Section 7.6.

7.1 Summary and Answers to the Research Questions

RQ1 What requirements should stimuli meet to investigate action event memory and verb learning with the help of iconic gestures?

We developed and normed the GestuRe and ACtion Exemplar (GRACE) stimuli database for which we identified four requirements that were described in Chapter 3. These requirements are briefly summarized here. First, clear matches and mismatches between iconic gestures and actions were important to create two contrasting choices in forced choice tasks. Such two-way forced choice tasks are commonly used for testing child participants in studies on language and gesture (e.g., Goodrich & Hudson Kam, 2009; Mumford & Kita, 2014). To this aim, we paired up our actions and designed a matching iconic gesture for each action. Mismatches were created by showing the iconic gesture for one action with the other action in the pair. Match and mismatch ratings from a first norming experiment demonstrated that actions and matching iconic gestures were rated as clear matches and actions and mismatching iconic gestures were rated as clear mismatches. Female actors and male actors with the best average rating scores for action pairs and matching iconic gestures were assigned to those action pairs to create an optimal stimulus set.

Second, it was crucial for the two-way forced choice tasks that female actors and male actors who were assigned to an action pair performed the same actions in a similar manner, and different actions in a dissimilar manner. Especially in our memory task, in which we tested children's actor memory, it was important that everything about the distractor events was the same in the two-way forced choice task, apart from the actors. Also in our verb learning tasks, generalizing an action label from one actor's movement to another actor's movement is too difficult for children if the movements are not performed in a highly similar manner. Moreover, the distractor movements and the target movements were required to be dissimilar to make it clear that the actors were performing different actions in the test. Rating scores from a second norming experiment showed that male actors and female actors, whose action videos matched the iconic gestures in the best possible way, performed the same actions in very similar manners and different actions in highly distinct manners.

Third, it was vital that the actions in the GRACE database were distinct from each other to avoid confusion. After all, it would be hard to test children's action recognition memory if they would confuse the actions with each other during a memory task. Similarly, for verb learning tasks, children's performance might be biased if they confuse the actions with one another. This could happen, for instance, if children think that they have seen an action before in the context of another heard label. According to the mutual exclusivity principle (Markman & Wachtel, 1988), it may be hard for children to accept a novel label for an already labeled action. The similarity ratings from a third norming experiment demonstrated that most of the actions were highly distinct.

Fourth, it was important that the actions events in the database were unusual. This is crucial for memory tasks and verb learning tasks, because it would be hard to test children's action recognition memory and verb learning performance if children see the stimuli actions daily and perhaps already know a label for them. The results from a fourth norming experiment showed that adult native English speakers did not converge on single-word or multi-word descriptions for the actions, indicating that the actions are unusual.

It must be noted here that based on the findings from the four norming tasks, we dropped action pair 7 from the stimulus set that was used in the memory and verb learning experiments with child participants. On average, this pair received the lowest score for the mismatch between iconic gestures and actions in the first norming experiment (see Figure 3.3). It also received relatively high similarity scores for two actors performing the two different actions in the pair (see Figure 3.4). In the third norming experiment, both actions appeared similar to other actions in the set (see Table 3.2). The similarity between action 7b and action 13a was rated, on average, with 6.63 out of 7 (with 7 being "very similar"). Finally, the fourth norming experiment revealed that adults often used the word "jump" to describe action 7a (see Table 3.3), which is a word that 3-year-old children may already know. Although most of the norms for action pair 7 were still acceptable, accumulatively, they led to the decision to exclude action pair 7 from the stimulus set. Overall, the four requirements were met for the remaining stimuli.

RQ2 Do children encode and remember action events differently when their attention is guided by iconic gestures?

We discussed an event recognition memory task in Chapter 4, which was designed to address RQ2. In training phase of this task, 3-year-old children watched videos of actors moving in unusual manners with an experimenter, who either produced iconic gestures that depicted how the actors moved, interactive gestures that did not depict any aspect of the videos, or no gesture at all. After a short break, children's memory of real-life actors and their actions was tested in a two-way forced-choice task in the following way. For action memory, the videos that children had seen before were paired up with videos of the same actors performing different actions. For actor memory, the videos that children had seen before were paired up with videos of novel actors performing the same action. Children who encoded the action events while seeing iconic gestures remembered more of those events in the event recognition memory task than children in the control groups. Children thus encoded and remembered the action events differently when their attention was guided by iconic gestures.

We also analyzed iconic gestures at the semiotic level to investigate *how* seeing iconic gestures influence children's action event memory. There were three categories of iconic gestures that differed in how much of the actors was represented by the experimenter's hands. The experimenter's hands mapped either directly to the feet, legs, or body of the actors. All gestures boosted children's action recognition memory. Importantly, this finding shows that children encode and remember those aspects of action events that are schematically depicted in iconic gestures. However, only hand-as-body gestures boosted children's actor recognition memory. This is probably the case because hand-as-body gestures draw children's attention to the actors' whole bodies, which gives children more opportunity to encode person-specific information than hand-as-feet gestures or hand-as-leg gestures. Specifically, iconicity in iconic gestures draws children's attention to part of a complex event where information relevant for the task can be seen. Thus, iconic gestures influenced children's action event memory in two different ways. First, they directly encoded task-relevant information, namely, distinctive features of actions, which led to better action recognition memory. Second, iconic gestures highlighted particular parts of the event (e.g., the actor's body) that included task-relevant information, and guided children's attention to these parts. This, in turn, helped children to find and encode actor-specific features, leading to better actor recognition memory.

RQ3 Do children use prior unlabeled exposures to actions for verb learning and does seeing iconic gestures influence this process?

We reported a novel verb learning task in Chapter 5, which was designed to answer RQ3. In this task, we manipulated pre-exposure to actions and the gesture type children saw with these pre-exposures to see how this influenced their verb learning. In the initial phase of this task, children observed videos of action events while an experimenter produced either iconic gestures depicting the actions in the events or interactive gestures that did not depict any aspect of the events. Note that no label was introduced in this part of the task. After a short break, children were presented with novel action exemplars (i.e., showing novel actors) which either showed the actions they had seen in the initial phase or novel actions. The experimenter now labeled each action with a novel verb. Immediately after children were taught a verb label, a generalization trial followed in which they were required to extend the newly learned verb to one of two videos: one showed the referent action performed by a novel actor, and the other showed a novel action performed by the actor children saw when the verb was taught. Only children who were pre-exposed to unlabeled exemplars of the referent actions ("retrospective exemplars") while the referent actions were highlighted by iconic gestures, generalized the novel verbs successfully to novel events (i.e., different actor) showing the referent actions.

In a follow-up experiment, we asked whether seeing retrospective exemplars with iconic gestures helps children to form action concepts with schematic information or whether iconic gestures simply give children extra opportunity to form action concepts (i.e., iconic gestures as extra retrospective exemplars). Two groups of children (different from the previous experiment) was taught novel verbs using the same novel verb learning task as in the previous experiment. This time, we manipulated the type of pre-exposure to actions. One group was taught verbs while pre-exposed to one retrospective exemplar with an iconic gesture, and the other group while pre-exposed to two different retrospective exemplars (i.e., two different actors performing the same action) side by side without gesture. Children learned novel verbs more successfully when pre-exposed to retrospective exemplars with iconic gestures than when pre-exposed to two different retrospective exemplars of the same action (i.e., different actors) presented side by side without gesture. However, children also performed reliably above chance in the verb generalization trials when pre-exposed to two different exemplars of the same action. Taken together, both experiments show that children can use prior unlabeled exposures to actions for verb learning, when iconic gestures helped them to individuate actions in the unlabeled action exemplars, and when two unlabeled exemplars of the same action were presented side by side. However, the follow-up experiment shows that iconic gestures work better than side by side presentation, suggesting that children formed action concepts based on schematic information in the iconic-gesture group.

RQ4 Do children gain word-specific knowledge of verbs and word-category knowledge about verbs from seeing iconic gestures during verb learning?

We discussed findings from a novel verb learning task in Chapter 6, which was designed to investigate RQ4. Preschool-aged children watched videos of action events while an experimenter labeled these events with novel verbs. The verbs were taught across two sets. We manipulated the gesture type that children saw with the first set of novel verbs (iconic gesture vs. interactive gesture). Children's knowledge of the novel verb meanings was tested in generalization trials in which children were required to extend each newly learned verb to one of two videos: one showed the referent action performed by a novel actor and the other showed a novel action performed by the actor children saw when the verb was taught. Children who saw iconic gestures depicting the referent actions outperformed children who saw interactive gestures in the verb generalization trials of the first set. This finding shows that seeing iconic gestures promotes word-specific knowledge of verbs. Crucially, the iconic-gesture group also outperformed the interactive-gesture group in subsequent trials, in which all children saw interactive gestures when a second set of different novel verbs was taught. This finding shows that seeing iconic gestures promotes word-category knowledge about verbs.

