

University of Groningen

How hand movements and speech tip the balance in cognitive development

de Jonge-Hoekstra, Lisette

DOI:
[10.33612/diss.172252039](https://doi.org/10.33612/diss.172252039)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2021

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):
de Jonge-Hoekstra, L. (2021). *How hand movements and speech tip the balance in cognitive development: A story about children, complexity, coordination, and affordances*. University of Groningen.
<https://doi.org/10.33612/diss.172252039>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

How hand movements and speech tip the balance in cognitive development

A story about children, complexity, coordination,
and affordances

Lisette de Jonge-Hoekstra

Colofon

How hand movements and speech tip the balance in cognitive development

A story about children, complexity, coordination, and affordances

Design/Lay-out

Lisette de Jonge-Hoekstra

Print

GVO drukkers & vormgevers B.V.

© 2021, Lisette de Jonge-Hoekstra



university of
 groningen

How hand movements and speech tip the balance in cognitive development

A story about children, complexity, coordination, and affordances

PhD thesis

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus Prof. C. Wijmenga
and in accordance with
the decision by the College of Deans.

This thesis will be defended in public on

Thursday 1 July 2021 at 16.15 hours

by

Lisette de Jonge-Hoekstra

born on 21 August 1990
in Dordrecht

Supervisor

Prof. P.L.C. van Geert

Co-supervisors

Dr. R.F.A. Cox

Dr. S. van der Steen

Assessment Committee

Prof. M.W.G. van Dijk

Prof. H.A. Ozyurek

Prof. J. Rączaszek-Leonardi

Table of contents

Chapter 1	General Introduction	7
Chapter 2	Asymmetric dynamic attunement of speech and gestures in the construction of children's understanding	27
Chapter 3	Easier said than done? Task difficulty's influence on temporal alignment, semantic similarity, and complexity matching between gestures and speech	61
Chapter 4	Movers and shakers of cognition: Hand movements, speech, task properties, and variability	95
Chapter 5	Get it together: How do school-aged children interact multimodally and at multiple time scales when they solve dyadic balance scale problems?	123
Chapter 6	General Discussion	157
	Appendices	177
	References	195
	Nederlandse Samenvatting [Dutch summary]	225
	Dankwoord [Acknowledgements]	235
	Curriculum vitae	244



1

General Introduction

General Introduction

When we educate children in primary education, we aim to induce development and teach them something new. While this development in children spans many, equally important, developmental areas, we tend to emphasize *cognitive development*. Inducing cognitive development requires educators and children to communicate. In other words, educators and children need to speak, gesture, and move together for cognitive development to happen (e.g. Novack & Goldin-Meadow, 2015; Pennings et al., 2018; van de Pol et al., 2010; Van der Steen et al., 2012). Moreover, children's hand movements in general, and gestures in specific, have been found to lead cognitive development, over speech (e.g. Adolph et al., 2015; Adolph & Franchak, 2017; Adolph & Kretch, 2015; Alibali & Goldin-Meadow, 1993a; Church & Goldin-Meadow, 1986; Fischer & Bidell, 2006; Goldin-Meadow et al., 1992, 1993; Perry et al., 1992; Roth, 2002).

Existing explanations for gestures' leading role in cognitive development center around abstract concepts, such as implicit and explicit knowledge (Broaders et al., 2007), conflicting cognitive representations (Church & Goldin-Meadow, 1986; Goldin-Meadow et al., 1993; Perry et al., 1992), and cognitive load (Cook et al., 2012; Melinger & Kita, 2007). However, these explanations disregard that moving your hands and speaking is not (only) abstract. Instead, hand movements and speech are 1) actions which involve many physical components at many scales which interact over time, 2) physically coupled to each other, and 3) related and adapted to the physical and social environment.

These three characteristics of hand movements and speech are related to complex dynamical systems, coordination dynamics, and affordances, respectively, which are the theoretical grounds on which this dissertation is build. Previous research from these theoretical perspectives has yielded crucial understanding about diverse areas of child development and skill acquisition (e.g. Adolph et al., 2018; Gibson & Pick, 2000; Smith & Thelen, 2003; Thelen et al., 1987; van Geert, 2008). My goal in this dissertation, based on these theoretical perspectives, is to understand how cognitive development is related to how children move their hands and how they speak during cognitive tasks -over time and at multiple scales-, and how their hand movements and speech relate to each other, and to the physical and social environment.

In this General Introduction, after giving a brief overview of hand movements, speech, and cognition in development, I will introduce the theoretical perspectives of complex dynamical systems, coordination dynamics, and affordances. As it is entirely possible to write whole books with sophisticated detail and mathematical precision about either of the topics that I address in the General Introduction (e.g. E. J. Gibson & Pick, 2000; J. J. Gibson, 1966; Kelso, 1995; Thelen & Smith, 1994; Van Geert, 2008), I will focus on the core ideas of these perspectives.

General Introduction

Hand movement and speech in (cognitive) development

The relation between hand movements and vocalizations starts early in life. Even long before birth, human fetuses have been shown to suck their fingers at 10 to 15 weeks of gestation (e.g. de Vries et al., 1982), and to coordinate hand moving and mouth opening at 19 to 35 weeks of gestation (e.g. Myowa-Yamakoshi & Takeshita, 2006). After birth, the frequency of this hand-mouth-coordination sharply increases (Butterworth & Hopkins, 1988; Sparling et al., 1999). Throughout the first year of life, infants' hand-mouth-coordination differentiates into new patterns, such as bringing objects to their mouth to explore them orally, rhythmical manual banging and vocal babbling, and pointing gestures and saying their first word (for an overview, see Adolph & Franchak, 2017; Iverson & Thelen, 1999). In line with differentiation into new and more patterns, Abney, Warlaumont et al., (2014) found both hand movements and vocalizations of one infant to become more flexible and context-dependent over time, from 51 to 305 days of age. In particular, changes in the variability of the infant's hand movements and vocalizations were related. These early couplings between infants' hand movements and vocalizations provide the basis for more adult-like gestures and speech in communication (Iverson, 2010; Iverson & Fagan, 2004; Iverson & Thelen, 1999).

Hand movements and speech continue to develop hand in hand, also after the first year of life. The coordination between children's hand movements – and body movements in general – and speech is pivotal for prosody development, i.e. how children learn speech rhythm and intonation (Esteve-Gibert & Guellai, 2018). With regard to semantic development, children's first pointing precedes saying their first word, whereby children who were early pointers also tend to be early speakers (for a review, see Goldin-Meadow & Alibali, 2013). Furthermore, the moment of children's first gesture + word combinations predicts the moment of their first word + word combinations (Iverson & Goldin-Meadow, 2005). While children's pointing initially is accompanied by some form of speech for only 40% of the time, gestures predominantly occur together with speech (Esteve-Gibert & Prieto, 2014) after a couple of months. This pattern, of hand movements leading and speech “catching up”, also extends to cognitive development more generally.

With regard to cognitive development, children use their hands to explore and gesture about the world around them (Adolph & Franchak, 2017; Adolph & Kretch, 2015; Chapter 4 of this dissertation). Children (and adults) from all ages reach for objects that interest them, and feel and manipulate these objects using their hands, in ways they are unable to do by speaking. Within primary education, hands-on learning activities also rely on such manual exploration (Fischer & Bidell, 2006; Roth, 2002). When children talk about objects, they also gesture and thereby extend their array of manual action (Roth, 2002). Encouraging children to gesture while

they reason about something they do not yet understand, such as conservation problems or mathematical equivalence problems, fosters their understanding (Broaders et al., 2007), particularly when children are instructed to shape these gestures according to relevant task properties (Brooks & Goldin-Meadow, 2015; Goldin-Meadow et al., 2009).

Moreover, children as young as 5 years old have been shown to convey their “new” understanding in gestures, while simultaneously putting their “old” understanding into words (Church & Goldin-Meadow, 1986; Pine et al., 2004). For example, in the context of a liquid conservation task, a child may still say that one glass contains more water because the level of water is higher (i.e. old understanding = only taking the water level into account), while simultaneously make a C-shape with their hand to indicate the width of the glass in gestures (i.e. new understanding = also taking the width of the glass into account). This phenomenon has been called a gesture-speech mismatch. However, we still grapple to understand how these gesture-speech mismatches fit with, and could originate from, an integrated and tightly coordinated gesture-speech system (Koschmann, 2017; Pouw et al., 2017). To better understand why hand movements in general, and gestures in specific, are leading over speech in cognitive development, we investigate them from a complex dynamical systems perspective in this dissertation.

Complex dynamical systems

Complex dynamical systems are systems that consist of multiple components, typically at multiple scales of a system, which interact and spontaneously coordinate over time by means of self-organization (e.g. Kelso, 1995; Smith & Thelen, 2003; Thelen & Smith, 2007; Van Geert, 1998; Van Geert, 2008; Van Orden et al., 2003; Van Geert, 2019). Examples of complex dynamical systems are weather systems, ant colonies, and the stock market, to name a few. Due to the interactions between components, a complex dynamical system is a whole greater than the sum of its parts. More specifically, interacting components self-organize into global patterns, whereby new patterns emerge. For example, weather systems self-organize into hurricanes, ant colonies self-organize into hyper-efficient trails to bring food into their nest, and the stock market self-organizes into sudden recessions (see Figure 1). Such global patterns, also known as attractors or collective states, are relatively stable, thus tending to resist perturbations – at least to a certain degree. During such stable states, the coupling between a system’s components is strong.

Albeit relatively stable, changes from one stable state to another can occur. This is characterized by a reorganization of a system’s components and their relations. For example, hurricanes tend to dissolve above land, ant colonies reorganize into different trails when they find new food sources, and stock markets reorganize into growth after a recession. Such a

General Introduction

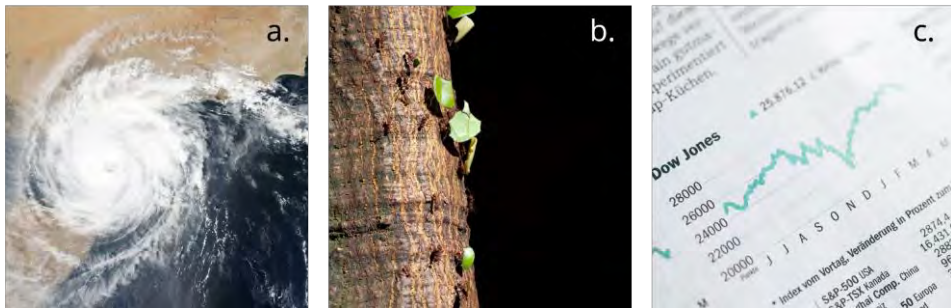


Figure 1. Examples of self-organization in complex dynamical systems. Panel a shows a hurricane, panel b shows an ant trail, and panel c shows a recession.

system reorganization and transition is typically accompanied by an increase of variability in the system's behavior, when the coupling between components weakens. Such an increase in variability is seen as a hallmark of change.

Attractor landscape

The stability of stable states, or attractors, and the variability surrounding transitions between them can be metaphorically described using an attractor landscape (see Figure 2)¹. In this landscape, an attractor is depicted as a well with a certain width and depth. Furthermore, there can be one or multiple wells, corresponding to the existence of an equal number of attractors. More attractors typically indicate that a system is capable of adapting to different circumstances. For example, if someone knows multiple ways to bring food to their mouth, they can adaptively use one to eat either soup or chocolate.

With regard to stability and variability of attractors, one can imagine what would happen with a ball rolling across the landscape. If a well is wide, the chance that the ball rolls in the well is relatively large, as compared to a narrow well. Analogously, some attractors are relatively stronger than others. For instance, when we were on a holiday, my daughter took off to the playground, which was to the right, about 50 times per day. However, when she needed to go to the bathroom, which was to the left, she still would take off to the right, indicating that the running-towards-the-playground-attractor was relatively strong.

Furthermore, if a well is deep, the chance that the ball will get out of the well is relatively small, as compared to a shallow well. This analogy corresponds to some attractors being more stable, or more resistant to perturbations, than others. With regard to the previous example of running

¹ It should be noted that the attractor landscape is only capable of describing a particular type of attractor, namely point attractors. Many other attractors also exist, but explaining them in detail would go beyond the scope of this General Introduction. For beautiful pictures, one can search the internet for "strange attractors".

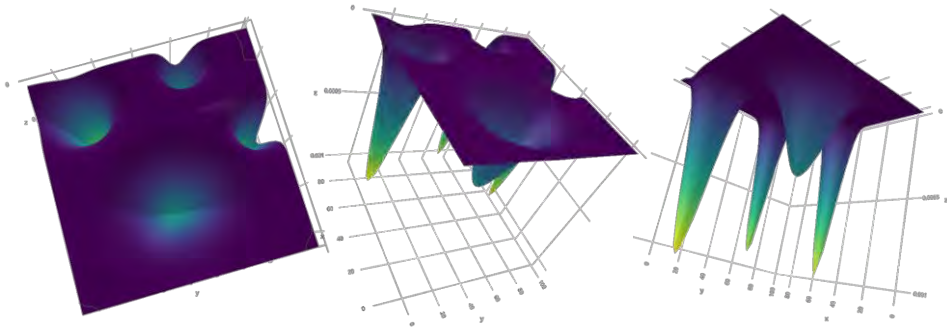


Figure 2. Example of an attractor landscape with four wells, or attractors. These wells differ in width and depth, which analogously correspond to attractors with a different strength and stability, respectively.

towards the playground, my daughter eventually took a U-turn and ran towards the bathroom. This indicates that she did not get stuck running towards the playground, and the attractor thus was not particularly stable.

Lastly, the attractor landscape changes over time, with some wells becoming wider or deeper, while other wells appear or disappear, in correspondence with what can happen with real attractors. Again returning to the previous example, over the course of a couple of days, I noticed that the time it took my daughter to make the U-turn towards the bathroom became less and less, and eventually she immediately ran left towards the bathroom and right towards the playground. In other words, next to the running-towards-the-playground-attractor also a running-towards-the-bathroom-attractor had emerged.

People as complex dynamical systems

As implied by the examples in the previous section, people are complex dynamical systems as well, as they also consist of multiple components at multiple scales which interact over time (e.g. Kelso, 1995; Smith & Thelen, 2003; Thelen & Smith, 2007; Van Geert, 1998; Van Geert, 2008; Van Orden et al., 2003; Van Geert, 2019). For example, people consist of different types of cells, which self-organize into different structures (systems) such as bones, muscles, blood vessels, or brain parts. These structures are self-organized in larger structures such as the skeleton, muscular system, circulatory system, or central nervous system, and these larger structures themselves are all self-organized in a coherent human being. We can scale this example even further up to larger systems (e.g., people are part of a family, which is part of a community), as well as identify intermediate scales, such as the musculoskeletal system, but also the cognitive system (Thelen & Smith, 1994; Thelen & Smith, 2007), or the gesture-system and the speech-system (Iverson & Thelen, 1999; Rusiewicz & Esteve-Gibert, 2018). Just like any complex dynamical system, people's components are coupled at many different scales and self-organize into coherent wholes and stable states, which is dazzlingly complex to realize.

General Introduction

Furthermore, similar to other complex dynamical systems, also people show a destabilization and increase in (various forms of; see previous section) variability upon changes from one stable state to another (e.g., Adolph et al., 2015; Bassano & van Geert, 2007; Shockley et al., 2002; Van der Maas & Molenaar, 1992; Van Geert & Van Dijk, 2002). For example, participants show an increase in hand movement variability upon discovering a new cognitive strategy (Stephen et al., 2009). Moreover, critical fluctuations precede large shifts in symptom severity of patients with a mood disorder (Olthof, Hasselman, Strunk, van Rooij, et al., 2020), and destabilization of self-ratings is related to better intervention outcomes (Olthof, Hasselman, Strunk, Aas, et al., 2020). A last example is that an increase in variability of utterance length precedes structural changes in language development (Bassano & van Geert, 2007; Van Dijk & van Geert, 2011). How change exactly arises from multiple components when they are coupled and “work together”, is the topic of the field of *coordination dynamics*.

Coordination dynamics

Coordination is everywhere. Examples of coordination include a couple dancing the tango, an acrobat juggling with six balls, male fireflies synchronizing their flashes to attract female fireflies, a baby learning to walk (or talk, etc.), Usain Bolt sprinting towards victory, bees running their hive together, or car drivers slowing down and accelerating together during a traffic jam. In all these examples, two or more things are *coupled*, be it physically and/or perceptually. Due to this coupling they adjust their actions to each other, and their behavior becomes coordinated. Moreover, often these coupled systems and their components behave as if they were one – a synergy (e.g. Haken, 1987; Kelso, 2013; Latash, 2008; Strogatz, 2012; Turvey, 2007; also see Warren, 2006).

A synergy is a functional grouping of systems that “work together” and self-organize in the service of a particular “goal” (Kelso, 2013; Latash, 2008; Turvey, 2007). In our previous examples, we can identify functional organizations, such as dancing, juggling, attracting female fireflies, walking, winning, running the hive, and driving somewhere while keeping the car in one piece, respectively. Within a synergy, fluctuations of one component are compensated for by fluctuations of other components, as to preserve the functional organization of the synergy. For example, if Usain Bolt steps on a stone with his foot, muscles in other parts of his body will compensate for this and he will still be able to maintain a stable running pattern, leading him to win the match. If babies step on a stone however, they still lack the ability to compensate for this fluctuation in one component, and they will probably fall. In other words, while the coordination of many components goes smoothly for Usain Bolt, this is not (yet) the case for the baby. Smooth coordination of many components is related to the problem of degrees-of-freedom in motor control.

The problem of degrees-of-freedom entails that any movement, no matter how big or small, entails the coordination of numerous diverse body components (Bernstein, 1967). For instance, uttering one syllable already involves the cooperation of more than 70 muscles (Turvey, 2007). In theory, the number of possible configurations (degrees-of-freedom) for each movement is astronomically large, and controlling each individual component that is involved in a movement would take an immense effort (hence the problem). However, instead of being individually controlled by some central command system, the components self-organize into collectives: Synergies (Haken, 1987; Kelso, 2013). Within a synergy, the degrees-of-freedom are compressed and the components are constrained to act as a functional unit. To maintain this functional unity, changes in degrees-of-freedom in one component of the synergy (e.g. Usain Bolt's foot) are compensated for by changes in degrees-of-freedom in other components of the synergy (e.g. Usain Bolt's muscles in his leg and back). In Kelso's words: "Retaining stability is, for a synergy, the retaining of functional integrity." (Kelso, 2013, p. 1541).

Behavior, however, not only shows stability, but also flexibility, and is adaptive to changing circumstances. That is, there are multiple attractors in an attractor landscape, from which the system can choose. Synergies are task specific and always ready to become something else in a moment. For example, writing this thesis involves the components of the neuromuscular system responsible for my hand movements to change between typing on a keyboard, moving the mouse, writing on paper, grabbing a coffee mug, and fidgeting my hair. This fits with degeneracy in complex dynamical systems, which means that multiple combinations of components can achieve one function, and one combination of components can achieve multiple functions (e.g. Edelman & Gally, 2001; Seifert et al., 2016; Whitacre, 2010). Furthermore, the coordination patterns between components differ for different functions. For example, typing on a keyboard involves different fingers of both my hands to engage in temporally and spatially tightly coordinated movement cycles of pressing and releasing keys (see Figure 3, left panel). Writing on paper, however, involves all the fingers of my right hand to



Figure 3. Coordinative patterns of typing on a keyboard (left panel) and hand writing (right panel).

General Introduction

engage in a cycling motion as well, albeit with a different spatial configuration for each finger (see Figure 3, right panel).

Following von Holst (1938), Kelso (2013) identifies three general patterns of coordination. During *absolute* coordination, components are locked in time - a pattern also known as *phase synchronization* (Pikovsky et al., 2001). The earlier example of fireflies' synchronous flashing illustrates absolute coordination. During *relative* coordination components become locked for some period of time and then unlock again, such as in the earlier example of car drivers slowing down and accelerating during a traffic jam. Lastly, components can go about independently, which is the case with *no* coordination. Furthermore, these three general patterns of coordination can also mix and coexist. In addition, more forms of coordination exist, which will be explained in Chapter 3 (Study 2). Changes between such coordination patterns are called *phase transitions*.

Phase transitions in human motor behavior have been extensively studied within rhythmic motor tasks, following the Haken-Kelso-Bunz (HKB) study paradigm (Haken et al., 1985). In bimanual coordination, people move their fingers in two distinct coordinative patterns at lower speeds: either *in phase* (or parallel; see Figure 4, left panel) or *anti phase* (or mirror; see Figure 4, right panel). However, when people increase their movement speed, they involuntarily switch from anti phase to in phase coordination at a certain threshold, while this is not the case for in phase coordination. In other words, at higher movement speeds only one coordinative pattern is possible, namely in phase coordination. Furthermore, when people lower their movement speed again, the threshold at which they switch back to anti phase coordination is lower than the threshold at which they switched to in phase coordination. This phenomenon is called hysteresis, and shows that coordinative patterns are dependent on what happened before, i.e. the history of the system. In terms of attractors, these findings have been modelled as two stable attractors at lower speeds: An in-phase and an anti-phase attractor. When the speed increases, the stability of the anti-phase attractor increasingly diminishes until it virtually disappears, and only the in-phase attractor exists.

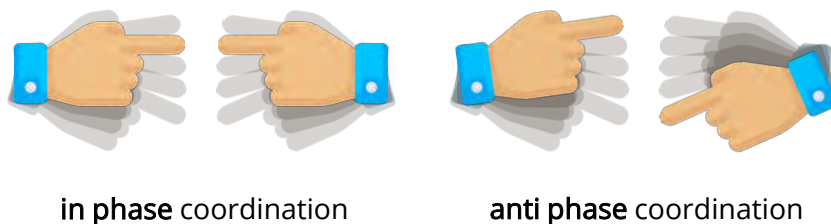


Figure 4. In phase (parallel) bimanual coordination pattern (left panel) and anti phase (mirror) bimanual coordination pattern (right panel) in the Haken-Kelso-Bunz study paradigm.

The findings from the HKB paradigm (i.e. inter-limb coordination), including transition and hysteresis phenomena, have been extended to coordination between manual and vocal actions (i.e. gesture-speech coordination; e.g. Treffner et al., 2008; Treffner & Peter, 2002) and to coordination between people (i.e. interpersonal coordination; e.g. Richardson et al., 2007; Schmidt & Richardson, 2008). Interestingly, Richardson et al. (2007) found that directly looking at each other while rocking in rocking chairs resulted in more stable interpersonal movement coordination, compared to peripherally seeing each other. In other words, *perception-action couplings* between people modulates their movements, and thus modulates the coordinative patterns that arise between them. (e.g. E. J. Gibson & Pick, 2000; J. J. Gibson, 1966; Marsh et al., 2009; Warren, 2006). Not only does this perception-action coupling allow us (and other animals) to adapt our actions to each other, but it also allows us to adapt our actions to our physical surroundings, which is captured by the concept of *affordances*.

Affordances – the match with the environment

Affordances are possibilities for action which the environment offers to the animal, thereby matching its capabilities (e.g. Adolph & Kretch, 2015; E. J. Gibson & Pick, 2000; J. J. Gibson, 1966, 1979). Given the importance of the match between animal and environment, animals (including humans) should be considered within their natural surroundings, doing the things they normally do. Moreover, animals and their surroundings are reciprocal. That is, animals adjust their actions to the environment and the environment offers possibilities for action accordingly. Furthermore, the environment provides information that specifies these action possibilities, which animals attune to and use to guide their actions.

Perception and action thus are reciprocal too, which is known as the *perception-action loop* (e.g. Adolph & Kretch, 2015; E. J. Gibson & Pick, 2000; J. J. Gibson, 1966, 1979). First, perceiving means that the animal *actively gathers information* about things and events in the environment, by means of looking, listening, feeling, tasting, and moving and manipulating the environment to optimize the information. Crucially, the information that flows through the senses is *rich*, in the sense that the energy in the form of light, sound, pressure, and chemicals is structured by objects and events in the environment in a specifying manner. Second, perception informs the animal about the actions it can perform with the objects and about what to expect in the environment (called *prospexion*), while the animal's own movements are informative about its (changing) relation to the environment.

To give an example of perception-action reciprocity in relation to the structure of the environment: When Usain Bolt runs through an area covered with obstacles, he will move and turn his head and body to change his angle of approach towards each obstacle on his way. These changes in visual angle enable Usain Bolt to regulate perception of distance between

General Introduction

him and the obstacles, their size, and time-to-contact, thereby enabling him to avoid them. On the other hand, a young toddler going through the same area will also turn their head and body to optimize perception, but the exact movements will be very different. These different movements are due to differences in the toddler's size, speed, strength, visual accuracy, and motor coordination, compared to Usain Bolt. Moreover, these differences in body capacities also make that the toddler will probably not run through the area and avoid obstacles, but will rather climb them, or walk around them, among other things that the toddler can and would like to do (see also next paragraph). In conclusion, affordances are animal-specific, which means that they depend on the match between an individual's bodily scales and action capabilities and the properties of the environment (e.g. Fajen et al., 2009).

Children need to learn to perceive and realize affordances (possibilities for action; Adolph & Kretch, 2015; E. J. Gibson & Pick, 2000). For example, while a couch affords sitting for an adult, it affords pulling up to stand for a baby (see Figure 5). Throughout development, children need to learn what their body can do and what the environment has to offer to make use of this. By means of exploration, such as a toddler doing different things on and around obstacles, children become increasingly better at attuning to the relevant information specifying the possibilities for action in a given situation. This is a lifelong process, whereby the match between a growing body and increasing action repertoire constantly changes, and new opportunities for actions in the environment continue to arise. Importantly, these new action opportunities, in turn, provide new things to be explored and new skills to be learned. For example, sitting requires strong core muscles to keep the torso stable. When that has been mastered, the child's hands free up, which gives them the possibility to reach for and grasp objects. With this new skill, the child can make all kinds of wonderful discoveries.



Figure 5. Affordances of a couch for an adult (sitting; left panel) and for a baby (pulling up; right panel).

Affordances and structure of the environment are also apparent in children's hand movements and speech, and social interactions. Regarding hand movements, children's hands are crucial to learning about affordances, especially for objects that require fine motor skills to handle (Adolph, 2019; Adolph & Franchak, 2017). Children use their hands to explore these objects: They feel its surface structure, size and weight. Furthermore, they pick it up and turn it in order to see it from different angles, hear the sounds the object is making, and put it in their mouth to taste it and explore its texture. In addition, they 'use' the objects to explore the surfaces around them, by bouncing on them or by using an object to change something in their surroundings (i.e. tool use; Lockman, 2000; Smitsman & Bongers, 2003). Gestures are also hand movements, whereby gestures can be thought of as moving one's hands according to the rhythmic structure of speech (e.g. Wagner et al., 2014, also see Pouw et al., 2018), as well as according to the spatial structure of objects and events in the environment.

With regard to vocalizations and speech, from early on vocalizations (e.g. crying) are very effective to elicit or stop someone else's actions within particular situations. Furthermore, babies very quickly learn that making sounds, such as cooing, captures their caregiver's attention for longer periods of time (e.g. Jaffe et al., 2001). During these interactions, caregivers actively and voluntarily as well as involuntarily structure children's vocalization patterns (e.g. reacting, turn taking, mimicking), and at the same time over-emphasize the relevant acoustic structure of their own speech (e.g. so-called *motherese*) (e.g. Stern et al., 1983). In other words, embedded within everyday social interactions, children learn to mutually and adaptively structure their vocal sounds on many levels with their interaction partners (e.g. Reed, 1995; van Dijk et al., 2013). Speaking thereby opens up many new possibilities for action together with other people, such as collaborating, sharing thoughts and feelings, and teaching and learning about cognitive tasks, which extend to both the past, present, and future (e.g. Smith & Gasser, 2005).

Cognitive development from the perspective of complex dynamical systems, coordination dynamics, and affordances

A recent review (Adolph & Hoch, 2019; also see Adolph, 2019; Adolph et al., 2018; Newen et al., 2018) summarizes the characteristics of motor development as embodied, embedded, enculturated and enabling. Embodied refers to the fact that the current specifics of the body determine possibilities for action, embedded implies that the environment opens up and constrains possibilities for action, enculturated indicates that motor development is shaped by social and cultural forces, and enabling means that each new skill opens up a whole new range of opportunities to learn other skills, and thereby can bring about a developmental cascade. This echoes the descriptions already given above about new possibilities for action, which

General Introduction

continuously arise throughout development. Following previous researchers (e.g. Kloos & Van Orden, 2009; Thelen & Smith, 1994; Thelen & Smith, 2007), I am convinced that these characteristics of motor development also apply to cognitive development in general.

Based on the framework above, cognitive understanding within cognitive development is the equivalent of what a motor skill is within motor development. This entails that cognitive understanding is a functional coordination pattern too, similar to motor skills. Functional hereby means that it arises when a particular child is in a particular physical and social environment, such as when an adult asks them to explain about a particular task (see e.g. Study 1 and 3/Chapter 2 and 4). Depending on the specifics of the environment, cognitive understanding can take many forms, such as talking and gesturing, but also writing on paper, or hands-on problem solving. Similarly, also motor skills come in many different forms, such as walking, running, climbing, or swimming, depending on the environment that someone is in and the particular motor problem one is confronted with, such as moving on a horizontal surface, a slanting or vertical surface, or in the water, respectively. This suggests that any form of cognitive understanding, just like any particular form of motor skills, only exists for a specific child doing a concrete task in a specific environment.

Opponents of such a view typically emphasize that viewing cognitive understanding about a particular concept as being similar to a motor skill ignores that cognitive understanding, at least in part, is abstract, symbolic, disembodied and ungrounded. This expresses that cognitive understanding about a particular task, once it is well-developed, is supposed to happen “in someone’s head”, and is thereby relatively independent from the specific environment that someone is in or in which the understanding emerged. However, I would like to challenge the idea that a motor skill is any less (or more) abstract or “in someone’s head” than cognitive understanding about a particular task, using the example of swimming.

Few people would disagree that swimming is a motor skill that depends just as much on the specifics of the environment as that it depends on a person’s capability to adjust to that in a very typical way. This specific environment is a pool of water large enough for a person to move about in. Swimming on land is, strictly speaking, impossible, because the resistance of air is much lower than the resistance of water, while a floor, on the other hand, is much too resistant. Furthermore, flapping your arms and legs in the air in a pattern that looks like swimming will not get you anywhere and is thus not functional (unless your aim was to make other people laugh). Swimming thus only exists and can be concretely defined in the water. In addition, learning to swim entails learning to coordinate many components of your body so that you stay afloat and move forwards or backwards while being in the water. When you have learned to swim, we expect you to be able to swim whenever you are in the water. However, when you are

not in the water and are thus not swimming, we do not think that you are not a skillful swimmer anymore. We typically do not ask “where your skill of swimming went”. No one considers it to be *abstract* or *in your head*, when you are not in the water.

Similar to swimming, cognitive understanding about a particular task only exists and can be concretely defined when a child is in a particular physical and social environment. For example, talking and gesturing about balance scale problems (see also Study 3 and 4, Chapter 4 and 5, respectively) only happens when a child is in a situation in which a balance scale and weights are present and an adult asks them to explain about balance scale problems. If a child would do a similar coordination pattern while playing hide and seek, this would give away their location, and would thus not be functional. Furthermore, having learned to correctly (from the perspective of the adult) explain about balance scale problems entails paying attention to, speaking, and gesturing about both mass of the weights and distance from the fulcrum whenever a child is in a situation that requires them to do so. This is thus similar to a skilled swimmer being able to swim whenever they are in the water. I therefore assert that asking “where the cognitive understanding about balance scale problems went” when a child is not in that particular situation is just as meaningful, or rather meaningless, as asking “where the skill of swimming went”.

One last counterargument, which is in favor of cognitive understanding being fundamentally different from motor skills, is that cognitive understanding about a particular task transfers to many other situations, while this is not the case for motor skills. However, this argument disregards that the ability to adaptively use a motor skill in an increasing number of diverse situations is inherent to learning a motor skill (e.g. Adolph, 2019; Adolph et al., 2018; Adolph & Hoch, 2019). With regard to the previous example of swimming, while children typically learn to swim in calm waters, such as a swimming pool, later on they will learn to swim in water with waves, or currents, such as in a sea or river. On the other hand, adverse circumstances, such as heavy clothing or stormy waters, will make swimming impossible for even the most skilled swimmers.

Moreover, cognitive understanding is known to be grounded and highly sensitive to environmental circumstances. I will illustrate this with the famous example of the A-not-B error (see Figure 6). The A-not-B error pertains to a classical Piagetian task, in which a toy is repeatedly hidden at a location A (the A-trials), where the child subsequently and correctly finds the toy. After a number of A-trials, the toy is hidden at location B. Children between 7 to 12 months old have been found to continue searching at location A, instead of location B. This has been coined as the A-not-B error (Piaget, 1954). Piaget attributed the error to the idea that children at that age have not yet developed the concept of *object permanence*. However, in a series of studies,

General Introduction

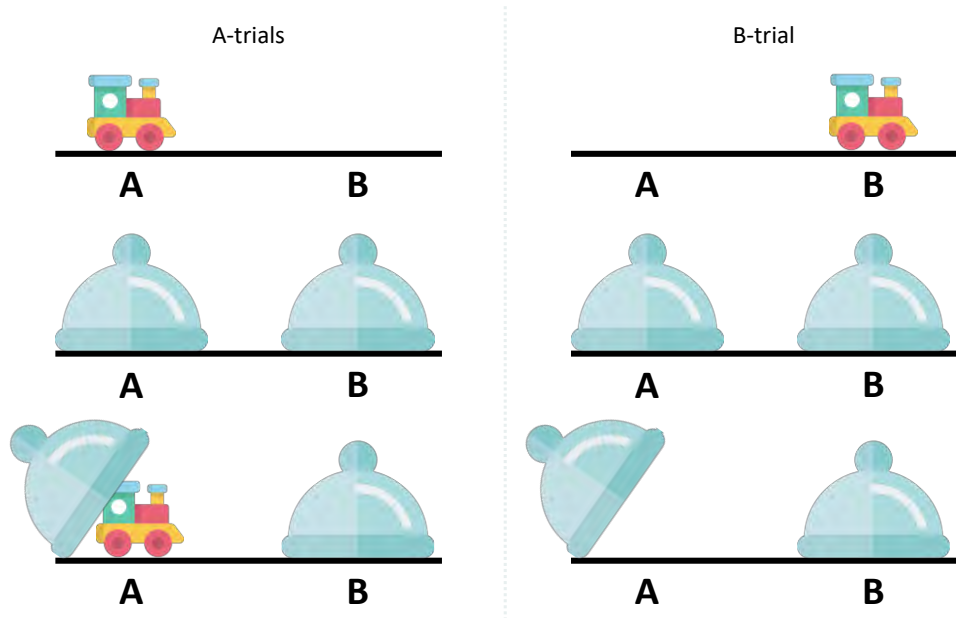


Figure 6. Visualization of the A-not-B-error task. In this task, initially a toy is repeatedly hidden at location A (the A-trials). The child correctly finds the toy at location A. After a number of A-trials, the toy is hidden at location B instead of location A (the B-trial). Yet the child still continues to search for the toy at location A.

inspired by complex dynamical systems theory, Smith et al., (1999), Spencer et al. (2001), Thelen et al. (2001), Schöner and Thelen (2006), and Schöner and Dineva (2007) showed that particular circumstances make the A-not-B error disappear in 10 month old children, while other circumstances elicit the A-not-B error in older children. To be specific, a salient visual difference between the locations, as well as a change in posture (i.e. sitting vs standing) made younger children correctly search at location B during B-trials (Smith et al., 1999), for example. Furthermore, a longer waiting time between hiding the toy at location B and searching for the toy elicited the A-not-B error in children who were older than 12 months (Spencer et al., 2001). This example of the A-not-B error again shows that the theoretical perspectives of complex dynamical systems, coordination dynamics, and affordances are just as useful for capturing cognitive development as they are for capturing motor development.

This dissertation

I started this General Introduction with describing that hand movements in general, and gestures in specific, have been found to lead cognitive development in children, over speech. Furthermore, I highlighted that a satisfying explanation for this phenomenon has been lacking, both from the perspective of gestures' and speech's typically tight integration, as well as from

the theoretical perspectives of complex dynamical systems, coordination dynamics, and affordances. Throughout the General Introduction, I showed how these three theoretical perspectives are very powerful in explaining a broad range of phenomena in many diverse systems and organisms, including human behavior, (cognitive) development and skill acquisition. As described before, my goal in this dissertation, based on these theoretical perspectives, is to understand how cognitive development is related to how children move their hands and how they speak during cognitive tasks -over time and at multiple scales-, and how their hand movements and speech relate to each other, and to the physical and social environment. By researching cognitive development in children, I will move beyond the topics which traditionally have been investigated from these perspectives, such as early motor development, and motor coordination, hereby following the footsteps of many inspiring researchers before me (e.g. Stephen et al., 2009; Thelen & Smith, 2007; Van Geert, 2019). Together with my supervisors and several collaborators I carried out four studies.

In Study 1 (Chapter 2) we investigated the **stability and variability of the coupling between children's gestures and speech**, in terms of level of understanding during a hands-on Science & Technology task, which children did together with an adult who provided support. We also investigated how these within-task measures of gesture-speech coupling predicted general measures of cognitive performance.

In Study 2 (Chapter 3) we investigated students' **gesture-speech synchronization** in an easy and a difficult cognitive task. We specifically researched gesture-speech synchronization in terms of temporal alignment (phase synchronization), semantic similarity (gesture-speech mismatches), and complexity matching (multiscale synchronization).

In Study 3 (Chapter 4) we investigated how children performed hands-on Science & Technology tasks with **different spatiotemporal** properties, and how these different properties of the environment were related to **differences between children's variability of hand movements and speech**. We conceptualized variability in terms of Diversity and Complexity.

In Study 4 (Chapter 5) we investigated how **dyads of children coordinate their speech, hand movements and head movements**, when they solve cognitive problems **together**. We researched the coherence and relative phase angle (which informs about leader- and follower-patterns, and in- and anti-phase coordination) of dyads' speech, hand movements and head movements at multiple timescales, and analyzed whether these measures predicted task performance and dyadic agreement.

In the **General Discussion** (Chapter 6) I will discuss what these studies have contributed to my aim of understanding of how children move their hands and speak during cognitive tasks, and how their hand movements and speech relate to each other, to the physical and social

General Introduction

environment, and to cognitive development. Furthermore, I will address what our findings mean in light of the more theoretical backgrounds of complex systems, coordination dynamics, and affordances. Lastly, I will discuss what our findings mean for educating children in primary education and inducing cognitive development.