In a follow-up experiment, we asked whether seeing iconic gestures promotes lasting word-category knowledge about verbs. Two groups of children (different from the previous experiment) were presented with the same novel verb learning task, but this time, the second set of verbs was taught one week (range=7–9 days) after the first set. Children in the iconic-gesture group outperformed the interactive-gesture group even in subsequent trials that were administered after a one-week delay, in which all children saw interactive gestures when a second set of different novel verbs was taught. Taken together, findings from both experiments show that seeing iconic gestures promotes word-specific knowledge of verbs and word-category knowledge about verbs in preschool-aged children, but the follow-up experiment also demonstrates that iconic gestures promote *lasting* abstract linguistic knowledge.

7.2 Theoretical and Practical Implications

This section starts with theoretical implications of the research presented in this dissertation for existing theories of gesture (Section 7.2.1) and verb learning (Section 7.2.2). It ends with practical implications of the research findings for language teaching (Section 7.2.3).

7.2.1 Implications for Gesture Studies

The research findings have three important theoretical implications for gesture studies. First, iconicity in iconic gestures can have a deictic function. In Section 2.2, it was explained that there are different categories of gestures, but that the distinction between these categories is a fuzzy one (Goodwin, 2003). Rather, gestures are best interpreted along the dimensions (e.g., deixis, iconicity, and "temporal highlighting" for beat gestures) of these categories (McNeill, 1992). Findings from Chapter 4 show that iconic hand-as-body gestures drew children's attention to the bodies of the actors, which gave them more opportunity to encode person-specific information than iconic hand-as-feet or hand-as-leg gestures. This led to better actor recognition memory performance. Thus, the iconicity in iconic gestures in this case, drew children's attention to parts of an action event, where information relevant for the task could be seen. It is well-documented that pointing gestures (Langton et al., 1996) and the deictic component of iconic gestures (i.e., location in gesture space, at which iconic gestures are produced) (e.g., Sekine & Kita, 2015), can direct the recipient's attention to particular areas of the interactional space. Chapter 4 demonstrates for the first time that iconicity in iconic gestures can also direct the recipient's attention to a particular part of an event that includes the referent of the gestures.

Second, iconic gestures help children to form action concepts, which they can use for subsequent verb learning. Studies on iconic gestures and verb learning always presented children with iconic gestures and action exemplars in the context of novel verbs (e.g., Goodrich & Hudson Kam, 2009; Mumford & Kita, 2014). This set-up did not allow children to form an action concept before receiving a linguistic action label. The study reported in Chapter 5 is the first to show that iconic gestures help children to form action concepts based on schematic information. This interpretation is based on the finding that children who saw iconic gestures with action exemplars performed better in a novel verb learning task than children who saw two different exemplars of the same action (i.e., two different actors performing the same action), which were presented side by side and without gesture. Importantly, action concepts that were formed based on iconic gestures proved to be more useful for verb learning and generalization than action concepts derived from comparing two different exemplars of the same action. This may be due to the schematic nature of iconic gestures (Kita et al., 2017). This effect is best explained by the scaffolding effect that iconic gestures may have. That is, iconic gestures provide "top-down" guidance in forming action concepts based on schematic information. Children used those action concepts to identify the same actions in subsequently labeled exemplars, which facilitated mapping a label to these actions and extending this label to novel events showing the referent actions. Children who saw two exemplars of the same action (i.e., two actors performing the same action) side by side and without gesture needed to extract invariance of action through a "bottom-up" process, in which they had to consider all components of the two exemplars as possibly invariant. Thus, they may not have extracted the invariance of action and possibly had traces of the actors' appearances lingering in their representation of the two action exemplars.

Third, iconic gestures help children to generate abstract *linguistic* knowledge. Gestures can help children to construct abstract *nonlinguistic* knowledge about how to solve mathematical equivalence problems (Cook et al., 2013), but the verb learning study in Chapter 6 is the first to present evidence that gesture also help children to generate abstract *linguistic* knowledge about verbs. There is yet another, more important difference between the study by Cook et al. (2013) and the study presented in Chapter 6 of this dissertation. The gesture and speech manipulation in the experiment by Cook et al. (2013) encoded *abstract* knowledge for the children, whereas in our verb learning task, gesture and speech encoded *concrete* knowledge about individual verb meanings. Nevertheless, such word-specific knowledge of individual verb meanings helped children to *generate* word-category knowledge about what verbs are and that verbs typically refer to actions.

7.2.2 Implications for Theories of Verb Learning

The research findings have two important theoretical implications for theories of verb learning. First, we present an important unexplored first step in the verb learning process: retrospective exemplars (i.e., exposures to unlabeled actions that children may have observed before a linguistic label is introduced). Studies on noun learning have suggested that pre-exposure to objects influences children's learning of labels for these objects (Clerkin et al., 2017), though it was unclear whether the objects infants were pre-exposed to in the study by Clerkin et al. (2017) were labeled or not. Moreover, children probably saw the same object exemplar multiple times in their study, which is different from a retrospective exemplar (i.e., pre-exposure to a different unlabeled exemplar). To our knowledge, retrospective exemplars have not been explored in the context of verb learning. In those studies that did focus on verb learning with multiple exemplars, children were always presented with labeled exposures to actions (Childers, 2011; Childers & Tomasello, 2002; Haryu et al., 2011; Maguire et al., 2008). As such, these studies investigated how initial *linguistic* representations of verbs influence how children process subsequent *linguistic* representations of those verbs. In Chapter 5 we show that children can use prior unlabeled exposures to actions for verb learning, though this finding was dependent on whether the action component was highlighted, either by iconic gestures that depicted the action components or by presenting two different action exemplars side by side. Nevertheless, this finding suggests that it is also important to consider that *nonlinguistic* representations of actions can influence initial *linguistic* representations of verbs that refer to those actions. Children's prior unlabeled exposures to actions may thus play a more important role in mapping novel verbs to their referents than was previously thought. More research is needed, however, to investigate what sources of (visual) information children may use in forming action concepts and learning verbs.

Second, children can construct word-category knowledge about verbs from learning individual verb meanings. Previous verb learning studies have mainly focused on how children acquire individual verb meanings (Childers & Tomasello, 2002; Forbes & Farrar, 1993, 1995; Kersten & Smith, 2002; Imai et al., 2005; Imai, Li, et al., 2008). However, in Chapter 6, we show that children can also deduce word-category knowledge about verbs during the individual verb learning process. Though this finding was dependent on the iconic gestures that children saw during the experiment, in natural situations children can undoubtedly use other sources of (visual) information to deduce word-category knowledge too. More studies are needed, though, to investigate what other sources of information children use to form word-category knowledge about verbs, and how children develop abstract linguistic knowledge about words from other categories (e.g., prepositions, nouns).

7.2.3 Practical Implications

The research findings presented in this dissertation may have important practical implications too. Seeing iconic gestures that depict actions has a positive effect on many aspects of children's verb learning. For example, seeing iconic gestures helps children to encode complex action events (Chapter 4 & Chapter 5), form action concepts (Chapter 5), learn the meanings of individual verbs (Chapter 5 & Chapter 6), and learn about verbs as a word category (Chapter 6). Iconic gestures may thus offer caregivers, nursery staff members, and teachers a useful tool for teaching children about action events and verbs. Scaffolding children's verb learning before they reach school age is important, because the vocabulary size and skills of preschool-aged children are major predictors of later school success (Anderson & Freebody, 1979; Morgan et al., 2015; Rescorla, 2005, 2009; Rowe et al., 2012).

The current research findings may also have implications for populations that were not investigated in this dissertation. For example, seeing iconic gestures may be particularly useful for facilitating the vocabulary growth of children with Specific Language Impairments (SLI) (e.g., Vogt & Kauschke, 2017), which cannot be accounted for by a general developmental delay. Furthermore, children who are learning more than one language at home or at nursery (school) may also benefit from the semantic information conveyed by iconic gestures in word learning contexts (e.g., Tellier, 2008).

7.3 Digital Resource for Language and Gesture Research

This dissertation presents a large database in Chapter 3, which contains stimuli videos that psychologists, linguists, and gesture researchers can openly access online. Extensive norming of this database led to a high-quality stimulus set for the memory and verb learning tasks presented in this dissertation. By publishing a paper on how the GRACE video database was developed and normed, and by making this digital resource openly available to the research community, we hope that the videos can serve as stimuli in many more experiments. As an openly accessible digital research for language and gesture research, the GRACE video database is an important research output of this dissertation.

7.4 A Methodological Note on Mixed-Effects Models

Throughout this dissertation, we used mixed-effects logistic regression models to analyze the empirical data obtained in experiments with child participants. Mixed-effects models are just one tool among many, but they have two clear advantages compared to other types of analyses and these advantages are highlighted in this section. First, mixed-effects logistic regression models allow one to analyze the data in its binary form, which preserves trial-by-trial information that would be lost in an aggregate. Thus, this analysis has an advantage over mixed-effects ANOVAs, or older types of ANOVAs, which are performed on the aggregate data (i.e., the sum of the correct number of trials or the average proportion of correct responses).