2

Asymmetric dynamic attunement of speech and gestures in the construction of children's understanding

This chapter is based on:

De Jonge-Hoekstra, L., Van der Steen, S., Van Geert, P., & Cox, R.F.A. (2016). Asymmetric Dynamic Attunement of Speech and Gestures in the Construction of Children's Understanding. *Frontiers of Psychology*, 7:473. doi: 10.3389/fpsyg.2016.00473

Asymmetric dynamic attunement of speech and gestures in the construction of children's understanding

How do children learn and develop understanding? How does cognitive change arise? In developmental psychology, this is one of the most intriguing questions, as evidenced by the considerable literature on the topic (see for instance, Anderson et al., 2012; Carey & Spelke, 1994; Gelman, 2004; Perry et al., 1988; Piaget, 1952; Siegler, 1989; Sternberg, 1984; Thelen, 2000; Van Der Steen et al., 2014; Vygotsky, 1994). In search for the mechanisms behind cognitive development, the hands of children have come up as a vital ingredient. As children learn new things, or when they communicate or explain things, they use both their speech for verbal utterances and their hands to gesture (Alibali & Nathan, 2012; Anastas, Stephen, & Dixon, 2011; Goldin-Meadow, Wein, & Chang, 1992).

Gestures and speech are coupled, and mostly they are well aligned, such that meaning expressed in gestures matches that expressed in speech. However, sometimes gestures and speech do not overlap, and a so-called gesture-speech mismatch occurs (Church & Goldin-Meadow, 1986; Goldin-Meadow, 2003; Perry et al., 1992). It has been demonstrated that during such gesture-speech mismatches, people (children and adults) express their cognitive understanding in gestures before they are able to put them into words (Crowder & Newman, 1993; Garber & Goldin-Meadow, 2002; Gershkoff-Stowe & Smith, 1997). Gesture-speech mismatches are especially likely to occur when a person is on the verge of learning something new. This makes them a hallmark of cognitive development (Goldin-Meadow, 2003; Perry et al., 1992), and shows that gestures and cognition are coupled as well. In the literature the explanation for this link has been attributed to gestures being a medium to express arising cognitive strategies (Goldin-Meadow et al., 1993), to highlight cognitively relevant aspects (Goldin-Meadow et al., 2012), to add action information to existing mental representations (Beilock & Goldin-Meadow, 2010), to simulate actions (Hostetter & Alibali, 2010), to decrease cognitive load during tasks (Goldin-Meadow, Nusbaum, Kelly, & Wagner, 2001) and to construct cognitive insight (Boncoddio et al., 2010; Stephen et al., 2009; Stephen et al., 2009; Trudeau & Dixon, 2007).

A conceptual framework which has been largely ignored in the research on gestures, and which follows from the work by Iverson and Thelen (1999), is that of synergetics and self-organization dynamics introduced by Haken (1977/1983), Kelso (1995), and Kugler and Turvey (1987). First of all, at the behavioral level, gestures and speech are considered to be action systems (Reed, 1982) That is, they are functional units organized to perform a specific task, like a hands-on science task in the present study. In addition, at the coordination level, we argue that gestures and speech form two coupled synergies. Within the context of action control, a synergy is a

Asymmetric dynamic attunement of speech and gestures

temporarily stable task-specific collective organization (Kelso, 1995), which emerges through self-organization out of a large set of underlying components distributed across body, brain and environment.

To elaborate, gestures and speech require the precise coordination of many different muscles, joints, neurons, as well as related perceptual subsystems. Speech articulation, even for the simplest utterances, involves well over 70 muscles in the respiratory, laryngeal ('voice box') and pharyngeal (throat) systems as well as of the mouth, the tongue, etcetera (Galantucci, Fowler & Turvey, 2006; Turvey, 2007). Moreover, speech is highly attuned, for instance, to auditory information, but also to vision (needed for e.g., interpersonal communication). Gesturing results from the coordinated contractions of tens of muscles in the shoulder, upper arm, forearm, hand and fingers of both upper limbs (Weiss & Flanders, 2004), and involves a tight informational link to proprioceptive as well as visual subsystems to stay attuned to the environment. Synergies for speech and gestures consist of several (overlapping) neural structures involved in information-motor couplings, across the central nervous system. Cognitive subsystems loosely associated with attention, memory and the planning of movements will play a role in gestures as well as in speech. Importantly, the gesture and speech synergies share several of these underlying components, and their recruitment will temporally overlap in any given task (cf. Wijnants, Cox, Hasselman, Bosman & Van Orden, 2012).

During communication or the expression of thoughts and ideas, the gesture and speech synergies synchronize to a high degree (McNeill, 1992). This synchronization reflects that the self-organizing process underlying the creation of both synergies is able to recruit the underlying components in the service of both gestures and speech adequately and synchronously. In fact, because of the tight coupling of the gesture and speech synergies, trying not to use either gestures or speech while communicating, or to desynchronize them, proves to be detrimental for the other (Goldin-Meadow, Cook & Mitchell, 2009). Moreover, Goldin-Meadow et al. (2001) found that if children or adults do not gesture -either by instruction or by choice- while they explain how they solved a mathematical problem, they perform worse on recalling a list of words or letters that they had to remember while they explained the mathematical problem. Goldin-Meadow et al. (2001, p. 521) conclude that "...gestures and speech form an integrated and, indeed, synergistic system in which effort expended in one modality can lighten the load on the system as a whole".

From the perspective of synergetics and self-organization dynamics, the decline in performance if one only speaks but does not gesture should be related to suboptimal coordination of the gesture and speech synergies. More generally, when demands on the action systems increase, such as, for instance, in a novel or challenging task, the synergies become relatively less stable

and less synchronized as compared to less challenging tasks. Novel and challenging tasks often have several new and (seemingly) conflicting task constraints. Since synergies are task specific, different task constraints lead to different collective organizations, competing for existence and the recruitment of (shared) components. Following Wijnants et al. (2012), who studied synergetic control under conflicting task constraints in the context of a Fitts task, we reason that the gesture-speech mismatch in a novel task (Goldin-Meadow, 2003) resides in a less optimal simultaneous organization and coordination of the gesture and speech synergies. As a result, the usually tightly coupled synergies of gestures and speech dissociate, due to overlapping recruitment of the underlying components involved, resulting in the observable gesture-speech mismatch. Consequently, a gesture-speech mismatch can take different forms, such as instances in which gestures convey different content than speech, in which there are only gestures but no speech, and in which there is only speech but no gestures, similar to what Goldin-Meadow et al. (2001) found.

Most studies examining the gesture-speech mismatch have thus far focused on series of problem solving events in which, across different trials with some time in between, children are asked to solve a certain problem and explain their solution. These studies have focused on children's solutions to, for instance, a series of mathematical equivalence problems (Alibali & Goldin-Meadow, 1993b), Tower of Hanoi-problems (Garber & Goldin-Meadow, 2002), conservation tasks (Goldin-Meadow et al., 1993), and gear solving tasks (Boncoddo et al., 2010). From these studies, it appears that children show new problem solving strategies by means of gestures in earlier trials, to be followed by speech one or multiple trials later. A more detailed understanding of how such patterns of gestures and speech arise, and how this relates to our proposal of suboptimal coordination of synergies and cognitive development, requires a study of children's verbal and nonverbal behaviors as they occur in real time (Pine et al., 2007), that is, during a task, considering their temporal order and coupling. The current study investigates the nonlinear, dynamic interplay of children's gestures and speech as they construct their cognitive understanding during a hands-on science task. Analysis tools will be employed which allow us to quantify the process of dynamic attunement between speech and gestures across all possible time scales during the task.

The current focus on the coupled dynamics of gestures and speech as it occurs in the moment and across time scales resonates with the relatively recent call for microgenetic studies to investigate the process (rather than just the outcome) of cognitive development (e.g. Cox & Van Dijk, 2013; Flynn et al., 2007; Grannot & Parziale, 2002; Siegler, 2006; Van der Steen et al., 2012). These microdevelopmental studies are exponents of the complex dynamical systems approach to behavior, cognition, and development (Smith & Thelen, 2003; Van Geert, 1998, 2011). This approach aims to infer the "why" and "how" of development (Thelen & Corbetta, 2002), using

Asymmetric dynamic attunement of speech and gestures

the language of complex dynamical systems: multi-causality, self-organization, variability, stability, non-linearity and so on, and the accompanying data-analytical tools.

To explain these terms in short, multi-causality pertains to the notion that development cannot be ascribed to one component or level of the developing system, but instead emerges from the continuous interaction of all the levels of the developing system (Thelen & Smith, 2007). Self-organization means that patterns and order emerge from the continuous interaction of all levels of the developing system, without external interference. Variability and stability follow from self-organization, as both variable and stable behavior occur within a developing system. For new stable behavior, i.e., new patterns, to emerge, a system typically displays variable behavior before settling in a new, more stable, pattern. Variability is thus a hallmark of developmental change. Moreover, this indicates that development is inherently non-linear, with periods of stable and variable behavior (Van Geert, 2008). Multicausality, self-organization and variability are also mechanisms that are apparent in our proposal that diverse components coordinate to form the synergies of gestures and speech, and that the dynamics within and between the synergies, under certain conditions, result in gesture-speech mismatches.

Dynamic skill theory is a theory of cognitive development encompassing dynamical system principles (Van Geert & Fischer, 2009). It provides a model that allows researchers to structurally investigate processes of cognitive development (Fischer, 1980; Fischer & Bidell, 2006). Dynamic skill theory states that the development of cognitive skills —defined as actions and thinking abilities, which includes verbalizations and gestures — proceeds through a series of hierarchically, ordered levels. That is, the development of cognitive skills follows a structure in which higher-order skills are constructed of a combination of skills at lower levels. According to dynamic skill theory, skills develop through a series of ten levels, divided over three tiers, although not in a simple linear fashion (see below). The first tier is the sensorimotor tier, which consists of perceptions, actions and observable relations between these perceptions and actions. The second representational tier goes beyond the observable relations between actions and perceptions, although still restrained to concrete situations. The last tier, abstractions, includes non-concrete rules that apply in general (Schwartz & Fischer, 2005). Each tier consists of three levels, single sets, mappings (relations between single sets), and systems (relations between mappings).

In accordance with the notion of nested timescales, which implies that development occurs at different, though tightly interconnected timescales, the levels as distinguished by dynamic skill theory are applicable to both macro (long term) and micro (short term) development (Fischer & Bidell, 2006; Schwartz & Fischer, 2004). This means that people also go through these levels on the short-term time scale, for example during a new task, in a nonlinear fashion, so that

drops, spurts and stable periods in understanding occur (Van der Steen et al., 2012). This makes this theory particularly suitable for detailed, within-task dynamical analyses. Furthermore, dynamic skill theory provides a structure in which the concepts expressed in and constructed by gestures and speech can be compared, as it can be applied to both actions and verbalizations (Granott et al., 2002; Hoekstra, 2012). Lastly, dynamic skill theory's model can grasp meaningful intra-individual variability on the short term timescale, by allowing for fluctuations in cognitive understanding during a single task, as well as the (sometimes differing) levels displayed by gestures and speech. This intra-individual variability has been linked to learning and transitioning to a higher (cognitive) level (Goldin-Meadow, 2003; Schwartz & Fischer, 2004; Siegler, 2007; Van Geert & Steenbeek, 2005; Van Geert & Van Dijk, 2002; Yan & Fischer, 2002). Although it has never been studied explicitly, understanding at the level of the sensorimotor tier might lead to a different interplay of gestures and speech, compared to understanding at the level of the representational tier.

As learning is an inherently nonlinear process (Van Geert, 2008), and intra-individual variability in cognitive understanding and strategies is a hallmark of transitioning to more advanced levels, non-linear time-series methods are needed to investigate these processes. One such method is Recurrence Quantification Analysis (RQA; Marwan et al., 2007; Webber Jr. & Zbilut, 2005). RQA originates from the study of natural systems, and has recently been applied to the study of human behavior and development (e.g., Aßmann et al., 2007; Shockley et al., 2002; Wijnants et al., 2009; 2012). RQA is based on the detection and quantification of recurrent (i.e. repeatedly occurring) behavioral states, one of the most fundamental and important properties of dynamic systems. By using RQA and the notion of recurrence, measures of interest in a dynamic analysis of the behavior of a system, such as stability, regularity, and complexity can be retrieved from the time series. For a full overview of the RQA method, see the paper by Marwan et al. (2007), and for a useful guide to applying it see the chapter by Webber and Zbilut (2005).

A methodological advancement of RQA, *Cross*-Recurrence Quantification Analysis (CRQA; Marwan et al., 2007; Shockley et al., 2002; Zbilut et al., 1998) will be used in this paper to study the interplay of gestures and speech. With CRQA, the shared dynamics of two coupled systems, such as, for instance, parent-child dyads (Cox & Van Dijk, 2013; Dale & Spivey, 2006; De Graag et al., 2012; Lichtwarck-Aschoff et al., 2012), staff-client dyads (Reuzel et al., 2013, 2014) and adult dyads (Louwerse et al., 2012; Richardson & Dale, 2005; Richardson et al., 2007; Shockley et al., 2003) can be studied. In CRQA, recurrence is generally defined as some match of behavioral state in the two systems under study. In RQA and CRQA alike, recurrence is not confined to states at exactly the same moment, but it is also noted when these particular matching states occur in the systems at either an earlier or later point in time, in fact across all possible time scales. These time scales range from the smallest time scale of the sample rate

Asymmetric dynamic attunement of speech and gestures

(seconds), to the duration of the entire observation. Linear tools fall short to fully capture the underlying dynamics of the cognitive system, which is fundamentally non-stationary and nonlinear, as well as continuously attuning to a changing environment. Recurrences of system trajectories, on the other hand, can provide important clues as to the system from which they derive, in this case, the cognitive system (cf. Marwan & Webber, 2014).

To summarize, children's use of gestures and speech is known to be informative about their cognitive capabilities, which change on a developmental time scale (Goldin-Meadow, 1998). As we have argued above, synergetic control and synergetic competition form a valuable explanatory framework for this research topic, which might lead to novel insights. As synergies are reflected in the dynamic organization of behavior (cf. Stephen et al., 2009), we will analyze children's gestures and speech as they construct understanding in real time. To this end, CRQA will be applied to the two time series of skill levels (based on dynamic skill theory) displayed in children's gestures and speech, while they are working on an educational science task. The main research question of this study is: How is the leading role of gestures over speech in children's cognitive change, as reported in previous studies, related to and reflective of an underlying dynamic interplay between gestures and speech during task performance? Research outcomes will pertain to the dynamic attunement of gestures and speech, focusing, for instance, on their temporal relation, leader-follower hierarchy, and asymmetric coupling. Furthermore, the dynamic interplay between gestures and speech during task performance will be related to age and more general measures of performance outside the task. Specific research questions, hypotheses, and their rationale will be given after a more detailed introduction of recurrence procedures and the derived measures of dynamic organization in the Method section.

Materials and methods

Participants

For this study, the data of 12 Dutch children, six boys and six girls, were analyzed. The participants took part in a larger longitudinal project (see Van der Steen, 2014), and were on average 39.1 months old ($SD = 3.8$) at the start of the longitudinal data collection. In this larger study, children individually worked on scientific tasks about air pressure and gravity, under guided supervision of a researcher, in four-month intervals. All children were recruited at their daycare centers or (pre)schools by asking their parents for a written consent. Parents were told about the nature of the study (children's longitudinal development of scientific understanding), but not about the specific tasks that were administered. The study was approved by the ethical committee of the Psychology Department of the University of Groningen.

For the current study, we chose to analyze children's (non)verbal behavior during an air pressure task administered at the sixth measurement (see below). We chose this task because the task protocol gradually builds up to a wrap-up question in which children are able to show their understanding of the task at that point. Our sample included five children from kindergarten ($M = 57.2$ months, $SD = 2.2$ months), and seven children from first grade ($M = 69.4$ months, $SD = 4.4$ months). Table 1 gives an overview of characteristics of each child, including children's early math- and language-scores on standardized tests from a national pupil-monitoring system that the children performed in kindergarten. These tests are administered twice a year to keep track of primary school children's progress on the subjects math and (Dutch) language. For the Kindergarten tests, children are asked to count, classify objects and phrase words. Scores can range from 1 to 5, with 1 as the lowest and 5 as the highest attainable score. In addition, Table 1 provides children's average skill level score during the past five measurements, as measured in their verbalizations.

Procedure

During the task, researcher and child were involved in a natural hands-on teaching-learning interaction. An adaptive protocol was constructed, which guaranteed that all children were asked the basic questions reflecting the core building blocks of the task and the incorporated scientific concepts (see Van der Steen et al., 2012 for an excerpt of an interaction). At the same time, the protocol left enough space for children to take initiative and manipulate the material. The researcher started by showing the task material to the child, asking about its purpose and

Table 1

Overview of characteristics of the 12 participating children.

Child	Grade	Age (months)	Math-score	Language- score	Average score past tasks
1	KG	58	5	-	2.65
2	KG	55	5	5	2.27
3	KG	60	2	3	0.77
4	KG	58	5	5	2.55
5	KG	55	5	4	2.45
6	1	64	4	5	2.31
7	1	64	5	5	2.56
8	1	69	4	4	2.42
9	1	76	4	4	2.27
10	1	69	3	3	1.98
11	1	73	4	4	2.75
12	1	71	5	5	2.79
<i>Mean</i>	-	<i>64.3</i>	<i>4.25</i>	<i>4.27</i>	<i>2.32</i>

Asymmetric dynamic attunement of speech and gestures

functioning. The child was then encouraged to explore the material, while the researcher asked questions, such as “What do you think we should use this for?” Furthermore, the researcher was allowed to provide guidance by asking follow-up questions, encouraging the child to try out his/her ideas using the material, and by summarizing the child’s findings or previous answers. The guidance never included statements indicating whether the child was right or wrong. We analyzed the interaction until the child answered a ‘wrap-up’ question (“After investigating all of this, can you now explain how this device works?”), after which the protocol prescribed the researcher to start with another topic. This part of the interaction (from the first question until the ‘wrap-up’ question) took 5 to 12 minutes (on average a little over 8 minutes). All interactions took place within children’s schools, always guided by the same researcher, and were recorded on video.

Materials

The task explored was called the “air canon”, specifically designed for this study. It was designed to let children explore how air pressure can be used to set materials in motion, and how air can be temporary stored in a balloon and released to have an even bigger impact on objects. The task consisted of wood, garden sprinkler parts, a transparent drainage tube, a gutter made from part of a room divider, a ball pump, balloon, and ping-pong balls (see Figure 1). There are three (sprinkler) taps on this device, one to (dis)connect the air pump, one to (dis)connect the balloon, and one to (dis)connect the drainage tube. Through questioning and exploring, children realize they have to open some taps (and close others) to make the canon work. There



Figure 1. The “air canon” and a close-up of the pump mechanism of this task.

are two ways to shoot a ping-pong ball down the tube: 1) simply opening the taps connected to the pump and tube (closing the tap to the balloon), and repeatedly pumping, and 2) by inflating the balloon first (closing the tap to the tube), and then releasing the air into the tube. The colors on the wood serve as a measuring device to see how far the ball goes.

Analysis

Coding procedure

The interactions were first coded for children's verbal utterances, and then for gestures/task manipulations. Both coding systems are described in more detail in Appendix A. The verbal utterances were coded in four steps using the computer program MediaCoder (Bos & Steenbeek, 2006). We started with the determination of the exact points in time when children's utterances started and ended. The second step involved the classification of these verbal utterances into categories (e.g., description, prediction, explanation). As a third step, meaningful units of the child's coherent task-related utterances were formed, so that utterances (sentences) about the same topic with only a short break in between were joined together for the fourth step. In this fourth and final step, the complexity of the child's verbalized understanding within a unit was determined, using a scale based on Dynamic skill theory. The dynamic skill levels ranged from the levels of the sensorimotor tier to single abstractions, with levels of the representational tier in between. For example, at the first level of the sensorimotor tier (level 1), the child states a single characteristic of the task, such as "This tube is long". At the first level of the abstract tier (level 7), the child mentions an abstraction that goes beyond the material, for example a statement about air pressure in general. This range of levels (1-7) approximately corresponds to the attainable levels for the children's age (see Fischer & Bidell, 2006). Only utterances that displayed correct characteristics or possible task operations or mechanisms were coded as a skill level. This verbal coding procedure is explained in more detail elsewhere (Van Der Steen et al., 2013; Van Der Steen et al., 2014).

In order to make sure that the codes of verbal utterances were reliable, a standardized codebook was used. For each step of coding, three raters went through a training of coding three video fragments of fifteen minutes and compared their codes with those of an expert-rater (who constructed the codebook and training). The codes of the third fragment were compared to the codes of the expert-rater and a percentage of agreement was calculated. The reliability of the percentage of agreement is based on Monte Carlo permutation testing. The codes of one of the raters were shuffled 1000 times, so that the order of the codes became random. The p -value is the amount of times that the percentage of agreement of the shuffled codes was the same (or higher) as the empirical percentage of agreement, divided by the times that the codes were shuffled (1000). On average, the empirical percentage of agreement was:

Asymmetric dynamic attunement of speech and gestures

Categories: 87% (range 81-93; $p < .01$), combining verbalizations into units: 93% (range 89- 96; $p < .01$), and level of understanding: 90% (range 83-95; $p < .01$).

The child's gestures and task manipulations (hereafter: gestures) were coded independently from the verbal utterances. The coding procedure for gestures also involved multiple steps. During the first step, the exact point in time when a gesture started and ended was determined, along with a broad categorization of the gesture into the categories *short answers*, *representations/manipulations*, and *emblems* (such as "thumbs up"). For the second step, the broad categories of the first step were refined to more specific categories. For example, *short answers* were allocated to *nodding yes*, *shaking no*, etc., *representations/manipulations* were split into *characteristic* (such as representing 'hard'), *movement* (such as representing 'fast', or the course of a ball), *representation* (such as representing relations among different objects), while *emblems* were kept undifferentiated. The third and last step involved assigning levels of complexity, based on Dynamic skill theory (similar to how the verbal utterances were coded), to all *representations/ manipulations*. For more details about the gesture codebook, see Appendix A, and Hoekstra (2012).

To ensure reliable coding of children's gestures, two raters coded four training video fragments of ten minutes independently, while following the standardized codebook, and their percentages of agreement were calculated for each step of coding. The reliability of the percentages of agreement was based on Monte Carlo permutation testing, like for the coding procedure for verbal utterances. On average, the percentages of agreement was: 97% (range 94-100; $p < .01$) for the first step (broad categorization), 86% (range: 78-91; $p < .01$) for the second step (refined categories), and 92% (range: 88-98; $p < .01$) for the third step (level of complexity).

Time series

Before performing CRQA on the data, the codes of the video fragments were transformed into a time series of the skill levels of speech, and a time series of the skill levels of gestures, with a sample rate of 1 second. If there was no event (i.e., no skill level), this was indicated with a 0 in the time series. In Figure 2, the time series of skill levels of gestures and skill levels of speech of one of the children in our sample is depicted. In order to be able to distinguish the lines in Figure 2 clearly, only the first 300 seconds of the 392 seconds in total are displayed.

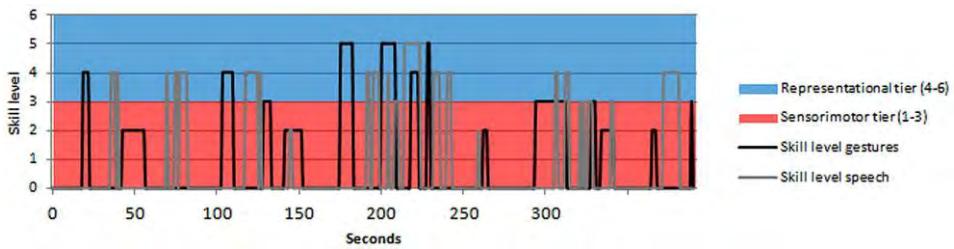


Figure 2. Time series of skill levels of gestures and speech of one child (boy, 55 months).

Cross Recurrence Quantification Analysis

For categorical data, CRQA starts by plotting in a plane (called the cross recurrence plot, CRP, see Figure 3) all congruent appearances of some pre-specified matching values within a pair of time series, by putting one of the time series along the horizontal axis and the other along the vertical axis. Specifically, the CRP represents all those instances when the behavioral state of one subsystem (e.g., skill level in verbalization) at some moment in time is matched by the behavioral state of another subsystem (e.g., skill level in gesture) at the same or any other moment in time during the observation. These instances are depicted as colored dots in the CRP, which are canonically referred to as ‘recurrent points’. From the spatial layout of these colored dots, several recurrence measures can be derived (see below). These CRQA-measures reveal hidden structure concealed in the shared dynamics of the two interaction subsystems (speech and gestures) across all possible time scales, which is informative about the dynamic organization of the cognitive system. Figure 3 illustrates the CRP of gestures and speech for the same child as the time series in Figure 2. The CRPs of the other children are available as supplementary materials. In this study, matching states (i.e. recurrent points) are defined as same-tier skill levels, and are color-coded in the CRP as follows: Blue dots represent instances in which gestures and speech both display a skill level from the sensorimotor tier (i.e. skill level 1, 2 or 3). Red dots represent instances in which the skill levels as displayed by gestures and speech are both from the representational tier (i.e. skill level 4, 5 or 6). Finally, yellow dots in the CRP represent a gesture-speech recurrence of the highest, abstract tier (i.e. skill level 7). The latter did not occur in our sample and these recurrences will therefore not appear in the analysis.

In Figure 3, the green diagonal line is the Line Of Synchrony (LOS), on which recurrent points have a delay of zero seconds. These represent instances when both speech and gestures display a skill level from the same tier at the *exact* same time. The percentage of recurrent points on this line is called the percentage of synchrony (%Sync), which is a measure of linear static synchrony of the two subsystems. The Recurrence Rate (RR) is a measure depicting the proportion of recurrent points in the entire CRP. Hence, RR reflects the extent to which

Asymmetric dynamic attunement of speech and gestures

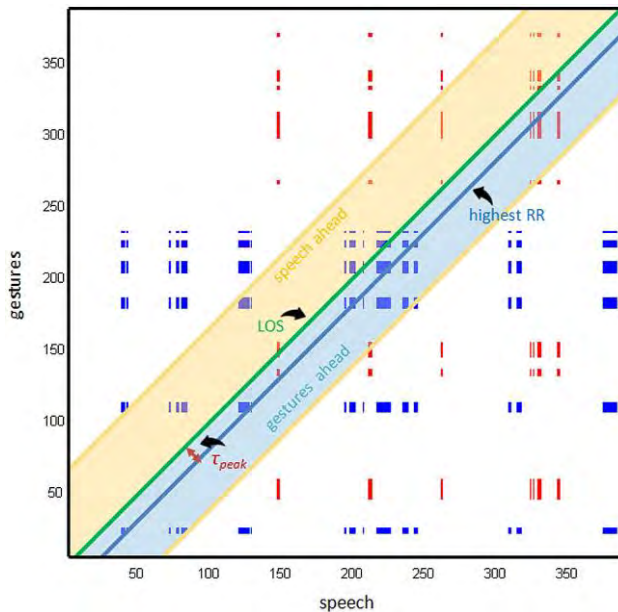


Figure 3. Cross Recurrence Plot (CRP) of one child (boy, 55 months).

behaviors of one subsystem are matched by those of the other subsystem across all possible time scales, from the high end determined by the sample rate of 1 second, up until the low end determined by the duration of the observation. As such, RR is a basic measure of the coupling and coordination of the two subsystems. In the CRP of Figure 3, the skill-level time series of gestures is plotted on the vertical axis and the skill-level time series of verbalizations on the horizontal axis. This means that all colored dots above the LOS represent instances in which a skill level expressed in speech earlier in time is matched by same-tier skill level expressed in gestures at a later moment. Congruously, colored dots below the LOS represent instances in which skill levels from the same tier are displayed by gestures at an earlier moment and matched by speech later.

As can be seen in Figure 3, most colored dots in the CRP align to form block and line structures. Generally, such structures indicate instances where behaviors which are briefly expressed by one subsystem are accompanied by episodes of lingering in the matching behavior by the other subsystem. This provides information about the shared dynamics of the gesture-speech interaction, and specifically about the strength and direction of the coupling between the two subsystems, as we shall demonstrate (see Cox et al., 2016). Thus far, research using CRQA has focused on diagonal and vertical lines. However, notice how the line structures in the CRP stretch into the horizontal and vertical direction (and not diagonal), which is quite common for

categorical time series. Analysis of the diagonal lines and the associated measures will therefore not be discussed here.

The different directions of the line structures (vertical and horizontal) provide differential and complementary information about the coupling between the two subsystems represented by the time series along the axes. For instance, a vertical line structure in the CRP (Figure 3) means that a brief skill-level expression in speech is followed (above LOS) or preceded (below LOS), with some delay, by a much longer same-tier skill level expression in gestures. Similarly, horizontal line structures represent instances in which a skill level that is expressed briefly in gestures, is followed (below LOS) or preceded (above LOS) by a much longer same-tier skill level in speech. More generally, line structures represent instances in which shortly expressed skill levels from a certain tier in one subsystem 'trapped' the other subsystem in a lingering same-tier expression for some time. In this study we will relate them to the relative strength and direction of the gesture-speech coupling, such that vertical line structures reflect the extent to which speech subsystems influence gestures, whereas horizontal line structures reflect the extent to which gestures subsystems influence speech.

To capture the asymmetric dynamic attunement between gestures and speech, we performed *anisotropic* CRQA (Cox et al., 2016), by calculating recurrence measures for the horizontal and vertical line structures separately and comparing them. The first measure derived from the line structures is 'Laminarity', defined as the proportion of recurrent points that are part of a vertical (LAM_V) or horizontal (LAM_H) line structure. Laminarity reflects the degree to which subsystems are trapped into expressing a same-tier skill level for some period of time. LAM_V depicts how much gestures constitute larger structures of points in the CRP, whereas LAM_H does so for speech. Second, 'Trapping Time' is the average length of either the vertical (TT_V) or horizontal (TT_H) line structures. TT is measured in units of time and estimates how long subsystems are, on average, trapped in a specific state. In our study, the higher TT is, the longer a same-tier skill level from one time serie lingers in the other one. If TT_V is high, gestures tend to be trapped in relatively long periods of same-tier skill levels that are also expressed by speech at some point, and for high TT_H speech tends to be trapped in relatively long periods of same-tier skill levels that are also expressed by gestures at some point. Finally, 'Maximum Line' also gives information about duration of line structures, with $MaxL_V$ the length of the longest vertical line and $MaxL_H$ the length of the longest horizontal line. In other words, $MaxL$ measures the duration of the longest same-tier skill-level expression for speech and gestures. High $MaxL_V$ means that gestures are trapped in a single tier of skill levels, and $MaxL_H$ means that speech is trapped strongly in a single tier.

Asymmetric dynamic attunement of speech and gestures

These three measures have been related to behavioral rigidity and regularity in previous studies (Cox & Van Dijk, 2013; De Graag et al., 2012). Accordingly, in the present study, we will interpret the CRQA-measures of horizontal and vertical line structures as ‘differential’ rigidity of speech and gestures, respectively. In addition, the relative size of these measures informs about the relative strength and direction of the coupling between speech and gestures.

LOS-profile analysis

Besides analyzing the global structure of the recurrence plot, we will also look in more detail at several recurrence measures within a smaller time window around the line of synchrony (LOS; see e.g. Reuzel et al., 2013; 2014; Richardson & Dale, 2005). Figure 4 depicts the so-called LOS profile of an interval of 60 seconds on each side of the LOS, derived from the CRP in Figure 3. The LOS profiles of the other children are available as supplementary materials. The interval of 60 seconds above and below the LOS is chosen intuitively, so as speech and gestures can either lead or follow each other with a maximum delay of one minute. In Figure 4, the position of the LOS, corresponding to a delay of zero seconds, is indicated with a green line. The LOS profile is drawn ‘from the perspective’ of gestures, in that a positive delay indicates instances of recurrence in which gestures are ahead of speech in time (blue area), whereas a negative delay indicates instances in which speech is ahead of gestures (yellow area). The orange envelope curve represents the Recurrence Rate at each delay; this delay is called τ (RR_{τ} ; see e.g. Marwan et al., 2007).

Several measures can be derived from this LOS profile, which inform about the coordination of the two subsystems within the chosen interval of two minutes around the LOS. Firstly, in Figure 4 the RR shows a clear peak of around 0.09 at a delay of 16 seconds. This maximum recurrence rate, defined as the highest proportion of recurrent points within the LOS profile, is called RR_{peak} , and is indicated with the blue line in Figure 4. The distance of this peak from the line of synchrony (in seconds), or in other words, the delay of RR_{peak} , is called τ_{peak} , and is indicated with the red arrows. Please note that τ_{peak} with a value of 16 seconds, is also visible in Figure 2,

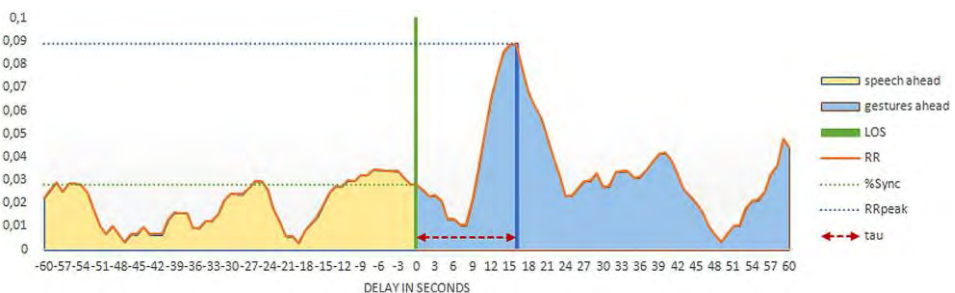


Figure 4. LOS (line of synchrony) profile plot of one child (boy, 55 months).

as the skill levels displayed in gestures are clearly ahead in time of the skill levels displayed in speech. An example of what a match between gestures and speech with a delay of 16 seconds could be is: With his hands, a boy depicts that if you turn a switch, the ball will roll down the tube (level 3, tier 1). Around 16 seconds later, he says: "It [the ball] rolls, because it is round" (level 3, tier 1). The final measure that we can derive from the LOS profile is Q_{LOS} . Q_{LOS} is the total proportion of recurrent points at the left side of the LOS (yellow area), divided by the total proportion of recurrent points at the right side of the LOS (blue area). If Q_{LOS} is lower than 1, this indicates that gestures are generally leading speech in time, whereas a Q_{LOS} with a value higher than 1 indicates the opposite.

Research questions and hypotheses

The research question of the current study is: Does the leading role of gestures over speech in children's cognitive change, as reported in previous studies, arise from and reflect an underlying dynamic interplay between gestures and speech during task performance? To answer this general question, four specific research CRQA questions and corresponding hypotheses were formulated, which will be introduced below.

Research Question 1

The first research question is: What is the temporal relation between gestures and speech, with regard to the displayed (skill) level of understanding? Studies thus far demonstrated that, across tasks, children express their cognitive insights in gestures before they are able to put them into words (Crowder & Newman, 1993; Garber & Goldin-Meadow, 2002; Gershkoff-Stowe & Smith, 1997). Here we will investigate whether these results can be extrapolated to a smaller (i.e. within-task) time scale, and whether theoretical claims of previous studies can be corroborated and possibly extended to the perspective of gesture-speech mismatches as originating from the suboptimal simultaneous coordination of the gestures- and speech synergies. To this end we performed LOS-profile analysis on the gesture-speech interaction. The associated measures should display a significant asymmetry in the amount of recurrence around the LOS (Q_{LOS}) and display a recurrence peak (RR_{peak}) at some delay (τ_{peak}) in the blue area of children's LOS profile (see Figure 4), indicating a leading role of gestures on speech.

Research Question 2

The second research question is: What is the relative strength and direction of the interaction coupling between the gesture and speech subsystems? For this we looked at LAM, TT, and MaxL for both vertical and horizontal line structures, across the entire CRP. The mutual, ongoing, possibly asymmetric influence between gestures and speech will be visible in the CRP by the isentropic patterns of colored line structures representing same-tier skill levels. Accordingly, we

Asymmetric dynamic attunement of speech and gestures

expect vertical and horizontal LAM, TT and MaxL, and especially their differences, to inform us about the coupled dynamics of gestures and speech, and its potential asymmetry with regard to strength and direction.

Research Question 3

The third research question is closely related to the second, but focused on the specific skill-level tiers: What is the relative strength and direction of the interaction between gestures and speech *for the different levels of understanding* (i.e. skill-level tiers)? To investigate this, two CRPs were analyzed and compared for each child. The first CRP *only* displayed matches of gestures and speech of a skill level from the sensorimotor (S-)tier (i.e. level 1, 2 or 3), while the second CRP *only* displayed matches of a skill level from the representational (R-)tier (i.e. level 4, 5 or 6). Subsequently, vertical and horizontal LAM, TT and MaxL were calculated from these CRPs, and compared on the group level. Furthermore, to capture the relative strength and direction of the coupling, that is, the asymmetry between gestures and speech within a child, we calculated a relative difference score for each measure, for each child. This relative difference score is defined as the standardized difference between the measures derived from the vertical lines minus the measures derived from the horizontal line, as follows: $V-H_{LAM}$ was calculated as $LAM_V - LAM_H$ (LAM is a proportion and can readily be compared), $V-H_{TT}$ as $(TT_V - TT_H)/(TT_V + TT_H)$, and $V-H_{MaxL}$ as $(MaxL_V - MaxL_H)/(MaxL_V + MaxL_H)$. A model simulation by Cox et al. (2016) of the relation between relative difference in coupling strength and relative difference in horizontal and vertical line measures showed a strong association between relative coupling strength and the difference between LAM and TT, but not for MaxL. The relative difference scores of the S- and R-tier scores were also compared on a group level.

There are two reasons to expect dynamic differences in the gesture–speech interaction for different levels of understanding. First, as explained, skill levels from the sensorimotor tier include expressions about perceptions, action, and observable relations between these perceptions and actions, whereas skill levels from the representational tier are assigned to expressions that go beyond these observable actions and perceptions. Previously, the link between gestures and cognition has been assigned to gestures adding action information to existing mental representations (Beilock & Goldin-Meadow, 2010) and gestures simulating actions (Hostetter & Alibali, 2010). This presumed close relation between actions and gestures might culminate in a different interplay between gestures and speech at the sensorimotor tier compared to the representational tier. Also, more complicated levels of understanding are likely to arise when the task is complicated, that is to say, when children perceive the task to be more challenging. A challenging task might trigger learning, and previously it has been shown that gesture–speech mismatches tend to occur when a child is on the verge of learning something new (Goldin-Meadow, 2003). As described earlier, we suggest that gesture–speech

mismatches in a difficult, new and/or challenging task, arise from suboptimal simultaneous coordination of the gesture and speech synergies. When this suboptimal simultaneous coordination happens, the tight coupling between the action systems breaks down and becomes less dynamically stable and strong than for a less challenging task. Together we are inclined to expect that vertical and horizontal LAM, TT and MaxL will show different patterns of values at different levels of understanding.