Second, mixed-effects models allow for the specification of a maximal random-effects structure for participants and items, which increases the *generalizability* of the findings beyond participants and items (Janssen, 2012). The random-effects structure for mixed-effects models is still under much debate, and there are several ways to specify random-effects structure (Janssen, 2012), but there is at least some agreement that models with a maximal random-effects structure models decrease the Type I error rate (i.e., false positives) (Barr et al., 2013). However, others argue that while a maximal random-effects structure decreases the Type I error rate, it increases the Type II error rate (i.e., loss of statistical power) (Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017). To avoid this, it is important to justify a maximal random-effects structure by the design of the experiment (Matuschek et al., 2017). In all our experiments, items and participants were crossed, since every participant saw at least one variant of each item. Note that we also had more trials and stimuli videos in each experiment, compared to Imai et al. (2005), who used similar paradigms to test recognition memory and verb learning. Therefore, we included a random slope and intercept for each participant and item, and for the co-variance between the slopes and intercepts. In other words, we account for the fact that some child participants perform very well (or very poor) overall (i.e., the random intercept) and thus they may be less (or more) influenced by the manipulation (i.e., the random slope), but overall performance and influence of the manipulation may differ per individual child (the co-variance between the slopes and intercepts). In the same way, our models take into account that children may respond differently to each stimulus event and may therefore be influenced by the manipulation differently (e.g., if they are struggling with a particular stimulus event then the manipulation may really help them, and vice versa). Thus, the design of our experiments justified a maximal random-effects structure for participants and items, which increased the generalizability of our findings beyond the participants and items used in the experiments, because these maximal models account for the variability within participants and items as a result of sampling from populations of participants and items.

7.5 Suggestions for Future Research

In the empirical chapters of this dissertation, several suggestions were made for future research. Here, we briefly summarize these suggestions. Chapter 3 suggested that the GRACE video database stimuli could be used for experimental research on language and gesture. Chapter 4 proposed to investigate whether iconic gestures can help adults to teach children about fundamental movement skills such as stability (e.g., balancing, twisting), physical fitness (e.g., stretching, bending), locomotor skills (e.g., running, jumping), object manipulation and control (e.g., throwing, catching), and the way the human limbs work. Chapter 5 suggested to examine whether seeing iconic gestures with unlabeled exemplars of word referents leads to categorical knowledge about these referents, and future referents that belong to the same category. Finally, Chapter 6 proposed to test whether iconic gestures could also help children to construct abstract linguistic knowledge about other word categories than verbs (e.g., prepositions), or nonlinguistic abstract knowledge about concepts such as balance and conservation. Additionally, it was suggested that sound symbolism could lead children to deduce word-category knowledge about verbs too, just like iconic gestures do.

Three more ideas for future work are presented in this section, which have arisen from serendipitous observations from a collection of studies. First, a possibly interesting direction for future research is to identify nonverbal and verbal markers that could reflect children's internal states during word learning. Children themselves produced a variety of utterances and ostensive signals (e.g., pointing gestures and eye gaze) across the verb learning experiments of this dissertation when giving a response in the two-way forced choice tasks. Occasionally, when children pointed at their answer in the test phases, they simultaneously produced a speech utterance (e.g., "It's this one', or "What's this one doing?"). Such utterances provide additional information about children's confidence levels and about whether children implicitly realize that they are observing a novel action. Importantly, some pointing gestures could be interpreted as if children used the gesture space in front of them to explore multiple possible answers with their index finger. An example of such a gesture is a pointing gesture with a swaying trajectory (i.e., their point moves from one answer to the other; following a zigzag pattern). Additionally, children who appeared uncertain about their answers seemed to keep their points close to their body and children who appeared confident about their answers seemed to fully extend their arm. Often, a shoulder flick was part of the gestures of children who seemed confident and performed well in the task, but not of the gestures of children who seemed less confident and performed poor in the task. In some cases, children pointed at one answer while clearly looking at the other answer over their pointing arm, which could indicate a mismatch between children's explicit and implicit knowledge. Mismatches between children's implicit and explicit knowledge revealed by their gestures, speech, and gaze orientation have been studied before in Piagetian conservation tasks (e.g., Church & Goldin-Meadow, 1986; Goldin-Meadow et al., 1993; Kelly & Church, 1998), false belief tasks (e.g., Clements, Rustin, & McCallum, 2000) and (mathematical) problem solving tasks (e.g., Alibali & Goldin-Meadow, 1993; Broaders et al., 2007; Heine et al., 2009), but not in word learning tasks. A future study could experimentally test the idea that young children's verbal and nonverbal behaviors reflect their psychological states during language learning, for instance, using one of the verb learning tasks reported in this dissertation. Identifying nonverbal markers of children's internal states during language learning is important to determine if children struggle to learn and generalize new words. If the difficulties that children may experience could be identified through (combinations of) particular utterances, pointing gestures, and gaze behaviors, then this knowledge could be used to design an intervention to help counteract children's struggle and short-circuit its impact on word learning and generalization. Such a study could also contribute to the literature by building on the existing view that gesture serves as a "window into the child's mind", which was discussed in Section 2.8 of Chapter 2.

Second, future studies could combine iconicity in spoken language and gesture in one study to investigate the development of children's level of symbolic understanding. It would be interesting to research a potential (super additive) effect that a combination of iconicity in spoken words and gestures may have on word learning. Section 2.6 in Chapter 2 outlined several studies which showed that both sound symbolism and iconic gestures separately facilitate learning, retention, and generalization of word-referent mappings. However, Section 2.5 showed that sensitivity to sound symbolism generally emerges earlier in development than sensitivity to iconic gestures. It may be the case that sound symbolic words lay the foundation for a lower level of symbolic understanding, which at a higher level, could be used to process and interpret iconic gestures as children get better at interpreting symbols. If this is true, then 2.5-year-old children, who are not yet equipped with the right level of symbolic understanding to map iconic gestures to verb meanings in a matching task (e.g., Mumford, 2014), may benefit from an additional sound-symbolic manipulation. We know that 2-year-old children are already reasonably good at mapping sound-symbolic words to verb meanings in a similar matching task (Imai, Kita, et al., 2008). In a future experiment with a 2 x 2 between-subjects design, sound symbolism (sound-symbolic vs. non-sound-symbolic) and iconic gesture (iconic vs. non-iconic) could be manipulated to investigate if sound symbolism bootstraps 2.5-year-old children's iconic gesture comprehension. If children in the sound-symbolic plus iconic-gesture condition outperform the children in the control groups, then this could mean that the sound-symbolic manipulation boosts children's interpretation of iconic gestures, which leads to a super additive effect of iconicity in the two manipulated modalities on word learning. Such a super additive effect could provide support for the hypothesis that sound symbolism in spoken language is a precursor for children's comprehension of iconic gestures at a later age. An initial experiment could introduce a match rating task, but ultimately, manipulating sound symbolism and iconic gestures would be particularly interesting in the context of verb learning.

Third, future studies could investigate at which stage in development children start to understand iconic gestures that are symbolically "removed from their referents" (Werner & Kaplan, 1963). Let us explain using an example. The "symbolic distance" (Werner & Kaplan, 1963) between the iconic manner gestures and referents in our studies may be larger than in studies in which iconic manner gestures represented a manual action (e.g., Namy, 2008; Mumford & Kita, 2014; Sekine et al., 2015; Stanfield et al., 2014). When iconic gestures depict manual actions, the hands used to produce iconic gestures that depict these manual actions often map directly to the hands used for performing the manual actions. For instance, if the action involved sprinkling sand on a surface (e.g., Mumford & Kita, 2014), then the hand used to produce the iconic gesture represented a sprinkling hand. This is an example in which the symbolic distance between iconic gesture form and its referent is small. The iconic gesture is not symbolically removed from its referent, because the gesture hand represents the hand in the manual action. In the case of iconic gestures presented in the empirical studies of this dissertation, the hands used to produce the iconic gestures always represented different body parts than the hands (i.e., feet, legs, body). Thus, the iconic gesture forms are symbolically removed from their referents; they are symbols that stand for something else. The distance between the iconic gestures and referents in the current example could thus be perceived as larger than in the previous example. Nevertheless, our iconic gestures led to increased event recognition memory performance and verb learning performance in the experiments with child participants, suggesting that, by age 3, children indeed understand iconic gesture forms that are symbolically removed from their referents. However, more research is needed to investigate the developmental stage at which children's level of symbolic understanding also includes those cases in which iconic gestures are symbolically removed from their referents. This could be investigated experimentally, for instance, by manipulating the symbolic distance between iconic gestures and their referents for different age groups (e.g., 14 months, 18 months, 24 months, 30 months) in a gesture comprehension task.

7.6 Conclusions

This dissertation examined if seeing iconic gestures influences 3-year-old children's action event memory and verb learning. The empirical studies with child participants provide evidence that seeing iconic gestures:

- positively influences the way children encode and remember action events (Chapter 4, Experiment 5)
- boosts children's memory of those aspects of action events that they schematically highlight (Chapter 4, Experiment 5)
- helps children to extract the invariance of action between unlabeled and labeled action exemplars, which helps them to form action concepts and facilitates subsequent verb learning (Chapter 5, Experiment 6)
- helps children to form action concepts based on schematic information via a top-down scaffolding process, which promotes verb learning more strongly than seeing extra action exemplars (Chapter 5, Experiment 7)
- promotes word-specific knowledge which is useful for learning individual verbs (Chapter 6, Experiments 8 & 9)
- promotes word-category knowledge about verbs which children can use for verb learning even in the absence of iconic gestures (Chapter 6, Experiments 8 & 9)

- promotes *lasting* word-category knowledge about verbs which children can use for verb learning even after a one-week delay (Chapter 6, Experiment 9).