Research Question 4

The final research question is: How are the measures of coordination between gestures and speech subsystems related to more stable child characteristics and school outcome measures, such as age and general level of cognitive performance? Children's use of speech and gestures is known to change over time (Goldin-Meadow, 1998). These changes are necessarily reflected in the dynamic organization of gestures and speech. Furthermore, as there is a link between gestures and cognition (Perry et al., 1988), children's general level of cognitive performance is also expected to be related to this dynamic organization. We investigate these possible relations by calculating correlations between Age, Math score, Language score, and Average skill level across the previous five interactions with the researcher and the LOS-profile measures (%Sync, RR_{peak} , Q_{LOS} , and τ_{peak}), the CRQA-measures (RR , LAM_V , LAM_H , TT_V , TT_H , $MaxL_V$, and $MaxL_H$) derived from the sensorimotor and representational tier, and the relative difference scores ($V-H_{LAM}$, $V-H_{TT}$ and $V-H_{MaxL}$) for each of the tiers.

Monte Carlo analysis

Throughout the Results section, p -values for differences between two measures were calculated by using Monte Carlo permutation tests (Todman & Dugard, 2001), which enabled us to reliably obtain significance levels with this relatively small sample (Ninness et al., 2002). Using this procedure, the probability that an empirically observed difference can be found was repeatedly calculated, in this case 1000 times, each time using a random distribution of the original data. If the average probability that the difference occurs in these random samples was small (i.e. $< .05$), we concluded that there is an actual difference present in the empirical data, which cannot be simulated using random samples, and hence, was not caused by chance. When a Monte Carlo permutation test was used to compare two values, we also calculated the effect size in the form of Cohen's d , that is, the observed difference divided by the pooled SD. A value of d between 0.2 and 0.3 is generally considered to be small, a value around 0.5 as medium, and a value of 0.8 and higher as large (Cohen, 1988).

Asymmetric dynamic attunement of speech and gestures

Results

Research Question 1: What is the temporal relation between gestures and speech, in terms of their displayed skill level?

For the first research question we expected that the LOS-profile analysis measures would display a significant asymmetry in the amount of recurrence around the LOS (Q_{LOS}) and display a peak in the recurrence (RR_{peak}) at some delay (τ_{peak}), indicating a leading role of gestures on speech. An overview of the values for Q_{LOS} , RR_{peak} and τ_{peak} in our sample can be found in Table 2. As described in the Method section, if Q_{LOS} is lower than 1, this suggests that gestures are leading speech in time. In our sample, Q_{LOS} ranged from 0.48 to 1.78, with an average of 1.08 which was not significantly higher than 1 ($p = .72$). The average Q_{LOS} ($M = 0.86$) of the children in Kindergarten was lower than the average Q_{LOS} ($M = 1.24$) of the children in first grade ($p = .04$, $d = 0.90$). This suggests that the gesture-speech dynamics had an opposite temporal pattern in the two age groups, with a leading role for speech for the first graders.

The observed RR_{peak} should exceed chance level, that is, there should be a real peak in the profile, for the observed τ_{peak} to make any sense. To verify this, a Monte Carlo procedure was performed to assess whether children's observed RR_{peak} significantly differed from chance. This was the case for all children in our sample (all p -values $< .01$), except for child 3 ($p = .63$). Therefore τ_{peak} of child 3 was not included in the subsequent analyses of this research question. On average τ_{peak} was 6.09 within the group, which was significantly higher than 0 ($p = .03$), indicating that gestures were ahead of speech in time. The average τ_{peak} of children in

Table 2

Overview of LOS-profile measures and CRQA-measures of all 12 children.

Child	Grade	LOS profile analysis measures				CRQA-measures over entire CRP					
		Q_{LOS}	RR_{peak}	τ_{peak}	RR	LAM _V	LAM _H	TT _V	TT _H	MaxL _V	MaxL _H
1	KG	0.46	.056	18	.013	.986	.910	5.2	3.4	21	7
2	KG	0.58	.089	16	.019	.996	.885	6.4	3.8	19	10
3	KG	0.91	.015	-	.004	.968	.687	4.3	2.6	12	3
4	KG	0.98	.076	2	.011	1.000	.885	7.4	5.1	26	11
5	KG	1.31	.012	36	.002	.893	.901	3.2	3.1	5	6
6	1	1.28	.034	-1	.010	.957	.701	6.6	2.6	16	5
7	1	0.48	.039	-1	.009	.979	.922	5.8	4.0	18	12
8	1	1.65	.034	0	.006	.973	.624	4.8	2.8	12	5
9	1	0.90	.140	0	.025	.992	.924	6.3	5.1	15	15
10	1	0.92	.053	-1	.016	1.000	.789	6.0	5.5	25	27
11	1	1.78	.021	-1	.002	.959	.632	5.4	2.7	18	3
12	1	1.66	.073	-1	.018	1.000	.793	8.3	3.6	24	6
Mean	-	1.08	.053	6.09	.011	.975	.805	5.8	3.7	17.6	9.2

Kindergarten ($M = 18$) differed from that of the first graders ($M = -.71$; $p < .01$, $d = 2.22$). In addition, the average τ_{peak} of children in Kindergarten was significantly higher than 0 ($p < .01$) and the average τ_{peak} of children in the first grade was significantly lower than 0 ($p < .01$). This is conform the earlier result (above), meaning that for the younger children in our sample gestures were ahead in time of speech (18 seconds on average), whereas, oppositely, gestures were behind in time of speech (0.71 seconds on average) for the older children.

Research Question 2: What is the relative strength and direction of the interaction between the gesture and speech subsystems?

See Table 2 for an overview of LAM, TT, and MaxL for both vertical and horizontal line structures. LAM_V ranged from .893 to 1.000 ($M = .975$), which means that 89.3% to 100% of the recurrent points comprised vertical line structures. TT_V ranged from 3.2 to 8.3 ($M = 5.8$), indicating that the average vertical lines in the recurrence plot consisted of 3.2 to 8.3 recurrent points. This reflects that gestures were trapped into same-tier skill-level episodes with average durations between 3 to 8 seconds for the different children. MaxL_V ranged from 5 to 26 ($M = 17.6$), which means that the maximum length of a vertical line in an individual recurrence plot ranged from 5 to 26 recurrent points. In other words, the maximum episode of gestures being trapped into a same-tier skill level lasted between 5 and 26 seconds. Calculations of the horizontal line structures revealed that the extent to which speech is trapped into displaying the same-tier skill level was somewhat less, with LAM_H ranging from .624 to .924 ($M = .805$), TT_H ranging from 2.3 to 5.5 ($M = 3.7$), and MaxL_H ranging from 3 to 27 ($M = 9.2$). At the group level, LAM_V , TT_V and MaxL_V were higher than LAM_H , TT_H and MaxL_H , respectively (all p -values $< .01$; $d_{\text{LAMV} > \text{LAMH}} = 2.01$; $d_{\text{TTV} > \text{TTH}} = 1.72$; $d_{\text{MaxLV} > \text{MaxLH}} = 1.31$). Interestingly, this is true for all children for LAM and TT, and for 9 out of 12 children also for MaxL. This finding clearly suggests an asymmetric dynamic attunement of gestures and speech, with gestures relatively more regularly and more rigidly displaying the same-tier skill level compared to speech.

Research Question 3: What is the relative strength and direction of the gesture-speech interaction for different skill-levels tiers?

We expected RR and vertical and horizontal LAM, TT and MaxL to be different for different levels of understanding. To analyze this, we first compared the averages of RR, LAM_V , LAM_H , TT_V , TT_H , MaxL_V , and MaxL_H on the sensorimotor (S)-tier with those on the representational (R)-tier. An overview of these CRQA-measures can be found in Table 3 (S-tier) and Table 4 (R-tier). The differences between the CRQA-measures of the S-tier or R-tier are weak to absent ($p_{RR} = .19$, $d = 0.31$; $p_{\text{LAM-V}} = .45$, $d = 0.05$; $p_{\text{TT-V}} = .45$, $d = 0.03$; $p_{\text{MaxL-V}} = .45$, $d = 0.05$; $p_{\text{LAM-H}} = .42$, $d = 0.08$; $p_{\text{TT-H}} = .91$, $d = 0.54$; $p_{\text{MaxL-H}} = .36$, $d = 0.12$). This means that there were no group-level

Asymmetric dynamic attunement of speech and gestures

Table 3

Overview of the CRQA-measures, calculated over skill levels 1 to 3 (sensorimotor tier).

Child	Grade	% RR*	LAM _V	LAM _H	V- H _{LAM}	TT _V	TT _H	V- H _{TT}	MaxL _V	MaxL _H	V- H _{MaxL}
1	KG	66.9%	.669	.595	.074	7.6	3.2	.41	21	7	.50
2	KG	29.3%	.289	.226	.063	8.3	2.5	.53	19	3	.73
3	KG	99.3%	.961	.687	.273	4.3	2.3	.31	12	3	.60
4	KG	7.2%	.072	.048	.024	6.0	3.2	.30	6	7	-.08
5	KG	73.3%	.733	.672	.061	3.2	3.1	.01	5	6	-.09
6	1	95.6%	.915	.672	.243	6.7	2.6	.44	16	5	.52
7	1	31.5%	.308	.248	.059	7.0	3.3	.37	18	5	.57
8	1	73.9%	.721	.480	.241	4.3	2.7	.24	8	4	.33
9	1	29.8%	.290	.267	.023	7.6	3.9	.32	15	15	.00
10	1	60.3%	.603	.539	.064	5.1	5.4	-.03	10	27	-.46
11	1	20.5%	.192	.103	.089	4.7	3.0	.22	10	3	.54
12	1	19.8%	.198	.161	.037	5.0	3.3	.21	9	5	.29
<i>M</i>	<i>KG</i>	<i>55.2%</i>	<i>.545</i>	<i>.446</i>	<i>.10</i>	<i>5.9</i>	<i>2.9</i>	<i>.31</i>	<i>12.6</i>	<i>5.2</i>	<i>.33</i>
<i>M</i>	<i>1</i>	<i>47.3%</i>	<i>.461</i>	<i>.353</i>	<i>.11</i>	<i>5.8</i>	<i>3.4</i>	<i>.25</i>	<i>12.3</i>	<i>9.1</i>	<i>.26</i>
<i>M</i>	<i>Overall</i>	<i>50.6%</i>	<i>.496</i>	<i>.391</i>	<i>.104</i>	<i>5.8</i>	<i>3.2</i>	<i>.28</i>	<i>12.4</i>	<i>7.5</i>	<i>.29</i>

*Note: % RR reflects the percentage of recurrence found on the S-tier, as compared to the overall recurrence rate on both the S- and R-tier, displayed in Table 2.

differences in the relative strength and direction of the interaction between gestures and speech for lower (S-tier) levels nor for higher (R-tier) levels of understanding.

Next, we analyzed whether the measures derived from the vertical and horizontal line structures showed the same pattern of differences for the S-tier and R-tier. LAM_V was not higher than LAM_H for both the S-tier ($M_{LAM-V} = .496$, $M_{LAM-H} = .391$, $p = .14$, $d = 0.38$) and the R-tier ($M_{LAM-V} = .479$, $M_{LAM-H} = .413$, $p = .30$, $d = 0.22$). However, the analysis revealed TT_V to be higher than TT_H for both the S-tier ($M_{TT-V} = 5.81$, $M_{TT-H} = 3.19$, $p < .01$, $d = 2.06$) and R-tier ($M_{TT-V} = 5.75$, $M_{TT-H} = 3.88$, $p = .01$, $d = 0.99$). In addition, MaxL_V was higher than MaxL_H for both the S-tier ($M_{MaxL-V} = 12.42$, $M_{MaxL-H} = 7.50$, $p = .03$, $d = 0.80$) and R-tier ($M_{MaxL-V} = 12.75$, $M_{MaxL-H} = 6.83$, $p = .02$, $d = 0.92$). Lastly, the relative difference scores between the S-tier and R-tier did not differ ($p_{V-H-LAM} = .15$, $d = 0.43$; $p_{V-H-TT} = .28$, $d = 0.22$; $p_{V-H-MaxL} = .38$, $d = 0.13$).

To summarize, the average differences between the CRQA-measures of vertical and horizontal lines showed the same pattern for the S-tier and R-tier. This means that the relative strength and direction of the coupling between gestures and speech did not differ between the levels of understanding. At the group level, they were similarly asymmetric for both tiers. Also,

Table 4

Overview of the CRQA-measures, over skill levels 4 to 6 (representational tier).

Child	Grade	% RR*	LAM _V	LAM _H	V- H _{LAM}	TT _V	TT _H	V- H _{TT}	MaxL _V	MaxL _H	V- H _{MaxL}
1	KG	33.1%	.316	.315	.002	3.1	3.9	-.11	5	7	-.17
2	KG	70.7%	.707	.660	.047	5.8	4.7	.11	9	10	-.05
3	KG	0.7%	.007	.000	.007	3.0	0.0	1.00	3	1	.50
4	KG	92.8%	.928	.837	.090	7.5	5.3	.17	26	11	.41
5	KG	26.7%	.160	.229	-.069	3.0	3.0	.00	3	5	-.25
6	1	4.4%	.042	.030	.013	4.8	4.0	.09	7	4	.27
7	1	68.5%	.671	.674	-.003	5.3	4.4	.09	11	12	-.04
8	1	26.1%	.252	.145	.107	7.0	3.3	.35	12	5	.41
9	1	70.2%	.702	.657	.045	5.8	5.9	-.01	10	10	.00
10	1	39.7%	.397	.250	.147	8.2	5.7	.18	25	8	.52
11	1	79.5%	.767	.530	.237	5.6	2.7	.35	18	3	.71
12	1	80.2%	.802	.632	.170	9.8	3.7	.45	24	6	.60
<i>M</i>	<i>KG</i>	<i>44.8%</i>	<i>.424</i>	<i>.408</i>	<i>.02</i>	<i>4.5</i>	<i>3.4</i>	<i>.23</i>	<i>9.2</i>	<i>6.8</i>	<i>.09</i>
<i>M</i>	<i>1</i>	<i>52.7%</i>	<i>.519</i>	<i>.417</i>	<i>.10</i>	<i>6.7</i>	<i>4.2</i>	<i>.21</i>	<i>15.3</i>	<i>6.9</i>	<i>.35</i>
<i>M</i>	<i>Overall</i>	<i>49.4%</i>	<i>.479</i>	<i>.413</i>	<i>.066</i>	<i>5.8</i>	<i>3.9</i>	<i>.22</i>	<i>12.8</i>	<i>6.8</i>	<i>.24</i>

laminarity (LAM) did not show the same asymmetry at the individual levels of understanding, as it did when the tiers were joined together for Research Question 2.

Does age play a role?

Prompted by the differences between younger and older children found for Research Question 1, we investigated whether similar age-group differences were present in the strength and direction of the interaction between gestures and speech for different levels of understanding. To this end, we compared the children in Kindergarten and first grade with regard to their CRQA-measures and relative difference scores on the S-tier and R-tier. These measures are displayed in Table 3 and 4.

For the S-tier, no clear differences between the CRQA-measures of younger and older children were found ($p_{RR} = .26$, $d = 0.34$; $p_{LAM-V} = .30$, $d = 0.26$; $p_{LAM-H} = .25$, $d = 0.37$; $p_{TT-V} = .46$, $d = 0.05$; $p_{TT-H} = .07$, $d = 0.73$; $p_{MaxL-V} = .50$, $d = 0.06$; $p_{MaxL-H} = .12$, $d = 0.57$). There were also no differences between the younger and older children with regard to the average relative difference scores on the S-tier ($p_{V-H LAM} = .41$, $d = 0.09$; $p_{V-H TT} = .24$, $d = 0.36$; $p_{V-H MaxL} = .35$, $d = 0.20$). For the R-tier, only TT_V of the older children was higher than TT_V of the younger children ($p_{TT-V} = .04$, $d = 1.12$). Even though the other CRQA measures on the R-tier might appear to be higher for the older children, no meaningful differences were found ($p_{RR} = .40$, $d = 0.17$; $p_{LAM-V} = .31$, $d = 0.29$; $p_{LAM-H} = .48$, $d = 0.03$; $p_{TT-H} = .17$, $d = 0.54$; $p_{MaxL-V} = .12$, $d = 0.73$; $p_{MaxL-H} = .51$, $d = 0.02$).

Asymmetric dynamic attunement of speech and gestures

Considering the relative difference scores, only $V-H_{LAM}$ was higher for older than for younger children ($p_{V-H_{LAM}} = .02$, $d = 1.11$). There were no clear difference for $V-H_{TT}$ ($p_{V-H_{TT}} = .46$, $d = 0.06$) and only slightly for $V-H_{MaxL}$ ($p_{V-H_{MaxL}} = .07$, $d = 0.85$).

In conclusion, for the less difficult levels of understanding on the S-tier, older and younger children did not differ in the strength and direction of the interaction between gestures and speech. However, for the more difficult levels of understanding there were age-differences in the asymmetry of the gesture-speech interaction: Gestures displayed longer average periods of lingering in the R-tier (TT_V) and were more regular ($V-H_{LAM}$) for the older children than for the younger children.

Research Question 4: How are the measures of coordination between gestures and speech subsystems related to more stable child characteristics and school outcome measures?

An overview of the significant correlations between child characteristics and school outcome measures, and the LOS-profile measures, CRQA -measures and relative difference scores can be found in Table 5. The entire correlation table is available in the supplementary materials. First we will describe the findings for the LOS-profile measures across both tiers, followed by the CRQA-measures and relative difference scores separately for each tier.

When recurrences on the sensorimotor and representational tier are combined, the correlation of %Sync and age had a value of .57. This means that relatively older children tended to show the same-tier skill level at the same time in gestures and speech. The correlation of -.73 between τ_{peak} and age in months corroborates to this finding, as it implies that younger children tended to show a more extensive delay between gestures and speech in displaying the same-tier skill level, with gestures being ahead of speech in time.

For the S-tier separately, LAM_V and $V-H_{LAM}$ were both negatively correlated with children's Math score and Average score on past tasks ($r = -.54$ and $r = -.58$, respectively). This means that for children who performed better on math and past tasks, gestures were being trapped into S-tier episodes less prominently. Moreover, for these children the asymmetry between gestures and speech was smaller. LAM_H was also negatively correlated with the average score on past tasks ($r = -.52$), which suggests that for children with a higher score on past tasks, speech was less prone to be trapped into S-tier episodes as well. Language score was correlated with TT_V ($r = .53$) and $V-H_{TT}$ ($r = .59$) on the S-tier, which shows that for children with a higher Language score, gestures were trapped into longer average S-tier episodes, and that the associated asymmetry between gestures and speech tends to be bigger.

Table 5

Significant correlations between child characteristics and CRQA-measures.

		Age (months)	Math score	Language score	Average score past tasks
Both tiers	%Sync	.57*			
	τ_{peak}	-.73**			
S-tier	LAM _V		-.54*		-.58**
	LAM _H				-.52*
	V-H _{LAM}		-.62**		-.58**
	TT _V			.53*	
	V-H _{TT}			.59*	
R-tier	LAM _V		.51*		.56*
	LAM _H		.57*		.54*
	V-H _{LAM}	.65**			
	TT _V	.51*			
	TT _H	.61**			.61**
	V-H _{TT}		-.68**		-.67**
	MaxL _H	.65**			
	V-H _{MaxL}	.52*	-.50*		

Note 1: Values marked with * are significant at $p < .1$, values marked with ** are significant at $p < .05$.

Note 2: The complete correlation matrix can be found in the supplementary materials.

For the more difficult skill-levels on the R-tier, it turns out that all CRQA and LOS profile measures are significantly correlated with age or measures of general performance. Both LAM_V and LAM_H are correlated with Math score ($r = .51$ and $r = .57$, respectively) and the average score on past tasks ($r = .56$ and $r = .54$, respectively). This suggests that for children with a higher score on math or past tasks, both speech and gestures were trapped into R-tier episodes more often. Age correlates with V-H_{LAM}, which means that the asymmetry between gestures and speech tended to be bigger for older children. TT_V was related to Age ($r = .51$), suggesting that older children were trapped into longer average R-tier gesturing episodes. Both Age and Average score on past tasks were correlated with TT_H ($r = .61$ and $r = .61$, respectively), which means that children who are older or who performed better on past tasks were trapped into longer average R-tier speech episodes. As V-H_{TT} is negatively correlated with both Math score and Average score on past tasks ($r = -.68$ and $r = -.67$, respectively), children who performed well on math or past tasks tended to display a smaller asymmetry in the average duration of gestures and speech R-tier lingering. MaxL_H and V-H_{MaxL} were related to age ($r = .65$ and $r = .52$, respectively), which suggests that older children had a longer maximum episode of speech being trapped at the R-tier, but at the same time, the asymmetry between gestures and

Asymmetric dynamic attunement of speech and gestures

speech tended to be larger for this. Finally, $V-H_{\text{MaxL}}$ was negatively correlated with Math score ($r = -.50$). So children with a higher score on math had a smaller asymmetry in the longest gestures and speech R-tier lingering episode.

Discussion

Summary of results

The present study concentrated on how the earlier reported leading role of gestures over speech in children's cognitive change arises from the asymmetries in the dynamic attunement of gestures and speech during task performance. Appreciating the dynamic nature of this issue naturally implied using of the language and methods of complex dynamical systems. Accordingly, we used Cross Recurrence Quantification Analysis (CRQA), a novel nonlinear time series method, to analyze the two skill-level time series as coded from children's gestures and speech while they were working on an educational science task. To be able to address this rather broad issue intelligibly we proposed four specific research questions, focusing on: 1) the temporal relation between gestures and speech, 2) the relative strength and direction of the interaction between gestures and speech, 3) the relative strength and direction between gestures and speech for different levels of understanding, and 4) the relations between measures of dynamic organization and more stable child characteristics and school outcome measures.

Firstly, regarding the temporal relation, older and younger children differed in the (temporal) asymmetry in the gestures–speech interaction. In the two minute window of the LOS-profile analysis, in younger, i.e. Kindergarten, children, the balance leant more towards gestures leading speech in time, whereas the balance leant more towards speech leading gestures in time for the older first-grade students. This difference between older and younger children is even more pronounced when we look at the actual temporal delay in seconds. While gestures are, on average, ahead of speech for 18 seconds for the younger children, speech only slightly precedes gestures for just under a second for the older children.

Secondly, we investigated the relative strength and direction of the interaction between gestures and speech as it plays out on all possible timescales, ranging from the sample rate (one second) to the entire interaction (~ 489 seconds). As described earlier, calculating and comparing recurrence measures of vertical and horizontal line structures is informative about the coordinative structures in the gesturing–speech interaction. At the group level, we found LAM, TT and MaxL to point towards speech influencing gestures more regularly and rigidly into displaying the same-tier skill level than vice versa. Moreover, when comparing the strength and direction for different levels of understanding (Research Question 3), this asymmetry in

gestures and speech extended to both the sensorimotor and representational tier. The relative difference scores did not differ for the S-tier and R-tier. In other words, there are no differences in the coupling between gestures and speech for different levels of understanding at the group level.

However, when we compared the CRQA measures for different levels of understanding of children from first grade and Kindergarten, an interesting pattern of differences appeared. Although no differences were present at the S-tier, at the more difficult R-tier level of understanding, older and younger children did differ in the coupling between gestures and speech. All CRQA measures were higher for the older children at the R-tier, suggesting that the coupling between gestures and speech was more rigid at higher levels of understanding.

The relation of age with the coupling between gestures and speech is also apparent when we relate the CRQA measures to individual child characteristics. The correlations between age and %Sync, and between age and τ_{peak} support the results from the LOS-profile analysis. This again shows that gestures are more ahead of speech in time when children are younger, and that they are more temporally aligned when children are older. The results reveal a larger asymmetry in the gesture-speech attunement for older children. A higher score on schools' standardized language tests is also related to more asymmetry between gestures and speech, but only for the less difficult levels of understanding (S-tier).

However, children's average score on past tasks and their scores on math seem to be related to speech attracting gestures less, and also to less asymmetry between gestures and speech for the less difficult levels of understanding. For the more difficult levels of understanding (R-tier), both speech and gestures tend to attract each other more for children with a higher score on math or past tasks, which points to more symmetry between speech and gestures. Moreover, a higher score on math or past tasks is also related to less asymmetry between gestures and speech at the R-tier.

Dynamic, entangled development of gestures, speech and cognitive skills

Earlier studies have shown that children express new cognitive insights by means of gestures before they are able to put them into words. An important nuance following from the present study is that although gestures might appear to be ahead in time of speech during children's learning, this does not imply that gestures influence speech to a larger extent. Learning is a process that occurs at multiple, nested time scales, by means of entangled processes of action, perception and cognition. In studies thus far, such a process approach has not been considered with respect to the interplay of gestures and speech in children's learning. At the very least our study shows that the relation between gestures, speech and cognition in our sample is much more dynamic and bidirectional than previously thought, with a high degree of

Asymmetric dynamic attunement of speech and gestures

inter-individual variability. In addition, children differ in how speech and gestures are coupled, whereby gestures are not always ahead of speech, or leading speech, as cognitive understanding unfolds. Moreover, the gestures-speech coupling is related to age and measures of scholastic and cognitive performance that exceed the time-span of a single task.

Age, Language score and the dynamic emergence of speech and gestures

One particularly prominent result is that, with increasing age, speech and gestures become more synchronized and tightly coupled. Within this tight coupling for older children, speech attracts gestures more than vice versa in displaying the same-tier skill level. A possible explanation for this finding can be found in Iverson and Thelen's (1999) account of the dynamic emergence of speech and gestures. They suggest that the link between speech and gestures starts with the hand-mouth linkage that is already apparent in newborns. Coordination between oral and manual actions is very common in newborn's spontaneous actions, such as bringing their hands to the facial area or sucking their fingers. These connections between oral and manual actions are characterized by a low threshold—as they are so easily and spontaneously performed—and high activation, because of their frequency. Around the age of 6 to 8 months, both rhythmical arm movement and rhythmical babbling emerge, through which coinciding vocal and manual activities are entrained.

The linkages between the hands and mouth become more controlled as children develop, with the emergence of the first gestures and words around 9 to 14 months of age. Typically, children's gestures precede and outnumber their spoken words tremendously during this period. To be more specific, children's pointing gestures precede the word for an object by, on average, 3 months, and gesture-plus-word combinations precede two-word combinations by an average of 4.7 months (Iverson & Goldin-Meadow, 2005). According to Iverson and Thelen (1999), the reason for this is that, in comparison to the vocal articulators, the control of the hands is more advanced and therefore it is easier for children to communicate by means of gestures. In other words, for gestures the threshold is low and activation is high, while for speech the threshold is high and activation is low. However, as children practice their vocal skills, the threshold of speech becomes lower and activation higher. The activation of speech eventually becomes so high, that it captures and concurrently activates gestures. Stated differently, as children's language skills become more advanced, their speech system activates their gesture system, and thereby the two motor systems become more synchronized.

Returning to our finding that older children in our sample show higher levels of synchronization and coupling between speech and gestures. It is safe to assume that older children have more developed speech and gesture synergetic control. The reason for this is that both action systems have been explored and practiced more, and under more different and variable task

conditions, than in the younger children (cf. Iverson & Thelen, 1999). Because of this, the speech and gesture synergies are more entrained, which means that older children can coordinate the two synergies more optimally and simultaneously. This reasoning and the finding that speech influences gestures more than vice versa in older children, is in line with Iverson and Thelen's (1999) notion of speech capturing gestures when vocal skills become more practiced. A final noteworthy observation in this respect is that the older children in our sample just entered first grade, in which they learn to read and write. Although speculative at this point, it is not farfetched to expect that this emphasis on language in the first grade increases how much speech is able to influence gestures (cf. Shanahan & Roof, 2013).

As already implied in the previous section, the explanation that gestures are ahead of speech in time for the younger children, with an average delay of 18 seconds, might also be found in the simultaneous coordination of the synergies of speech and gestures. For the younger children the task might be more difficult than for the older children, and pose considerably more conflicting task constraints. These conflicting task constraints may cause the two synergies to be unable to simultaneously exist in an optimal way. This makes the tightly coupled synergies dissociate, with the gesture synergy being created first and the speech-synergy later. The average lag of 18 seconds between speech and gestures might intuitively seem hard to understand, but such contingencies over relatively large timescales have been found before in the context of communication, albeit with younger children. For example, Jaffe, Beebe, Feldstein, Crown and Jasnow (2001) report a 20 to 30 second lag between contingencies in the vocal patterns of 4-month old infants and their mothers or strangers. Moreover, Jaffe et al. point to other studies, which found a 20 to 30 second cycle in infant attention (Brazelton, Kozlowski and Main, 1974), a 10 to 45 second cycle in coordination of facial engagement (Lester, Hoffman and Brazelton, 1985), and a 20 second lag in facial engagement correlation (Cohn and Tronick, 1988). Although this concerns interpersonal coordination, these studies demonstrate that latencies of this magnitude are not extreme.

Jaffe et al. propose that the 20 to 30 second lag between contingencies in the vocal patterns of the infants and their mother or a stranger is an indication for a slow rhythm in the interaction. This slow rhythm can only be found by analyzing the data in much detail, as opposed to rhythms such as vocalization-pause or turn taking, which are detectable for untrained observers. To return to our study: both speech and gestures are suggested to originate from coinciding rhythmical activities (Abney, Warlamout, Haussman, Ross & Wallot, 2014; Iverson & Thelen, 1999), and in fact, speech and gestures are rhythmical activities in itself (Loehr, 2007). The average delay between speech and gestures of 18 seconds that we found for the younger children might be a slow rhythm in the gesture-speech interaction. This slow rhythm may reside

Asymmetric dynamic attunement of speech and gestures

in a process on a larger timescale, in which both the synergies of gestures and speech are nested. Which specific process this would be remains a question for future research.

The relation we found between a higher language score and more asymmetry in the speech-gesture coupling fits with the dynamic emergence account of speech and gestures as outlined above. With an explained variance of 25%, better language skills are associated with a stronger influence of speech on gestures. Interestingly, the relation between a higher language score and more asymmetry is not apparent for the higher levels of understanding on the R-tier. An explanation for this might be that the levels of understanding on the R-tier go beyond the skill of naming observable task characteristics, but rather involve relations among task elements, and relations among relations (cf. Fischer & Bidell, 2006; see below). A second explanation might be that understanding on the R-tier is more difficult, which causes a different interplay between the synergies of speech and gestures than on the S-tier - in this case less asymmetry in influence.

Average score on past tasks, math-score and higher-order understanding emerging from actions

Contrary to age and language score, a higher average score on past tasks is not related to a leading role of speech over gestures, but instead to a more symmetric interaction between speech and gestures. In order to grasp this finding, consider how higher-order understanding emerges from actions. In a previous study, participants were asked to perform a gear task and predict the turning direction of a target gear (Trudeau & Dixon, 2007). At first, all participants simulated the motions of the gears with their hands, i.e. force tracing, to predict in which direction the target gear would turn. After a certain number of these problems, participants discovered a higher-order relation, alternation, which is concealed in the task. Alternation, like all higher-order relations, is a relation among relations and requires coordinating two or more lower-order relations and integrating multiple actions over time. Participants varied considerably in how many simulations they performed before discovering alternation, and Trudeau and Dixon (2007) found that the number of alternating actions performed before discovering the higher-order rule predicted the likelihood of generalization of this rule to new problem types. Trudeau and Dixon (2007) explained this finding by stating that for participants who made more correct alternating actions before discovering the higher order rule, the representation of alternation stems from a larger corpus of actions. This larger corpus of relevant (i.e. task-related) actions increases the chance of discovering and being able to generalize the higher-order relation. Extrapolating on this thought, children for whom speech is less leading over gestures might be more open to gesturing, that is, they might gesture more. This provides them with the larger corpus of actions, which increases their chance of discovering higher-order relations by means of actions, resulting in a higher score on (past) tasks.

Even more so, the gestures of children may also elicit discovering these higher-order relations in other tasks. Indeed, Smith (2010) has emphasized the essence of the motor system in learning higher-order regularities. She states that “It is action that creates a task, that couples component systems in the moment, and that selects and creates the momentary dynamic input on which learning must depend” (p. 264). In the context of action, component systems become coupled and coordinated within diverse tasks, which makes action essential for learning higher-order relations and generalizing learning such relations to other tasks. With respect to our findings, gesturing may also elicit the discovery of higher-order relations in other tasks, which might explain why children for whom speech was less leading over gestures performed better on past tasks.

Next to a higher average score on past tasks, a higher match score is also related to a more symmetric interaction between gestures and speech, whereby speech is less leading over gestures. Note that these two scores were highly correlated ($r = .84$), meaning that children who scored high on math were also likely to have done well on previous tasks. It is well-known that gestures are beneficial for math learning (e.g., Alibali & Nathan, 2012; Broaders et al., 2007; Cook et al., 2013; Cook et al., 2012; Cook & Goldin-Meadow, 2006; Ehrlich et al., 2006; Novack et al., 2014). The reason why gestures are related to math might be the same as why gestures are related to a higher average score on past tasks: From gestures, higher-order (mathematical) understanding can emerge and generalize. Indeed, Novack et al. (2014) and Cook et al. (2013) found that gestures are related to the generalization and transfer of mathematical knowledge to new problem types.

To summarize this subsection, children within our study for whom speech is less leading over gestures may perform better on both math and past tasks because they are more open to gesturing, from which higher-order understanding is thought to emerge and generalize. A reason for this might lie in the importance of variability in learning (e.g., Van Geert & Van Dijk, 2002). If the first system influences the second system to a lesser degree, that second system is obviously less constrained by the first, and can adhere more adaptively to task requirements. In other words, the coupling of the two systems can be characterized as more flexible, which allows for different types of coordination between them and with the environment.

Following the framework introduced earlier, we interpret the finding that speech is less leading over gestures in terms of synergetic competition. Accordingly, a decreasing leading role of speech over gestures indicates a more optimal and efficient (simultaneous) coordination of both synergies. However this optimal coordination of both synergies does not necessarily have to be simultaneous since we found that more temporal synchrony of speech and gestures is not related to a better score on past tasks or a better math performance in our study. Future

Asymmetric dynamic attunement of speech and gestures

studies could focus more specifically on how improved understanding of concepts and/or performance on a task is related to a more optimal (possibly but not necessarily simultaneous) coordination of both the speech and gestures synergies. This could become visible, for instance, by a decrease in the temporal delay between gesture and speech behaviors (Goldin-Meadow et al., 1993), a phase-transition like change from a period of suboptimal coordination of one or both synergies to state of simultaneous optimal coordination (cf. Haken, Kelso & Bunz, 1985), or perhaps by a change in the temporal structure of gestures and speech (cf. Den Hartigh et al., 2015; Wijnants et al., 2009; 2012).

Conclusion

Our results suggest that speech and gestures may be more *tightly coupled* for the older children in first grade and children with a high language score, because their speech and gesture systems are more developed. The reason that speech *leads* over gestures for these children may as well stem from this developmental process, and might be enhanced by the emphasis on language in first grade. For children with a higher average score on past tasks and math score, speech is leading less over gestures, possibly leaving more room for higher-order understanding to emerge from their action experiences through gesturing. Because of the time-intensive coding procedures and the in-debt nature of our analyses, this study used a small N . Note that we used Monte Carlo permutation tests, which are particularly strong in the case of small sample sizes. The credibility of our results is further strengthened by the relatively large effect sizes we found (Cohen, 1988). Nonetheless, this study deserves replication to check whether the findings can be verified and eventually also further refined and strengthened.

It is important to note that speech leading less over gestures is not the same as *less speech* or *less coupling between gestures and speech*. In fact, higher-order understanding, and more broadly speaking, cognition itself, resides in and emerges from the coupling between a multitude of perception-action subsystems, such as those related to speech and gestures (Goldin-Meadow, 1998). Congruously, within our study, the child for whom there was no RR_{peak} , that is, for whom coupling between gestures and speech was weaker, had low scores on all the other variables of cognitive performance. To elaborate, it is not the mere presence or absence of coupling between subsystems that is important, but rather the nature of their coupling, in the sense of interaction-dominant dynamics (Van Orden et al., 2003; 2005). How the subsystems are coupled determines how development will progress, and whether and how higher-order understanding will occur. Our findings suggest that a coupling in which the influence of gestures and speech is more balanced (i.e. where speech is less leading), seems to be beneficial for higher-order understanding to develop in this respect, in a hands-on science and technology task.

As cognition resides in and emerges from the dynamic coupling between perception-action subsystems, and learning is a non-linear process with variability as its hallmark, methods that capture this coupling over time are necessary to understand how development comes about. The complex dynamical systems approach provides a framework for asking question and interpreting answers pertaining to how higher-order relations can emerge from perception-action couplings. In our study, we investigated how the speech and gesture subsystems of children are coupled during a hands-on educational science task. Among other things, we found this coupling to be related to other measures of cognitive performance. Instead of gestures expressing or adding to a rather disembodied cognitive insight before speech is able to express it, we outlined how higher-order understanding might emerge from the changing coupling between gestures and speech over time. Moreover, we proposed a new mechanism, of competing and suboptimal coordinated synergies resulting in gestures-speech mismatches, that builds a bridge between the existing research on gestures and recent views on cognition as fundamentally embedded and embodied. Future studies should investigate if the dynamic organization of gestures and speech indeed points to gesture-speech mismatches as originating from competing synergies of gestures and speech.



3

Easier said than done? Task difficulty's influence on temporal alignment, semantic similarity, and complexity matching between gestures and speech

This chapter is based on:

De Jonge-Hoekstra, L., Cox, R., van der Steen, S., & Dixon, J. A. (2021). Easier said than done? Task difficulty's influence on temporal alignment, semantic similarity, and complexity matching between gestures and speech. *Cognitive Science*. doi: 10.1111/cogs.12989

Easier said than done? Task difficulty's influence on temporal alignment, semantic similarity, and complexity matching between gestures and speech

Gestures and speech are two salient aspects of multimodal communication in humans. When people tell a story, explain a difficult problem, or talk about daily affairs, they tend to move their hands in all kinds of ways. Many researchers have therefore proposed that gestures and speech are tightly coupled (e.g. Goldin-Meadow, 2003; McNeill, 1985). Moreover, this tight coupling has been conceptualized as gesture-speech synchronization (e.g. Iverson & Thelen, 1999; Pouw & Dixon, 2019b; Treffner & Peter, 2002). Gestures and speech synchronize in time, semantic content, emphasis, and emotional valence (for a comprehensive review, see Wagner et al., 2014).