Taken together, the studies reported in this dissertation demonstrate beneficial effects of seeing iconic gestures on children's action event memory and verb learning. Some of the findings also provide new insights into *how* iconic gestures function in this regard. Iconic gestures depict actions in a schematic way, and this schematic information is "light-weight" and easy to remember (Kita et al., 2017). As such, iconic gestures help children to form action concepts, even before a linguistic label for these actions is introduced. When children have access to an action concept when they hear an action label for the first time, they form an initial semantic representation of this label that is ready to be generalized to novel events. In conclusion, seeing iconic gestures facilitates action event memory and learning of individual verbs in 3-year-old children. Moreover, children can deduce information from word-specific iconic gestures that shows them *how* to learn verbs.

References

Ahlner, F., & Zlatev, J. (2010). Cross-modal iconicity: A cognitive semiotic approach to sound symbolism. *Sign Systems Studies*, *38*, 298–348.

Alcock, K., Meints, K., & Rowland, C. (2017). *UK-CDI Words and Gestures - Preliminary norms and manual.* Retrieved from `http://lucid.ac.uk/ukcdi`

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*(4), 468–523.

Anderson, R., & Freebody, P. (1979). Vocabulary knowledge. In J. Guthrie (Ed.), *Reading comprehension and education* (pp. 77–117). Neward, DE: International Reading Association.

Andrén, M. (n.d.). *ISGS: International Society for Gesture Studies.* Retrieved on 04-09-2017, from `http://www.gesturestudies.com/`

Aussems, S., & Kita, S. (in press). Seeing iconic gestures while encoding events boosts children's memory of these events. *Child Development*.

Aussems, S., Kwok, N., & Kita, S. (2017). Digital resource to support: GestuRe and ACtion Exemplar (GRACE) video database [online]. University of Warwick, Department of Psychology. *Warwick Research Archive Portal*. Retrieved from `http://wrap.warwick.ac.uk/78493`

Aussems, S., Kwok, N., & Kita, S. (in press). GestuRe and ACtion Exemplar (GRACE) video database: Stimuli for research on manners of human locomotion and iconic gestures. *Behavior Research Methods*. doi: 10.3758/s13428-017-0942-2

Austin, E., & Sweller, N. (2017). Getting to the elephants: Gesture and preschoolers' comprehension of route direction information. *Journal of Experimental Child Psychology*, *163*, 1–14. doi: 10.1016/j.jecp.2017.05.016

Barr, D., Levy, R., Scheepers, C., & Tily, H. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. doi: 10.1016/j.jml.2012.11.001

Bates, D., Mächeler, M., Bolker, B., & Walker, S. (2013). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. doi: 10.18637/jss.v067.i01

Bates, E. (1976). *Language and context: The acquisition of pragmatics.* New York, NY: Academic Press.

Bates, E., Beeghly, M., Bretherton, I., Harris, C., Marchman, V., McNew, S., ... Whitesell, K. (1986). *The Language and Gesture Inventon: Unpublished checklist.* University of California, San Diego.

Bates, E., Benigni, L., Bretherton, I., Camaioni, L., & Volterra, V. (1979). *The emergence of symbols: Cognition and communication in infancy.* New York, NY: Academic Press.

Bavelas, J., Chovil, N., Lawrie, D., & Wade, A. (1992). Interactive gestures. *Discourse Processes*, *15*(4), 469–489.

Behrend, D. (1990). The development of verb concepts: children's use of verbs to label familiar and novel events. *Child Development*, *61*(3), 681–696. doi: 10.2307/1130953

Bergelson, E., & Swingley, D. (2012). At 6 to 9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences of the USA*, *109*, 3253–3258. doi: 10.1073/pnas.1113380109

Booth, A., & Waxman, S. (2002). Object names and object functions serve as cues to categories for infants. *Developmental Psychology*, *38*(6), 948–957.

Booth, A., Waxman, S., & Huang, Y. (2005). Conceptual information permeates word learning in infancy. *Developmental Psychology*, *41*(3), 491–505. doi: 10.1037/0012-1649.41.3.491

Booth, A. E., McGregor, K. K., & Rohlfing, K. J. (2008). Socio-Pragmatics and Attention: Contributions to Gesturally Guided Word Learning in Toddlers. *Language Learning and Development*, *4*(3), 179-202.

Boyatzis, C., & Watson, M. (1993). Preschool children's symbolic representation of objects through gestures. *Child Development*, *64*(3), 729–735. doi: 10.1111/j.1467-8624.1993.tb02939.x

Bredin, H. (1996). Onomatopoeia as a figure and a linguistic principle. *New Literary History*, *27*(3), 555-569. doi: 10.1353/nlh.1996.0031

Broaders, S., Cook, S., Mitchell, Z., & Goldin-Meadow, S. (2007). Making children gesture brings out implicit knowledge and leads to learning. *Journal of Experimental Psychology: General*, *136*(4), 271–288. doi: 10.1177/0956797610366082

Broaders, S., & Goldin-Meadow, S. (2010). Truth is at hand: How gesture adds information during investigative interviews. *Psychological Science*, *21*, 623–628. doi: 10.1177/0956797610366082

Burdett, E., Lucas, A., Buchsbaum, D., McGuigan, N., Wood, L., & Whiten, A. (2016). Do children copy an expert or a majority? Examining selective learning in instrumental and normative contexts. *PLoS ONE*, *11*(10), e0164698. doi: 10.1371/journal.pone.0164698

Butterworth, G. (2003). Pointing is the royal road to language for babies. In S. Kita (Ed.), *Pointing. Where language, culture, and cognition meet* (pp. 3–99). Mahwah, NJ: Erlbaum.

Butterworth, G., & Morisette, P. (1996). Onset of pointing and the acquisition of language in infancy. *Journal of Reproductive and Infant Psychology*, *14*, 219–231.

Camurri, A., Mazzarino, B., Ricchetti, M., Timmers, R., & Volpe, G. (2004). Multimodal analysis of expressive gesture in music and dance performances. In A. Camurri & G. Volpe (Eds.), *Gesture-Based Communication in Human-Computer Interaction: 5th International Gesture Workshop* (pp. 20–39). Berlin, Heidelberg: Springer Berlin Heidelberg. doi: 10.1007/978-3-540-24598-8_3

Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Proceedings of the Stanford Child Language Conference*, *15*, 17–29.

Cassell, J., McNeill, D., & McCullough, K. (1999). Speech-gesture mismatches: Evidence for one underlying representation of linguistic and nonlinguistic information. *Pragmatics and Cognition*, *7*(1), 1–34. doi: 10.1207/s15327973rlsi2703_4

Childers, J. (2011). Attention to multiple events helps two-and-a-half-year-olds extend new verbs. *First Language*, *30*(1), 3–22. doi: 10.1177/0142723710361825

Childers, J., & Tomasello, M. (2002). Two-year-olds learn novel nouns, verbs, and conventional actions from massed or distributed exposures. *Developmental Psychology*, *38*(6), 967–978.

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. doi: 10.1037/a0013157

Church, R., & Goldin-Meadow, S. (1986). The mismatch between gesture and speech as an index of transitional knowledge. *Cognition*, *23*(1), 43–71. doi: 10.1016/0010-0277(86)90053-3

Clark, E. (1995). *The lexicon in acquisition.* Cambridge, UK: Cambridge University Press.

Clements, W., Rustin, C., & McCallum, S. (2000). Promoting the transition from implicit to explicit understanding: A training study of false belief. *Developmental Science*, *3*(1), 81–92. doi: 10.1111/1467-7687.00102

Clerkin, E., Hart, E., Rehg, J., Yu, S., & Smith, L. (2017). Real-world visual statistics and infants' first-learned object names. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *372*(1711), 20160055. doi: 10.1098/rstb.2016.0055

Coletta, J., Guidetti, M., Capirci, O., Cristilli, C., Demir, O., Kunene-Nicolas, R., & Levine, S. (2014). Effects of age and language on co-speech gesture production: an investigation of french, american, and italian children's narratives. *Journal of Child Language*, *41*(1), 122–145. doi: 10.1017/S0305000913000585

Cook, S., Duffy, R., & Fenn, K. (2013). Consolidation and transfer of learning after observing hand gesture. *Child Development*, *84*(6), 1863–1871. doi: 10.1111/cdev.12097

Cook, S., Yip, T., & Goldin-Meadow, S. (2010). Gesturing makes memories that last. *Journal of Memory and Language*, *63*, 465–475. doi: 10.1016/j.jml.2010.07.002

Davis, R. (1961). The fitness of names to drawings. A cross-cultural study in Tanganyika. *British Journal of Psychology*, *52*(3), 259–268. doi: 10.1111/j.2044-8295.1961.tb00788.x

de Ruiter, J. (2000). The production of gesture and speech. In D. McNeill (Ed.), *Language and Gesture: Window into Thought and Action* (pp. 284–311). Cambridge, UK: Cambridge University Press.

de Saussure, F. (1916). *Course in general linguistics.* New York, NY: McGraw-Hill.