However, the semantic similarity between gesture and speech has been shown to break down as people approach transitions in understanding (e.g., an insight into a difficult problem; Church & Goldin-Meadow, 1986; Goldin-Meadow, 2003). For instance, in a liquid conservation task a researcher pours equal amounts of water into a wide glass and a narrow glass and asks a child which glass contains more water. When a child is about to learn the concept of conservation, they might say that there is more water in the narrow glass because the level of water is higher, while they gesture about the width of the glasses (Church & Goldin-Meadow, 1986). These instances of semantic dissimilarity are called *gesture-speech mismatches*.

Different explanations exist for the breakdown in semantic similarity between gesture and speech when people approach transitions in understanding. Goldin-Meadow and colleagues' (Church & Goldin-Meadow, 1986; Goldin-Meadow, 2003) explanations center around participants' conflicting cognitive strategies and hypotheses that are thought to exist just before participants achieve a new insight into the problem they are working on (e.g. liquid conservation task). These conflicting strategies and hypotheses are then somehow differently expressed in gestures than in speech, during gesture-speech mismatches. However, Koschmann (2017) questions the existence of gesture-speech mismatches in the first place, and suggests that they are an artefact of the disintegrated methodological coding systems that led to their discovery. Furthermore, Pouw et al. (2017; also see Pouw et al., 2014) highlight an explanatory gap in how an integrated gesture-speech system could produce disintegrated gesture-speech mismatches, and suggest taking a dynamically embodied perspective to address this gap.

From a dynamically embodied, complex system's perspective, a change in understanding can be seen as a system of interrelated components which transitions from one stable state to a new, likely more advanced, stable state (Smith & Thelen, 2003; Stephen, Boncoddio, et al., 2009;

Task difficulty's influence on gesture-speech synchronization

Stephen, Dixon, et al., 2009; Thelen & Smith, 1994; Thelen & Smith, 2007; Van Geert, 2008; Van Geert, 2011). A transition from one stable state to another entails a reorganization of a system's components and their relations. This reorganization is elicited by a perturbation, that is, the learning situation. As put forward by De Jonge-Hoekstra et al. (2020), a metaphor for this reorganization is building a new LEGO-structure from an old structure, which is only possible when you break (perturb) the old structure and use the bricks to create a new structure. Taking such a dynamically embodied, complex system's perspective, De Jonge-Hoekstra et al. (2016) suggest that difficult tasks perturb a system, thereby inducing a suboptimal coordination between gestures and speech, which could then lead to various forms of gesture-speech mismatches.

In this study, we empirically address whether task difficulty indeed affects gesture-speech synchronization. We will approach gesture-speech synchronization in three ways: 1) temporal alignment, 2) semantic similarity, and 3) complexity matching (explanation follows below). We will investigate how task difficulty affects temporal alignment, semantic similarity, and complexity matching between gestures and speech, and how these different forms of gesture-speech synchronization are related. In addition, we will investigate how these three gesture-speech synchronization measures predict task performance.

Synchronization

Synchronization usually means that two (or more) systems start to behave in a similar way due to coupling (Pikovsky et al., 2001). In cognitive science, synchronization comes in different forms, including temporal alignment, semantic similarity, and complexity matching. We will explain these three forms below, and describe how they have been linked to gesture-speech synchronization.

Temporal alignment

Temporal alignment is a well-known form of synchronization. A simple and widely used example of temporal alignment are two asynchronously ticking metronomes, which start to tick in synchrony when they are placed on a shelf on top of two cans that act like wheels (the movement of each metronome is transmitted through the wheels thus providing coupling; see Figure 1). Also within humans, body parts such as fingers (e.g. Haken et al., 1985; Kelso, 1994) and legs (Clark et al., 1988) have been shown to synchronize and temporally align in rhythmic patterns. Moreover, a recent study by Pouw et al. (2018) shows that speech is more rhythmic when it goes together with more gestures, suggesting a rhythmic synchronization between gestures and speech within humans. This paradigm of one-to-one temporal alignment of behavior has been applied to coordination between humans, where it has been found that humans tend to move in synchrony while rocking in rocking chairs (Richardson et al., 2007),

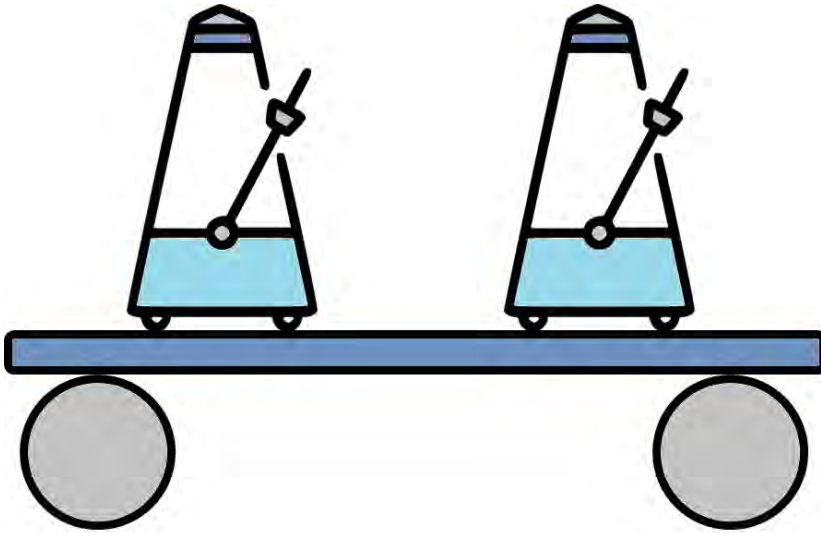


Figure 1. Synchronization of coupled metronomes.

swinging pendulums (Richardson et al., 2005; Schmidt & O'Brien, 1997), or telling jokes (Schmidt et al., 2014), to name a few examples.

With regard to temporal alignment between gestures and speech, adult's gestures and speech are highly aligned in time (see Wagner et al., 2014, for an overview). In other words, most gestures beat in-phase, and at the same rhythm as speech (Prieto & Roseano, 2018; also see Pouw et al., 2018). For gestures, this rhythm consists of changes in hand-movement velocity over time, while for speech this rhythm consists of changes in the amplitude of the sound produced by the speaker (also see Fowler, 2010). To support the existence of temporal alignment between gestures and speech, several studies indicated that the moment of *maximum effort* in gestures goes together with changes in pitch (i.e. relative frequency, "highness" or "lowness") of speech (Kendon, 1972; Kita et al., 1998; Leonard & Cummins, 2011). Recent studies by Pouw and colleagues (Pouw et al., 2018; Pouw & Dixon, 2019a, 2019b) showed that this relation between maximum gestural effort and speech is actually a tight alignment of peak velocity in gestures and peak pitch in speech.

Some circumstances affect the temporal alignment between gestures and speech. Children's age is a robust correlate with the temporal alignment between gestures and speech. According to Iverson and Thelen (1999), the coupling between gestures and speech in infants emerges from natural oscillations of hand movements and vocal acts, which synchronize and become entrained over time (see also Esteve-Gibert & Prieto, 2014; Iverson & Fagan, 2004). As a consequence of this entrainment, the temporal alignment between gestures and speech

Task difficulty's influence on gesture-speech synchronization

becomes higher when infants and toddlers grow older (Butcher & Goldin-Meadow, 2000; also see Iverson & Thelen, 1999). Adults' gestures and speech are so tightly coupled in time, that perturbing and delaying speech with a Delayed Auditory Feedback also delays gestures (e.g. Rusiewicz et al., 2013, 2014). Pouw and Dixon (2019b) found that a Delayed Auditory Feedback actually increases the temporal alignment between gestures and speech. Lastly, Bergmann et al., (2011) found that gestures and speech were more temporally aligned when their semantic content was more similar.

Semantic similarity

Semantic similarity refers to similarities in *meaning*. Humans can synchronize on a semantic level, whereby they align their "[...] understanding of the world with others [...]" (Dumas & Fairhurst, 2019, p. 10). Important to note is that semantic synchronization is not confined to (spoken) language, but can take other action forms involving other body movements as well (Dumas & Fairhurst, 2019). Bodily forms of semantic, meaningful synchronization, such as playing give-and-take-games, or interpersonal movement coordination when a parent dresses their child, seem to be essential for language development. Furthermore, differences in the semantic similarity of two people's words influence their bodily synchronization (see Shockley et al., 2009, for an overview).

Gestures and speech are considered to be semantically similar when a gesture is temporally aligned with a word or phrase, and both gesture and word/phrase convey the same meaning (cf. Wagner et al., 2014). Based on this definition, a distinction has been made between gestures that convey either *redundant*, *complementary*, *non-redundant*, or *mismatching*¹ semantic content to speech. Most of our gestures are either redundant (e.g. saying "The shelf is long" while gesturing that something is long) or complementary (e.g. saying "The shelf is [this] long" while gesturing the length of the shelf) to speech. Studies with participants from different languages show that the typical structure and semantic content of a language influence the semantic content of gestures (Allen et al., 2007; Kita & Özyürek, 2003), highlighting the usually strong semantic similarity between gesture and speech.

However, sometimes the semantic content of gestures and speech does not overlap, and is thus non-redundant in general (Goldin-Meadow et al., 1993). Examples of non-redundant semantic content are a child who points to a cup while saying that they are thirsty, or a teacher who explains two strategies for a problem at the same time: One in speech, and the other in gestures. In these examples, the semantic content of gestures and speech does not overlap, but their meaning is related and falls within an overarching theme (resp. "drinking", and

¹ Studies differ in whether they differentiate between non-redundant or mismatching content (Wagner et al., 2014).

“problem solutions”). Mismatches between gestures and speech are a specific kind of non-redundant semantic content. As previously described, mismatches are known to occur when a child (or adult) learns a new strategy for a difficult, cognitive problem (e.g. Church & Goldin-Meadow, 1986; Goldin-Meadow et al., 1993; Goldin-Meadow, 2003). Similar to non-redundant semantic content, the meaning of gestures and speech during mismatches does not overlap, but is related.

Complexity matching

Notwithstanding the impact and relevance of the synchronization examples above, involving temporal alignment and semantic similarity, complex systems in the real world often do not synchronize as one-to-one matching of behavior (Delignières et al., 2016). Complex systems, such as gestures and speech, can synchronize on many (time) scales of organization, which is called *complexity matching* (West et al., 2008; Stephen et al., 2008; see also Abney, 2016; Abney, Paxton, et al., 2014; Den Hartigh et al., 2018). During complexity matching, the information exchange between complex systems is maximized (West et al., 2008). Complexity matching occurs when both systems are complex, and the degree of the two systems' complexity is similar.

Gestures and speech as complex systems. Gestures and speech are complex systems. They consist of many different and interacting components and scales, and involve coordination of all these different components and scales of a system over time (e.g., Van Orden et al., 2003). Gestures' and speech's scales range from action potentials of neurons to overarching conversational goals, and beyond (see also De Jonge-Hoekstra et al., 2016). For example, numerous muscles and bones in a person's arms, chest, and even legs, as well as the lungs and central nervous system are involved in each gesture. Importantly, speaking also involves a large number of components; it is estimated that we use more than 70 muscles for each syllable that we utter (e.g., Turvey, 2007).

Infants clearly show how complex gesturing and speaking actually is. Before the first pointing gestures emerge, infants have learned to control their eye movements to focus on an object (Adolph & Franchak, 2017), to use their hands to grasp things, and have learned about distances by crawling forward (Clearfield, 2004). All these actions and perceptions, which are great coordinative accomplishments in themselves, come together in their first pointing gestures. When infant's first words emerge, infants partly 'build' on what they had accomplished for their first gestures (Esteve-Gibert & Prieto, 2014; Goldin-Meadow, 2007). However, uttering their first word involves another set of challenges too. Coordinating all different components to pronounce a specific syllable is a complex task as well, and it usually takes an infant many tries before they grasp the correct configuration. This process is nicely illustrated by Roy (Roy

Task difficulty's influence on gesture-speech synchronization

et al., 2015; also see Roy, 2011), who showed how his son went from saying 'gaaaa' to the word 'water', over numerous trials, in about 6 months' time.

Complexity and fractal scaling. A complex system's coordination over time (e.g. Van Orden et al., 2003) can be more or less fluent. When the coordination of components and layers of a system over time is fluent, the changes of behavior at all different scales are related (e.g. Wijnants, 2014). In other words, variability across time scales is related and dependent, which means that changes on smaller time scales (e.g. neuronal level) influence changes on larger time scales (e.g. conversational goals) and vice versa. If one would plot that system's behavior over time (e.g. time between word onsets during an affective conversation), one would see that small changes in the time series (visible as small *waves*) are nested within larger changes (larger *waves*) (see Figure 2, panel a, for an example). Furthermore, if one would zoom in or out, the plotted time series would look similar at different levels of magnification. In other words, the variability at the level of milliseconds looks like the variability at the level of seconds, which looks like the variability at the level of minutes, etc. Objects that show such self-similarity, such as the Koch snowflake (Figure 2, panel c) or Romanesco broccoli (Figure 2, panel d) are also called *fractal* objects. Similarly, a nested, and self-similar², structure of variability in the temporal domain is called (*mono*)*fractal* or *pink noise* (see Figure 2, panel a). Monofractal variability has been proposed as an index of optimal balance between rigid and random behavior, and is often found in complex systems that change over time (Van Orden et al., 2011; Van Orden et al., 2003; Wijnants, 2014). Indeed, many studies found that expert performance on repetitive motor tasks is more *pink* than non-expert behavior (e.g. Den Hartigh et al., 2015; Kloos & Van Orden, 2010; Van Orden et al., 2011). Monofractal variability has thus been considered as an identifying feature of complex systems, corresponding to a systems' degree of complexity.

However, different from relatively repetitive motor tasks, more diverse human behavior shows sudden jumps, and periods of relative stability mixed with intermittent bursts of variability (Dixon et al., 2012; Ihlen & Vereijken, 2010; Kelty-Stephen et al., 2013; Stephen et al., 2012). Moreover, these increases in variability have been related to transitions, which are a hallmark of human (and other complex systems') development. Examples of a sudden jump, which would go along with a burst in variability, are an abrupt change in conversation goals, or the ("aha"-)moment of acquiring new understanding (Dixon et al., 2012). Delignières et al. (2016), Dixon et al. (2012), Ihlen and Vereijken (2010), Kelty-Stephen et al. (2013), and Stephen et al. (2012) argue that timescales *themselves* also interact, and that these interactions between

² Strictly speaking, time series' variability usually is self-affine instead of self-similar, because its dimensions are scaled by different amounts in the x- and y-directions. For purposes of brevity and clarity, we will use the term self-similar throughout the paper.

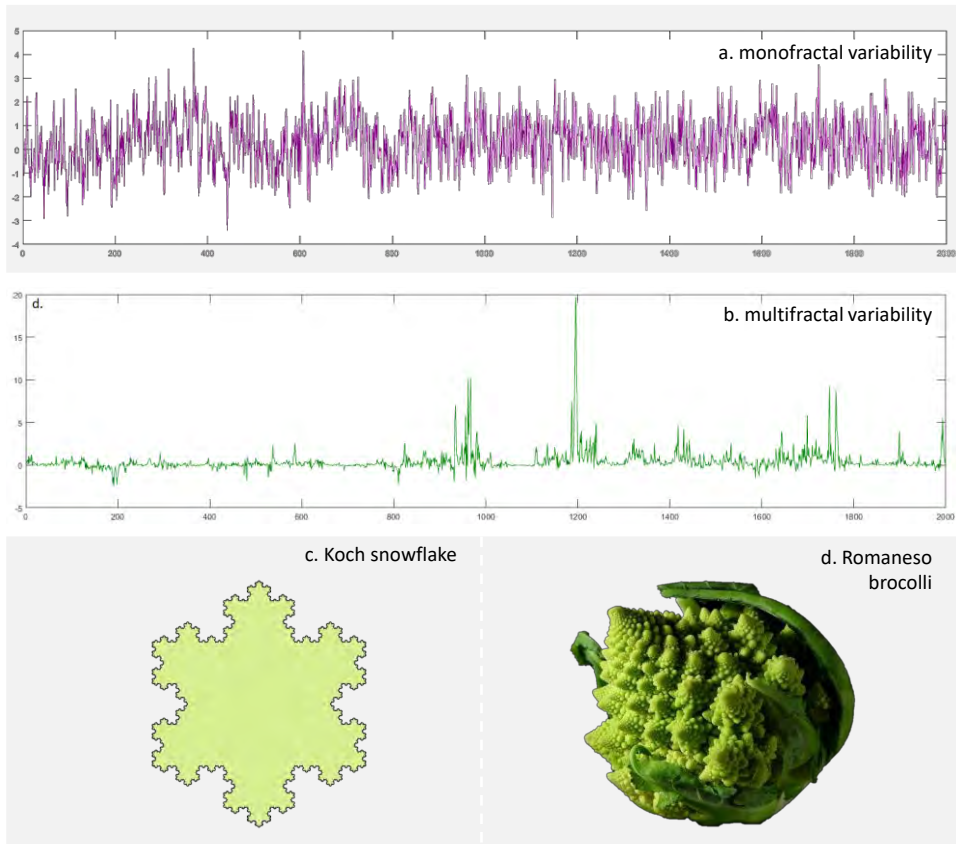


Figure 2. Examples of fractal structures. Panel a shows a timeseries with a monofractal structure of variability (source: script in [doi:10.3389/fphys.2012.00141](https://doi.org/10.3389/fphys.2012.00141)). Panel b shows a timeseries with a multifractal structure of variability, whereby periods of monofractal variability are intermitted by periods of large fluctuations and periods of small fluctuations (source: script in [doi:10.3389/fphys.2012.00141](https://doi.org/10.3389/fphys.2012.00141)). Panel c displays the Koch snowflake (7th iteration; source: bit.ly/2PGerAd). Panel d displays Romanesco broccoli (source: bit.ly/2wiEccN). The monofractal structures in panel a, c and d are self-similar, which means that they look the same at different levels of magnification. The multifractal structure in panel d is less self-similar.

timescales lead to these large changes in variability (for a clear and more in-depth explanation, please see Kelty-Stephen et al., 2013). When variability with a monofractal (pink noise) structure is mixed with periods of changes in variability, these time series display a *multifractal* structure (see Figure 2, panel b). Therefore, identifying complex systems and establishing a system's degree of complexity should also incorporate multifractal variability (Delignières et al., 2016; Dixon et al., 2012; Ihlen & Vereijken, 2010; Kelty-Stephen et al., 2013; Stephen et al., 2012).

Task difficulty's influence on gesture-speech synchronization

When does complexity matching occur? As previously described, complexity matching means that the degree of system's complexity is similar, due to coupling. In other words, when coupled systems match their complexity, the fractal structure of their temporal variability is alike.

Circumstances influence complexity matching. Abney, Paxton et al. (2014) found that type of conversation influences whether complexity matching between two participants in speech occurs. Specifically, when participants discussed things that they had in common, the fractal scaling of participants' acoustic onset events was similar, and their speech thus showed complexity matching. However, no complexity matching was found when participants who discussed issues on which they had different opinions. Furthermore, Almurad et al. (2017) investigated complexity matching between participants who were instructed to walk in synchrony. Participants walked either side-by-side or arm-in-arm (or independently), and the researchers measured the duration of the intervals between their strides. Participants in both (non-independent) conditions showed high levels of complexity matching, whereby arm-in-arm walking led to slightly higher levels of complexity matching than walking side-by-side. With regard to manual coordination between participants in terms intervals between finger taps, fractal hand movements, and a larger magnitude of hand movement's variation, leads to stronger complexity matching between a leader and a follower than random hand movements (Coey et al., 2016). In addition, complexity matching is stronger when participants coordinate movements of both their hands, than when they coordinate the movements of one of their hands to those of a partner (Coey et al., 2018). Most of these studies show that stronger coupling between systems goes together with higher levels of complexity matching (Cox, 2016).