Enfield, N. (2001). 'Lip-pointing': A discussion of form and function with reference to data from Laos. *Gesture*, *1*(2), 185–212.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149–1160. doi: 10.3758/BRM.41.4.1149

Feldman, H., Dale, P., Campbell, T., Colborn, D., Kurs-Lasky, M., Rockette, H., & Paradise, J. (2005). Concurrent and predictive validity of parent reports of child language at ages 2 and 3 years. *Child Development*, *76*(4), 461–468. doi: 10.1111/j.1467-8624.2005.00882.x

Fenson, L., Marchman, V., Thal, D., Dale, P., Reznick, J., & Bates, E. (2006). *Macarthur-bates communicative development inventories: User's guide and technical manual* (2nd ed.). Baltimore, MD: Brookes Publishing Co.

Feyereisen, P. (2006). Further investigations on the mnemonic effect of gestures:

Their meaning matters. *European Journal of Cognitive Psychology*, *18*(2), 185–205. doi: 10.1080/09541440540000158

Forbes, J., & Farrar, M. (1993). Children's initial assumptions about the meaning of novel motion verbs: Biased and conservative? *Cognitive Development*, *8*(3), 273–290.

Forbes, J., & Farrar, M. (1995). Learning to represent word meaning: What initial training events reveal about children's developing action verb concepts. *Cognitive Development*, *10*(1), 1–20. doi: 10.1016/0885-2014(95)90016-0

Fulkerson, A., & Haaf, R. (2003). The influence of labels, nonlabeling sounds, and source of auditory input on 9- and 15-month-olds' object categorization. *Infancy*, *4*(3), 349–369. doi: 10.1207/S15327078IN0403_03

Furman, R., Küntay, A., & Özyürek, A. (2014). Early language-specificity of children's event encoding in speech and gesture: evidence from caused motion in Turkish. *Language, Cognition and Neuroscience*, *29*(5), 620–634. doi: 10.1080/01690965.2013.824993

Gentner, D. (1978). On relational meaning: The acquisition of verb meaning. *Child Development*, *49*, 988–999.

Gentner, D. (1981). Some interesting differences between verbs and nouns. *Cognition and Brain Theory*, *4*(2), 161–178.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. Kuczaj (Ed.), *Language development: Language, thought and culture* (Vol. 2, pp. 301–334). Hillsdale, NJ: Erlbaum.

Gentner, D., & Boroditsky, L. (2001). Individuation, relational relativity and early word learning. In M. Bowerman & S. Levinson (Eds.), *Language acquisition and conceptual development.* Cambridge, England: Cambridge University Press.

Gillibrand, R., Lam, V., & O'Donnell, V. (2011). *Developmental Psychology.* Pearson Prentice Hall.

Goldfield, B., & Reznick, J. (1990). Early lexical acquisition: rate, content, and the vocabulary spurt. *Journal of Child Language*, *17*(1), 171–183. doi: 10.1017/S0305000900013167

Goldin-Meadow, S. (2003). *Hearing gesture.* Cambridge, MA: Harvard University Press.

Goldin-Meadow, S. (2015). From action to abstraction: Gesture as a mechanism of change. *Developmental Review*, *38*, 167–184. doi: 10.1016/j.dr.2015.07.007

Goldin-Meadow, S., Alibali, M., & Church, R. (1993). Transitions in concept acquisition: Using the hand to read the mind. *Psychological Review*, *100*(2), 279–297.

Golinkoff, R., Chung, H., Hirsh-Pasek, K., Liu, L., Bertenthal, B., & Brand, R. (2002). Young children can extend motion verb labels to point-light displays. *Developmental Psychology*, *38*(4), 604–614. doi: 10.1037//0012-1649.38.4.604

Goodrich, W., & Hudson Kam, C. (2009). Co-speech gesture as input in novel verb learning. *Developmental Science*, *12*(1), 81–87. doi: 10.1111/j.1467-7687.2008.00735.x

Goodwin, C. (2003). Pointing as situated practice. In S. Kita (Ed.), *Pointing: Where Language, Culture, and Cognition Meet* (pp. 217–241). Mahwah, NJ: Erlbaum.

Graham, S., Turner, J., & Henderson, A. (2005). The influence of object pre-exposure on 2-year-olds' disambiguation of novel labels. *Journal of Child Language*, *32*, 207–222. doi: 10.1017/S030500090400666X

Gu, Y., Mol, L., Hoetjes, M., & Swerts, M. (2017). Conceptual and lexical effects on gestures: the case of vertical spatial metaphors for time in Chinese. *Language, Cognition and Neuroscience*, 1–16. doi: 10.1080/23273798.2017.1283425

Haryu, E., Imai, M., & Okada, H. (2011). Object similarity bootstraps young children to action-based verb extension. *Child Development*, *82*(2), 674–686. doi: 10.1111/j.1467-8624.2010.01567.x

Heine, A., Thaler, V., Tamm, S., Hawelka, S., Schneider, M., Torbeyns, J., ... Jacobs, A. (2009). What the eyes already 'know': using eye movement measurement to tap into children's implicit numerical magnitude representations. *Infant and Child Development*, *15*(2), 175–186. doi: 10.1002/icd.640

Hendriks, P. (2014). *Assymmetries between language production and comprehension* (Vol. 42) (No. 1). Springer Netherlands. doi: 10.1007/978-94-007-6901-4

Hirsh-Pasek, K., & Golinkoff, R. (2006). *Action Meets Word: How children learn verbs.* New York, NY: Oxford University Press.

Hockett, C. (1960). The origin of speech. *Scientific American*, *203*, 89–96.

Hodges, L., Özçalışkan, S., & Williamson, R. (2015). How early do children understand different types of iconicity in gesture? In E. Grillo & K. Jepson (Eds.), *Proceedings of the 39th annual boston university conference on language development* (p. 70-77). Somerville, MA: Cascadilla Press.

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development*, *74*(5), 1368–1378.

Hoff, E. (2006). How social contexts support and shape language development.

Developmental Review, *26*, 55–88. doi: 10.1016/j.dr.2005.11.002

Hoff, E., Laursen, B., & Tardif, T. (2002). Socioeconomic status and parenting. In M. Bornstein (Ed.), *Handbook of parenting. ecology and biology of parenting* (Vol. II). Mahwah, New Jersey: Lawrence Erlbaum.

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

Huttenlocher, J., Vasilyeva, M., Cymerman, E., & Levine, S. (2002). Language input at home and at school: Relation to child syntax. *Cognitive Psychology*, *45*, 337–374.

Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., & Hedges, L. (2010). Sources of variability in children's language growth. *Cognitive Psychology*, *61*(4), 343–365. doi: 10.1016/j.cogpsych.2010.08.002

Imai, M. (2008). Children's use of argument structure, meta-knowledge of the lexicon, and extra-linguistic contextual cues in inferring meanings of novel verbs. In S. Müller (Ed.), *Proceedings of the 15th international conference on head-driven phrase structure grammar* (pp. 399–416). Stanford, CA: CSLI Publications.

Imai, M., Haryu, E., & Okada, H. (2005). Mapping novel nouns and verbs onto dynamic action events: are verb meanings easier to learn than noun meanings for Japanese children? *Child Development*, *76*(2), 340–355. doi: 10.1111/j.1467-8624.2005.00849.x

Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *369*(1651), 20130298. doi: 10.1098/rstb.2013.0298

Imai, M., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition*, *109*(1), 54–65. doi: 10.1016/j.cognition.2008.07.015

Imai, M., Li, L., Haryu, E., Okada, H., Hirsh-Pasek, K., Golinkoff, R., & Shigematsu, J. (2008). Novel noun and verb learning in Chinese-, English-, and Japanese-speaking children. *Child Development*, *79*(4), 979–1000.

Imai, M., Miyazaki, M., Yeung, H., Hidaka, S., Kantartzis, K., Okada, H., & Kita, S. (2015). Sound symbolism facilitates word learning in 14-month-olds. *PLoS ONE*, *10*(2), e0116494. doi: 10.1371/journal.pone.0116494

Irwin, F., & Newland, E. (1940). A genetic study on the naming of visual studies. *The Journal of Psychology*, *9*(1), 3–16. doi: 10.1080/00223980.1940.9917674

Iverson, J., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science*, *16*(5), 367–371.

Janssen, D. (2012). Twice random, once mixed: Applying mixed models to simultaneously analyze random effects of language and participants. *Behavior Research Methods*, *44*(1), 232–247. doi: 10.3758/s13428-011-0145-1

Kantartzis, K., Imai, M., & Kita, S. (2011). Japanese sound-symbolism facilitates word learning in english-speaking children. *Cognitive Science*, *35*(3), 575–586. doi: 10.1111/j.1551-6709.2010.01169.x

Kelly, S., & Church, R. (1998). A comparison between children's and adults' ability to detect conceptual information conveyed through representational gestures. *Child Development*, *69*, 85–93.

Kelly, S., Özyürek, A., & Maris, E. (2010). Two sides of the same coin: speech and gesture mutually interact to enhance comprehension. *Psychological Science*, *21*(2), 260–267. doi: 10.1177/0956797609357327

Kendon, A. (1982). The study of gesture: some remarks on its history. *Semiotic Inquiry*, *2*, 45–62.

Kersten, A., & Smith, L. (2002). Attention to novel objects during verb learning. *Child Development*, *73*(1), 93–109. doi: 10.1111/1467-8624.00394

Kirk, E., Gurney, D., Edwards, R., & Dodimead, C. (2015). Handmade memories: The robustness of the gestural misinformation effect in children's eyewitness interviews. *Journal of Nonverbal Behavior*, *39*, 259–273. doi: 10.1007/s10919-015-0210-z

Kita, S. (2000). How representational gestures help speaking. In D. McNeill (Ed.), *Language and Gesture: Window into Thought and Action* (pp. 162–185). Cambridge, UK: Cambridge University Press. doi: 10.1017/cbo9780511620850.011

Kita, S., Alibali, M., & Chu, M. (2017). How do gestures influence thinking and speaking? The gesture-for-conceptualization hypothesis. *Psychological Review*, *124*(3), 245–266. doi: 10.1037/rev0000059

Kita, S., & Eggesbey, J. (2001). Pointing left in Ghana: How a taboo on the use of the left hand influences gestural practice. *Gesture*, 73–94.

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*(1), 16-32. doi: 10.1016/S0749-596X(02)00505-3

Köhler, W. (1929). *Gestalt psychology.* New York, NY: Liveright Publishers.

Kucker, S., & Samuelson, L. (2011). The first slow step: Differential effects of object and word-form familiarization on retention of fast-mapped words. *Infancy*, *17*(3), 295–323. doi: 10.1111/j.1532-7078.2011.00081.x

Kuhn, H. (1956). Variants of the Hungarian method for assignment problems. *Naval Research Logistics Quarterly*, *3*, 253–258.

Kuhn, H., & Yaw, B. (1955). The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, *2*(1-2), 83–97.

Kunihira, S. (1971). Effects of the expressive voice on phonetic symbolism. *Journal of Verbal Learning and Verbal Behavior*, *10*, 427–429. doi: 10.1016/S0022-5371(71)80042-7

Langton, S., O'Malley, C., & Bruce, V. (1996). Actions speak no louder than words: Symmetrical cross-modal interference effects in the processing of verbal and gestural information. *Journal of Experimental Psychology-Human Perception and Performance*, *22*(6), 1357–1375. doi: 10.1037/0096-1523.22.6.1357

Lubans, D., Morgan, P., Cliff, D., Barnett, L., & Okely, A. (2010). Fundamental movement skills in children and adolescents: review of associated health benefits. *Sports Medicine*, *40*, 1019–1035. doi: 10.2165/11536850-000000000-00000

Lüke, C., & Ritterfeld, U. (2014). The influence of iconic and arbitrary gestures on novel word learning in children with and without SLI. *Gesture*, *14*, 204–225. doi: 10.1075/gest.14.2.04luk

Madan, C., & Singhal, A. (2012). Using actions to enhance memory: effects of enactment, gestures, and exercise on human memory. *Frontiers in Psychology*, *3*, 507. doi: 10.3389/fpsyg.2012.00507

Maguire, M., Hennon, E., Hirsh-Pasek, K., Golinkoff, R., Slutzky, C., & Sootsman, J. (2002). Mapping words to actions and events: How do 18-month-olds learn a verb? In B. Skarabela, S. Fish, & A. Do (Eds.), *Proceedings of the 27th Annual Boston University Conference on Language* (pp. 371–382). Somerville, MA: Cascadilla Press.

Maguire, M., Hirsh-Pasek, K., Golinkoff, R., & Brandone, A. (2008). Focusing on the relation: Fewer exemplars facilitate children's initial verb learning and extension. *Developmental Science*, *11*(4), 628–634. doi: 10.1111/j.1467-7687.2008.00707.x

Malt, B., Ameel, E., Imai, M., Gennari, S., Saji, N., & Majid, A. (2014). Human locomotion in languages: Constraints on moving and meaning. *Journal of Memory and Language*, *74*(7), 103–127. doi: 10.1016/j.cognition.2008.08.009

Malt, B., Gennari, S., Ameel, E., Tsuda, N., & Majid, A. (2008). Talking about walking: Biomechanics and the language of locomotion. *Psychological Science*, *19*(3), 232–240. doi: 10.1111/j.1467-9280.2008.02074.x

Marchman, V., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science*, *11*(3), 461–468. doi: 10.1111/j.1467-7687.2008.00671.x

Marentette, P., Pettenati, P., Bello, A., & Volterra, V. (2016). Gesture and symbolic representation in Italian and English-speaking Canadian 2-year-olds. *Child Development*, *87*(3), 944–961. doi: 10.1111/cdev.12523

Markman, E., & Wachtel, G. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*(2), 121–157. doi: 10.1037/a0014838

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

Maurer, D., Pathman, T., & Mondloch, C. (2006). The shape of boubas: soundshape correspondences in toddlers and adults. *Developmental Science*, *9*(3), 316–322. doi: 10.1111/j.1467-7687.2006.00495.x

McCarthy, D. (1954). Language development in children. In L. Carmichael (Ed.), *Manual of child psychology* (pp. 476–581). New York, NY: Wiley. doi: 10.1017/cbo9780511620850.011

McGregor, K., Rohlfing, K., Bean, A., & Marschner, E. (2009). Gesture as a support for word learning: The case of under. *Journal of Child Language*, *36*(4), 807–828. doi: 10.1017/S0305000908009173

McNeill, D. (1985). So you think gestures are nonverbal? *Psychological Review*, *92*(3), 350-371.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought.* Chicago, IL: University of Chicago Press.

McNeill, D. (2000). *Language and Gesture: Window into Thought and Action.* Cambridge, UK: Cambridge University Press.

McNeill, D. (2005). *Gesture and thought.* University of Chicago Press.

McNeill, D. (2009). Gesture and Communication. In K. Brown & J. Mey (Eds.), *Concise encyclopedia of pragmatics* (2nd ed., pp. 299–307). Elsevier Science.

McNeill, D., Cassell, J., & McCullough, K. (1994). Communicative effects of speech-mismatched gestures. *Research on Language and Social Interaction*, *27*(3), 223–237.

Meir, I., & Tkachman, O. (2014). Iconicity. In *Oxford bibliographies online.* Oxford University Press. doi: 10.1093/OBO/9780199772810-0182

Mervis, C., & Bertrand, J. (1995). Early lexical acquisition and the vocabulary spurt: a response to goldfield & reznick. *Journal of Child Language*, *22*(2), 461–468.

Milward, S., Kita, S., & Apperly, I. (2014). The development of co-representation effects in a joint task: Do children represent a co-actor? *Cognition*, *132*(3), 269–279. doi: 10.1016/j.cognition.2014.04.008

Morgan, P., Farkas, G., Hillemeier, M., & Scheffner Hammer, C. (2015). 24-month-old children with larger oral vocabularies display greater academic and behavioral functioning at kindergarten entry. *Child Development*, *86*(5), 1351–1370. doi: 10.1111/cdev.12398

Mumford, K. (2014). *The relationship between vocabulary and gesture development in early childhood and infancy.* (PhD Dissertation). University of Birmingham.

Mumford, K., & Kita, S. (2014). Children use gesture to interpret novel verb meanings. *Child Development*, *85*(3), 1181–1189. doi: 10.1111/cdev.12188

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*, 133–142. doi: 10.1016/j.jml.2013.11.001

Namy, L. (2008). Recognition of iconicity doesn't come for free. *Developmental Science*, *11*, 841–846. doi: 10.1111/j.1467-7687.2008.00732.x.

Novack, M., Congdon, E., Hemani-Lopez, N., & Goldin-Meadow, S. (2014). From action to abstraction: Using the hands to learn math. *Psychological Science*, *25*, 903–910. doi: 10.1177/0956797613518351

Novack, M., & Goldin-Meadow, S. (2016). Gesture as representational action: A paper about function. *Psychonomic Bulletin Review*, *24*, 652–656. doi: 10.3758/s13423-016-1145-z

Ota, M., & Skarabela, B. (2016). Reduplicated words are easier to learn. *Language Learning and Development*, *12*(4), 380–397. doi: 10.1080/15475441.2016.1165100

Overton, W., & Jackson, J. (1973). The representation of hand-as-hands in action sequences: A developmental study. *Child Development*, *44*, 309–314. doi: 10.2307/1128052

Owens, R. (2008). *Language development: An introduction (7th ed.).* Needham, MA: Alynn and Bacon.

Özçalışkan, S., Gentner, D., & Goldin-Meadow, S. (2014). Do iconic gestures pave the way for children's early verbs? *Applied Psycholinguistics*, *35*(6), 1–20. doi: 10.1017/S0142716412000720

Özçalışkan, S., Lucero, C., & Goldin-Meadow, S. (2016). Does language shape silent gestures? *Cognition*, *148*(1), 10–18. doi: 10.1016/j.cognition.2015.12.001

Ozturk, O., Krehm, M., & Vouloumanos, A. (2012). Sound symbolism in infancy: evidence for sound-shape cross-modal correspondences in 4-month-olds. *Journal of Experimental Child Psychology*, *114*(2), 173–186. doi: 10.1016/j.jecp.2012.05.004

Özyürek, A., & Kita, S. (1999). Expressing manner and path in English and Turkish: Differences in speech, gesture, and conceptualization. In M. Hahn & S. Stoness (Eds.), *Proceedings of the 21st Annual Conference of the Cognitive Science Society* (pp. 507–512). London: Erlbaum.

Özyürek, A., Kita, S., Allen, S., Furman, R., Brown, A., & Ishizuka, T. (2008). Development of cross-linguistic variation in speech and gesture: Motion events in English and Turkish. *Developmental Psychology*, *44*(4), 1040–1054. doi: 10.1037/0012-1649.44.4.1040

Özyürek, A., Willems, R., Kita, S., & Hagoort, P. (2007). On-line integration of semantic information from speech and gesture: Insights from event-related brain potentials. *Journal of Cognitive Neuroscience*, *19*(4), 605–616. doi: 10.1162/jocn.2007.19.4.605

Peirce, C. (1955). *Collected papers of Charles Sanders Peirce* (Vol. 8; C. Hartshorne, P. Weiss, & A. Burks, Eds.). Cambridge, MA: Harvard University Press.

Perniss, P., Thompson, R., & Vigliocco, G. (2010). Iconicity as a general property of language: evidence from spoken and signed languages. *Frontiers in Psychology*, *1*, 227. doi: 10.3389/fpsyg.2010.00227

Perry, L., Perlman, M., & Lupyan, G. (2015). Iconicity in English and Spanish and its relation to lexical category and age of acquisition. *PLoS ONE*, *10*, e0137147. doi: 10.1371/journal.pone.0137147

Pine, K., Lufkin, N., & Messer, D. (2004). More gestures than answers: Children learning about balance. *Developmental Psychology*, *40*(6), 1059–1067. doi: 10.1037/0012-1649.40.6.1059

Ping, R., & Goldin-Meadow, S. (2008). Hands in the air: Using ungrounded iconic gestures to teach children about conservation of quantity. *Developmental Psychology*, *44*, 1277–1287. doi: 10.1037/0012-1649.44.5.1277

Pulverman, R., Golinkoff, R., Hirsh-Pasek, K., & Sootsman Buresh, J. (2008). Infants discriminate manners and paths in non-linguistic dynamic events. *Cognition*, *108*(3), 825–830. doi: 10.1016/j.cognition.2008.04.009

Pulverman, R., Hirsh-Pasek, K., Golinkoff, R., Pruden, S., & Salkind, S. (2006). Conceptual Foundations for Verb Learning: Celebrating the Event. In K. Hirsh-Pasek & R. Golinkoff (Eds.), *Action Meets Word: How children learn verbs.* Amsterdam / Philadelphia: Oxford University Press. doi:

10.1093/acprof:oso/9780195170009.003.0006

Pulverman, R., Song, L., Golinkoff, R. M., & Hirsh-Pasek, K. (2013). Preverbal infants' attention to manner and path: Foundations for learning relational terms. *Child Development*, *84*(1), 241–252. doi: 10.1111/cdev.12030.

Quine, W. (1960). *Word and object.* The MIT Press.

R Development Core Team. (2011). *R: A Language and Environment for Statistical Computing.* Vienna, Austria. Retrieved from `http://www.R-project.org` (ISBN 3-900051-07-0)

Ramachandran, V., & Hubbard, E. (2001). Synaesthesia: A window into perception, thought and language. *Journal of Consciousness Studies*, *8*(12), 3–34. doi: 10.1016/S0010-0277(02)00108-7

Rescorla, L. (2005). Age 13 language and reading outcomes in late-talking toddlers. *Journal of Speech, Language, and Hearing Research*, *48*, 459–472. doi: 10.1044/1092-4388(2005/031)

Rescorla, L. (2009). Age 17 language and reading outcomes in late-talking toddlers: Support for a dimensional perspective on language delay. *Journal of Speech, Language, and Hearing Researchl*, *52*, 16–30. doi: 10.1044/1092-4388(2008/07-0171)

Roseberry, S., Hirsh-Pasek, K., Parish-Morris, J., & Golinkoff, R. (2009). Live action: Can children learn verbs from video? *Child Development*, *80*(5), 1360–1375. doi: 10.1111/j.1467-8624.2009.01338.x

Rowe, M., & Goldin-Meadow, S. (2009). Early gesture selectively predicts later language learning. *Developmental Science*, *12*(1), 182-189.

Rowe, M., Raudenbusch, S., & Goldin-Meadow, S. (2012). The pace of vocabulary growth helps predict later vocabulary skill. *Child Development*, *83*(2), 508–525. doi: 10.1111/j.1467-8624.2011.01710.x

Salkind, S. (2003). Do you see what I see? In *Paper presented at the 4th Annual University of Delaware Linguistics and Cognitive Science Graduate Student Conference.* Neward, DE.

Salkind, S., Golinkoff, R., & Brandone, A. (2005). Infants' attention to novel actions in relation to the conflation patterns of motion verbs. In R. Golinkoff & K. Hirsh-Pasek (Eds.), *Action packed for language: Prelinguistic foundations for learning relational terms.* Atlanta, GA: Society for Research in Child Development Biennial Meeting.

Schmidt, M., Rakoczy, H., & Tomasello, M. (2016). Young children force social norms selectively depending on the violator's group affiliation. *Cognition*, *123*(3), 325–333. doi: 10.1016/j.cognition.2012.06.004

Scott, R., & Fisher, C. (2012). 2.5-year-olds use cross-situational consistency to learn verbs under referential uncertainty. *Cognition*, *122*(2), 163–180. doi: 10.1016/j.cognition.2011.10.010

Sekine, K., & Kita, S. (2015). Development of multimodal discourse comprehension: cohesive use of space by gestures. *Language, Cognition and Neuroscience*, *30*(10), 1245–1258. doi: 10.1080/23273798.2015.1053814

Sekine, K., Sowden, H., & Kita, S. (2015). The development of the ability to semantically integrate information in speech and iconic gesture in comprehension. *Cognitive Science*, *39*(8), 1855–1880. doi: 10.1111/cogs.12221

Slobin, D. (2004). The many ways to search for a frog: Linguistic typology and the expression of motion events. In S. Strömqvist & L. Verhoeven (Eds.), *Relating events in narrative: Vol. 2: Typological and contextual perspectives* (pp. 219–257). Mahwah, NJ: Lawrence Erlbaum Associates.

Slobin, D. (2006). What makes manner of motion salient? Explorations in linguistic typology, discourse, and cognition. In M. Hickmann & S. Robert (Eds.), *Space in languages: Linguistic systems and cognitive categories* (pp. 59–81). Amsterdam / Philadelphia: John Benjamins.

Slobin, D., & Hoiting, N. (1994). Reference to movement in spoken and signed languages: Typological considerations. In *Proceedings of the 20th Annual Meeting of the Berkeley Linguistics Society* (pp. 487–503). Berkeley, CA: Berkeley Linguistics Society.

Slobin, D., Ibarretxe-Antuñano, I., Kopecka, A., & Majid, A. (2014). Manners of human gait: a crosslinguistic event-naming study. *Cognitive Linguistics*, *25*(4), 701–741. doi: 10.1515/cog-2014-0061

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568. doi: 10.1016/j.cognition.2007.06.010

So, W., Sim, C., & Low, W. (2012). Mnemonic effect of iconic gesture and beat in adults and children. Is meaning important for memory recall? *Language and Cognitive Processes*, *5*, 665–681.

Spencer, 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. doi: 10.1080/01690960802365567

Stanfield, C., Williamson, R., & Özçalışkan, S. (2014). How early do children understand gesture-speech combinations with iconic gestures? *Journal of Child Language*, *41*(2), 462–471. doi: 10.1017/S0305000913000019

Supalla, T. (2009). Theoretical Issues in Sign Language Research: Linguistics. In

S. Fisher & P. Siple (Eds.), (pp. 127–152). Chicago, IL: University of Chicago Press.

Tellier, M. (2008). The effect of gestures on second language memorisation by young children. *Gesture*, *8*, 219–235. doi: 10.1075/gest.8.2.06tel

Tolar, T., Lederberg, A., Gokhale, S., & Tomasello, M. (2008). The development of the ability to recognize the meaning of iconic signs. *Journal of Deaf Studies and Deaf Education*, *13*(2), 2240–240. doi: 10.1093/deafed/enm045

Valenzeno, L., Alibali, M., & Klatzky, R. (2003). Teachers' gestures facilitate students' learning: A lesson in symmetry. *Contemporary Educational Psychology*, *28*(2), 187–204. doi: 10.1016/S0361-476X(02)00007-3

Vogelsang, M., & Tomasello, M. (2016). Giving is nicer than taking: Preschoolers reciprocate based on the social intentions of the distributor. *PLoS ONE*, *11*(1), e0147539. doi: 10.1371/journal.pone.0147539

Vogt, S., & Kauschke, C. (2017). Observing iconic gestures enhances word learning in typically developing children and children with specific language impairment. *Journal of Child Language*, *27*, 1–27. doi: 10.1017/S0305000916000647

Vouloumanos, A., & Werker, J. (2009). Infants' learning of novel words in a stochastic environment. *Developmental Psychology*, *45*(6), 1611–1617. doi: 10.1037/a0016134

Washington, D., & Naremore, R. (1978). Children's use of spatial prepositions in two- and three-dimensional tasks. *Journal of Speech, Language, and Hearing Research*, *21*, 151–165. doi: 10.1044/jshr.2101.151

Waxman, S., & Markow, D. (1995). Words as invitations to form categories: Evidence from 12- to 13-month-old infants. *Cognitive Psychology*, *29*(3), 257–302. doi: 10.1006/cogp.1995.1016

Werner, H., & Kaplan, B. (1963). *Symbolic formation. An organismic-developmental approach to language and the expression of thought.* New York, NY: John Wiley & Sons.

Wood, D., Bruner, J., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, *17*, 89–100.

Wood, J. (2008). Visual memory for agents and their actions. *Cognition*, *108*(2), 522–532. doi: 10.1016/j.cognition.2008.02.012

Yoshida, H. (2012). A cross-linguistic study of sound-symbolism in childrens verb learning. *Cognition*, *13*(2), 232–265. doi: 10.1080/15248372.2011.573515

Yu, C., & Smith, L. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568. doi: 10.1016/j.cognition .2007.06.010

Appendix A

⟶ action pairs

female actors ⟵

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	5.79	6.14	5.90	5.78	6.27	6.33	6.18	5.05	5.26	5.41	5.95	5.77	6.05
2	5.77	5.93	6.01	5.67	5.86	6.63	6.48	5.50	5.82	5.86	5.92	5.05	6.27
3	6.18	6.32	5.69	6.10	5.68	5.88	5.94	5.17	3.99	5.37	6.13	4.62	6.42
4	5.71	5.59	5.73	5.05	5.90	6.49	6.06	5.16	6.18	5.71	6.54	5.92	6.09
5	5.74	6.36	6.05	5.27	6.03	6.10	6.36	5.21	5.38	5.96	5.99	4.64	6.44
6	6.00	6.23	5.95	3.86	5.57	6.29	5.91	5.51	5.20	5.45	5.79	4.87	6.14
7	6.33	6.10	5.82	5.23	6.04	6.33	5.73	5.68	4.71	5.84	6.27	5.84	5.98
8	5.73	5.27	5.69	5.95	5.87	6.51	6.14	5.56	5.09	5.38	6.23	5.45	5.94
9	5.92	6.27	6.20	5.68	5.63	6.36	6.05	5.32	4.98	5.84	6.00	5.80	6.48
10	5.86	6.22	5.36	5.40	5.87	6.43	6.03	5.64	4.93	4.68	5.18	5.05	6.54
11	5.96	5.86	5.92	4.99	5.72	6.45	5.69	4.91	4.98	5.71	6.36	5.71	6.13
12	6.20	5.94	5.25	5.52	6.16	5.86	6.13	5.27	4.73	5.82	5.82	5.70	6.32
13	5.11	6.20	6.09	6.24	6.05	6.71	6.43	5.05	5.20	5.41	5.86	5.83	6.32

Figure A.1: Profit matrix for action–gesture matches of female actors. Stimuli were rated on a seven-point scale, where 1 indicated a very bad match, 4 indicated neither a good nor a bad match, and 7 indicated a very good match. Ratings were averaged over each action pair and actor combination. Column numbers correspond to action pairs and row numbers correspond to the female actors in the database. Grey rectangles indicate the ratings that were selected by the Hungarian algorithm, which maximized the total profit of a one-to-one assignment of female actors to action pairs.

Appendix B

⟶ action pairs

male actors ⟵

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	6.23	5.74	5.70	6.44	5.73	6.59	6.16	5.03	5.64	6.05	5.88	5.65	6.05
2	5.77	6.59	5.26	5.67	6.14	6.10	6.13	6.06	4.68	5.55	5.93	4.53	6.12
3	5.45	5.95	5.82	5.64	5.77	6.18	6.32	5.08	5.14	5.44	5.79	4.66	5.33
4	6.00	6.05	5.77	5.95	6.34	6.00	6.20	5.88	4.55	4.92	6.00	5.50	6.63
5	6.01	6.41	5.91	5.36	5.53	6.54	6.39	5.32	5.09	5.21	5.67	5.71	6.45
6	6.61	6.18	5.45	5.23	5.26	6.17	5.69	5.63	4.36	5.23	5.68	5.19	5.64
7	5.70	6.27	6.05	6.39	5.68	6.08	5.76	4.73	5.13	5.25	6.07	5.29	6.13
8	6.18	6.13	5.77	5.45	5.75	6.74	6.59	5.84	5.38	6.04	6.18	5.86	6.40
9	5.95	6.34	5.73	5.84	6.04	6.39	6.38	4.68	4.87	5.17	6.14	5.41	6.23
10	5.73	5.78	6.05	6.07	6.04	6.23	5.74	4.91	5.29	5.94	6.58	4.91	6.34
11	5.81	5.97	5.72	5.89	6.42	6.36	6.15	5.49	4.89	4.60	5.67	5.73	6.23
12	5.87	5.62	5.67	5.79	6.23	6.14	6.18	5.33	5.14	5.09	5.58	5.63	6.17
13	5.23	5.72	4.72	5.67	5.50	6.41	6.32	5.42	5.09	5.59	6.34	5.23	6.27

Figure B.1: Profit matrix for action–gesture matches of male actors. Stimuli were rated on a seven-point scale, where 1 indicated a very bad match, 4 indicated neither a good nor a bad match, and 7 indicated a very good match. Ratings were averaged over each action pair and actor combination. Column numbers correspond to action pairs and row numbers correspond to the male actors in the database. Grey rectangles indicate the ratings that were selected by the Hungarian algorithm, which maximized the total profit of a one-to-one assignment of male actors to action pairs.

Appendix C

Table C.1: List of Video Files that Were Taken from the GRACE Database for Experiment 5. Column 1 Describes the Action labels for Reference to the Video Files in the Database. Columns 2 and 3 List Numbers 01–13 in the File Names, which Represent Female Actors and Male actors. Column 4 Indicates whether the Experimenter's Hands Represented the Actors' Feet, Legs, or Body.

Actions	Female actor	Male actor	Semiotic type
bowing	12	06	Body
skating	12	06	Feet
wobbling	06	09	Body
marching	06	09	Legs
mermaiding	09	03	Legs
overstepping	09	03	Legs
creeping	03	07	Feet
crisscrossing	03	07	Feet
turning	01	11	Body
hopscotching	01	11	Legs
swinging	13	05	Legs
skipping	13	05	Legs
dropping	08	02	Body
folding	08	02	Legs
twisting	04	01	Body
stomping	04	01	Feet
trotting	05	08	Legs
hopping	05	08	Body
flicking	11	10	Legs
dragging	11	10	Feet
grapevining	07	12	Legs
shuffling	07	12	Feet
groining	10	04	Legs
scurrying	10	04	Feet

Appendix D

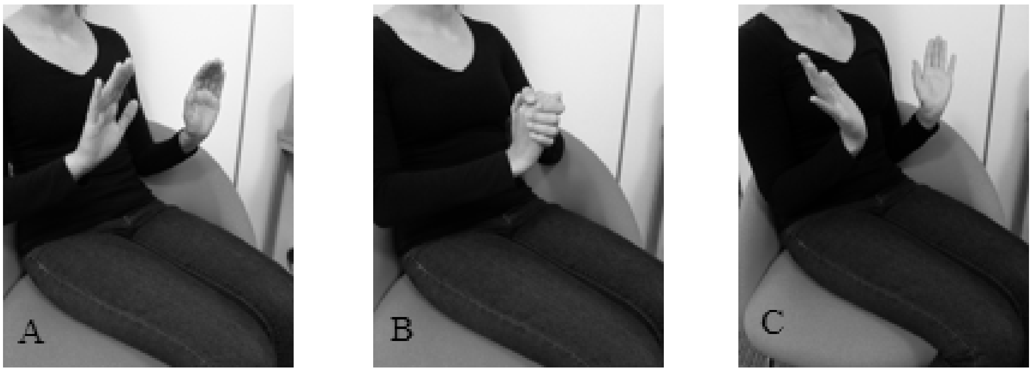


Figure D.1: Three interactive gestures (Bavelas et al., 1992) used in Experiments 5-9. The gestures indicate surprise (Panel A) and excitement (Panel B and C).