Interestingly, research findings are mixed about whether complexity matching is functional in terms of task performance: While Fusaroli et al. (2013) and Abney, Paxton et al. (2014) found better task performance with higher levels of complexity matching, Schloesser et al. (2019) and Abney et al. (2015) found an inverse relation. With regard to gestures and speech, De Jonge-Hoekstra et al. (2016) suggest that difficult tasks may influence whether and how gestures and speech synchronize on multiple scales. This would imply that difficult tasks influence complexity matching between gestures and speech.

Current study

In this study, we investigated how a difference in task difficulty influences the synchronization between participant's gestures and speech, in terms of temporal alignment, semantic similarity, and complexity matching. We asked participants to repeatedly match targets of the same colors presented on a tablet with touch screen, by means of pointing to these targets and saying their location. Participants were assigned to either a predictable, easy condition, or to an unpredictable, difficult condition.

Our first research question is: How does task difficulty influence temporal alignment, semantic similarity, and complexity matching between participant's gestures and speech? With regards to *temporal alignment*, Pouw and Dixon (2019b) found that gestures and speech became more synchronized in the more difficult Delayed Auditory Feedback condition. We therefore expected that gestures and speech would be more synchronized in the difficult than in the easy condition (hypothesis 1A). Regarding *semantic similarity*, Goldin-Meadow and colleagues (e.g. Church & Goldin-Meadow, 1986; Goldin-Meadow et al., 1993; Goldin-Meadow, 2003) found that gestures and speech mismatch in semantic content when people are about to understand a task which they do not understand yet, thus when the task is difficult for them. We therefore expected less semantic similarity between gestures and speech in the difficult than in the easy condition (hypothesis 1B). With respect to *complexity matching*, there are no studies that directly investigated how task difficulty influences complexity matching. As described above, we do know that the level of complexity matching increases when the coupling between systems is stronger (Abney, Paxton et al., 2014; Almurad et al., 2017; Coey, 2016, 2018). Our previously stated hypothesis 1A suggests that gestures and speech become more temporally aligned in the difficult condition, and thus a *stronger* coupling. However, our previously stated hypothesis 1B suggests that gestures and speech become less semantically similar in the difficult condition, and thus a *weaker* coupling. Because of this contradiction, we have no specific hypothesis for the influence of task difficulty on the level of complexity matching between gestures and speech.

Our second research question is: How are temporal alignment, semantic similarity, and complexity matching between gestures and speech related in the easy and difficult condition? Bergmann et al. (2011) found that gestures and speech were more synchronized in time when their semantic content was more similar. This suggests that a higher temporal alignment between gestures and speech would go together with a higher semantic similarity. On the other hand, hypotheses 1A and 1B suggest a higher temporal alignment and a lower semantic similarity in the difficult condition. We therefore expected a positive relation between gestures' and speech's temporal alignment and semantic similarity in the easy condition (hypothesis 2A), and a negative relation between temporal alignment and semantic similarity in the difficult condition (hypothesis 2B). In line with a higher level of complexity matching when coupling is stronger (Abney, Paxton et al., 2014; Almurad et al., 2017; Coey, 2016, 2018), and in line with hypothesis 2A (positive relation between temporal alignment and semantic similarity in easy condition), for the easy condition we expected a positive relation between gestures' and speech's temporal alignment, semantic similarity, and complexity matching as well (hypothesis 2C). Our expected negative relation between temporal alignment and semantic similarity (hypothesis 2B) in the difficult condition suggests an inverse relation in coupling strength.

Task difficulty's influence on gesture-speech synchronization

Therefore, we have no specific hypotheses about how complexity matching is related to either temporal alignment or semantic similarity in the difficult condition.

Our third research question is: How do temporal alignment, semantic similarity, and complexity matching between gestures and speech predict task performance? We assessed task performance in terms of *time needed to finish the task*. Our experimental manipulation of task difficulty will influence task performance, as difficult tasks typically take longer to perform. Therefore we controlled for the influence of condition (task difficulty) when we investigated whether temporal alignment, semantic similarity, and complexity matching between gestures and speech predict task performance. According to Iverson and Thelen (1999; also see Butcher & Goldin-Meadow, 2000; Esteve-Gibert & Guellaï, 2018) the temporal alignment between gestures and speech becomes higher when infants and toddlers grow older. As children's language skills change and become more advanced during that time too (e.g. Tamis-LeMonda et al., 1998; Tamis-LeMonda et al., 2001), this could imply that more temporal alignment goes together with a better language performance. Mismatches are a form of semantic dissimilarity, and predict better performance on *subsequent* tasks (e.g. Church & Goldin-Meadow, 1986; Goldin-Meadow et al., 1993; Goldin-Meadow, 2003). Findings about a link between complexity matching and task performance are mixed, whereby some studies found a positive relation (Abney, Paxton et al., 2014; Fusaroli et al., 2013) while others found a negative relation (Abney et al., 2015; Schloesser et al., 2019). Taken together, these findings are not sufficiently conclusive to formulate hypotheses about how temporal alignment, semantic similarity, and complexity matching predict task performance.

Method

Participants

We included³ 30 participants (20 F, 10 M) between 18 and 27 years ($M = 20.70$, $SD = 2.39$) in our study. All participants were students with a Dutch nationality at a University in the Netherlands, who participated in the experiment in exchange for course credit or monetary compensation. The participants provided written consent. The ethical committee of Psychology department of the University of Groningen approved of the study.

³ We recruited a total of 59 participants to participate in this experiment. However, due to technical issues with the tablets, for 29 participants the audio was either not recorded, or recorded with insufficient quality (e.g. loud ticks on the screen, background noise). After rigorous checks of the quality of all the audio recordings, we decided to include the 30 participants of which the audio-recordings were of high quality. For the analyses that we will conduct, with many data points, a sample of 30 participants is sufficiently large. We have the pointing data for all 59 participants, and we will use this data for other studies and research questions that do not involve speech.

Materials

Participants performed the task on a tablet (Lenovo MIIX 320-10ICR 1.44GHz x5-Z8350) with a 10.1 inch touchscreen (1280 x 800 pixels) and Windows 10 operating system. To facilitate pointing, the tablet was positioned in a 45 degree angle from the table using a tablet stand (see Figure 3). The experiment was programmed using OpenSesame [version 3.0.0] (Mathôt et al., 2012), which is an open-source program to build (social science) experiments. Using OpenSesame, we could run the task at the tablet (a detailed description follows below), and simultaneously record the time and x- and y-coordinates of participants' pointing (touching) at the screen, as well as participants' speech-signal.

Participants' speech was recorded at 44.1 kHz using a basic, hands-free microphone that was plugged into the 3.5mm audio jack of the tablet. We used Audacity [version 2.2.2] to normalize the volume of the speech-signal and to filter out background noise. Furthermore, we used Praat (Boersma & Weenink, 2018) [version 6.0.42] and RStudio [version 1.1.456] to calculate the *amplitude envelope* of the speech signal (resp. He & Dellwo, 2016; Pouw & Trujillo, 2019; a detailed description follows below). The amplitude envelope that is calculated by the R-script is identical to the amplitude envelope that is calculated by the Praat-script (Pouw & Trujillo, 2019). We used a custom script in Matlab [version 2018a] to identify the start of syllables in the speech signal, and to cut the audio recordings into smaller parts of one syllable each (a detailed description follows below). We used OpenSesame [version 3.0.0] (Mathôt et al., 2012) to manually code the semantic content of these syllables.

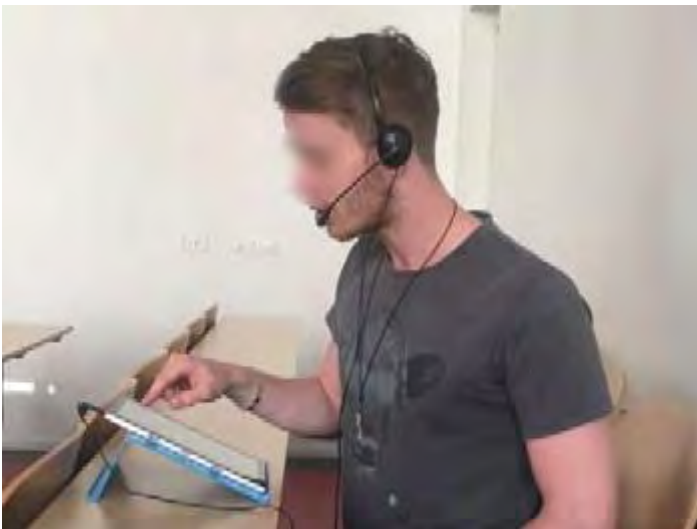


Figure 3. Set-up of experiment.

Task difficulty's influence on gesture-speech synchronization

We used Matlab to carry out the analyses on the time series of pointing and the amplitude envelope of speech. We specifically used the MFDFA-package by Ihlen (2012) to perform Multifractal Detrended Fluctuation Analysis, to estimate the *temporal multifractality* of participant's gestures and speech. Furthermore, we used RStudio to carry out inferential statistics, and the R-package ggplot2 (Wickham, 2016) to create plots of our data.

Procedure

Participants performed a tablet task (see Figure 3 and 4), which can be found here: osf.io/dj5vr/ (Scripts & Materials > Tablet task). We instructed the participants to repeatedly (virtually) put a ring on a bar of the same color, by first pointing (touching) to the ring on screen and thereafter to the top of the corresponding bar. Furthermore, each time that a participant pointed, we instructed them to say out-loud the location of the ring and bar (left, middle, right) that they were pointing to, in Dutch ("links", "midden", "rechts", resp.). In addition, we instructed participants to perform the task as fast and accurate as possible (in accordance with Fitts, 1954). We randomly assigned the participants to either the easy ($n = 14$; see Figure 4, left panel) or the difficult condition ($n = 16$; see Figure 4, right panel). In the easy condition, the color of

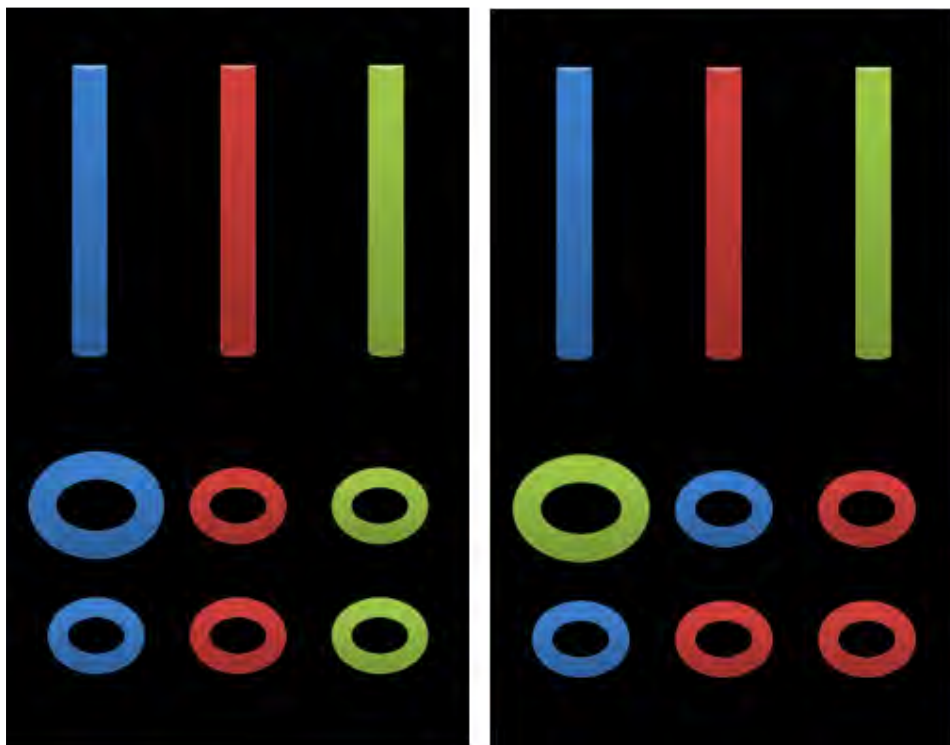


Figure 4. Example of tablet task. The left panel displays the easy task, and the right panel displays the difficult task.

the ring always corresponded to the color of the above bar (see Figure 4, left panel). In the difficult condition however, the color of the rings was random (see Figure 4, right panel). Participants were not informed about the pattern being either random or non-random. Since it is impossible to understand a random pattern, participants in the difficult condition constantly needed to reorganize to the new spatial arrangement. This state of reorganization shares similarities with the state of reorganization that precedes learning something new (see Kello et al., 2007; Stephen, Dixon, et al., 2009).

To register participants' pointing, we divided the screen into $3 \times 6 = 18$ (invisible) areas. Each top of the bar was positioned in an area at the top row of the screen, while each ring was positioned in an area at the second row from the bottom of the screen. The correct ring, that is, the ring that participants needed to point to during that trial, appeared larger on screen, as shown in Figure 4 (upper left ring in both panels). Please note that the participants did not have to point to the correct ring or bar for the task to proceed. However, if participants failed to click on a *ring-area* or a *top of the bar-area*, the task did not proceed and the time and location of every first error was recorded.

During the task, the order in which the rings were presented alternated between left to right and right to left. For example, the correct order of the task in the left panel of Figure 4 would be: [first row] left – left – middle – middle – right – right – [second row] right – right – middle – middle – left – left. The correct order of the task in the right panel of Figure 4 would be: [first row] left – right – middle – left – right – middle – [second row] right – middle – middle – middle – left – left. Each time a participant finished with the last ring of a row, that row disappeared from screen, the second row moved up, and a new row appeared at the bottom of the screen. The participants performed 540 repetitions of the task, which is identical to 180 rows of 3 rings and corresponding bars, or a total of 1080 times pointing and saying the location of either a ring or a bar. Before starting with the actual task, the participants completed a trial phase with 15 repetitions of the task, to get used to the set-up. The recordings of this trial phase were not included in the analysis.

Data preparation

To investigate the coupling between participants' gestures and speech, we recorded the time (ms), location (left/middle/right) and position (x- and y-coordinates) of their pointing, and their speech signal.

Task difficulty's influence on gesture-speech synchronization

Gestures

For gestures, the above resulted in a time series⁴ of a) the duration between pointing to rings and bars, and vice versa, b) a time series of the location of the pointing, and c) a time series of distances between the exact locations that participants pointed to. With regards to distances between rings and bars, there are three possible distances⁵ that participants' fingers needed to travel while pointing: 1) a short distance of 608 pixels, when the ring and bar are vertically aligned, 2) a middle distance of 664 pixels, when the ring and bar are one location off (i.e. from the left ring to the middle bar), or 3) a long distance of 809 pixels, when the ring and bar are two locations off (i.e. from the left ring to the right bar). This third, long distance can only occur in the difficult condition, and therefore the frequency distribution of distances between targets differs between the two conditions.

From the work by Fitts (1954) we know that the distance (D) between targets, combined with the width (W) of targets (ring: 167 pixels; bar: 61 pixels), influences how difficult the movement between two targets (i.e. from ring to bar or vice versa) is to perform. Fitts referred to this as the Index of Difficulty (ID), which is given by the following formula: $ID = \log_2\left(\frac{2D}{W}\right)$. Using this formula, from ring to bar the index of difficulty for the short, middle and long distance is 4.317, 4.444, and 4.729, respectively. From bar to ring the index of difficulty for the short, middle and long distance is 2.864, 2.991, and 3.276, respectively. In the current study, we aim to manipulate task difficulty by changing the overall task demand of matching targets of the same color when one of the targets' color was either random (difficult) or non-random (easy). However, any difference in movement time could potentially result from the difference in ID between targets. To remove this possible confound, and standardize this influence of the ID on each duration in our movement timeseries, we divided each duration between pointing to two targets (Movement Time; MT) with the Index of Difficulty of that particular movement. These corrected durations between pointing to two targets corrected with the Index of Difficulty of each movement yielded a time series of MT/ID.

Speech

We recorded participant's speech from the moment that the first experimental trial was presented until the moment that the participant finished with the last experimental trial. This yielded one long sound recording of what the participant said during the task. To increase the quality of the sound recording, we used Audacity to normalize the sound volume and to filter out background noise. We subsequently used PRAAT (He & Dellwo, 2016) or R (Pouw & Trujillo,

⁴ A time series is a sequence of datapoints in chronological order.

⁵ The distances are calculated between the middle of the ring-area and the middle of the top of the bar-area.

2019) to calculate the *amplitude envelope* of the speech signal. The amplitude envelope basically is a smoothed outline of a speech signal's intensity (He & Dellwo, 2016), and its structure corresponds to the lower lip kinematics (He & Dellwo, 2017). In addition, we calculated the velocity of the speech signal's amplitude envelope, which captures how the amplitude envelope increases and decreases.

We identified the start of syllables by extracting the peaks in velocity of the amplitude envelope, using a custom MATLAB script (osf.io/dj5vr/; Scripts & Materials > Data preparation), and saved the audio between two velocity peaks as audio segments (i.e. one syllable per audio segment). The Dutch word “links” has one syllable, “midden” has two syllables, and “rechts” has one syllable. Due to individual differences in speaking, extracting one word or syllable per audio segment did not work perfectly for each participant, however⁶. To ensure that MATLAB was not too sensitive, so as to cut one syllable into multiple audio segments, yet sensitive enough, so as to aggregate a maximum of five words into one audio segment, we manually tweaked a sensitivity parameter in the script (osf.io/dj5vr/; Scripts & Materials > Data preparation) for each audio recording. We subsequently coded the semantic content of the audio segments to identify the starting times of actual words.

We coded the semantic content of the audio segments using OpenSesame (osf.io/dj5vr/; Scripts & Materials > Data preparation). We loaded the audio segments into OpenSesame and coded whether a segment was A) [the first half of] “links”, B) [the first half of] “midden”, C) [the first half of] “rechts”, D) the second half of a word, E) a sequence of multiple words, or F) something else (i.e. other speech, a sigh). If a segment was E) a sequence of multiple words, we coded the semantic content of the sequence of words in that segment. This coding of audio segments yielded a time series of word (segments) and their starting time. For E) sequences of multiple words, we used the amount of words in an audio segment to extract the same amount of velocity peaks of the amplitude envelope in that particular audio segment. We replaced the word sequences in the time series with the individual words and their velocity peaks. We removed the F) other speech/sighs from the time series.

⁶ Some participants pronounced a very loud “s” at the end of “links” and “rechts”, and therefore the MATLAB script identified two syllables within these words, instead of one. Conversely, some participants mumbled the word “midden” (which is quite typical for people from the Northern part of the Netherlands), and therefore the MATLAB script identified one syllable within this word, instead of two. In addition, participants differed in their range of speech amplitude during the task: Some spoke evenly loud during the whole task, while others intermitted softer and louder periods of speaking. Therefore, for some participants, a velocity peak in a softer part of the audio recording is not recognized as a velocity peak in a louder part of the audio recording. This resulted in MATLAB identifying multiple words as one syllable in the loud periods of speaking, and multiple words per audio segment in the softer periods.

Task difficulty's influence on gesture-speech synchronization

Combining gestures and speech

To investigate the temporal alignment and semantic similarity between gestures and speech, we aligned the time series of gestures and speech by linking the gestures to the word that was closest in time. To find the correct delay for each participant, we aligned the time series of gestures and speech for every delay between 10 ms and 1000 ms, with steps of 10 ms, and calculated the amount of semantic content-differences, and the average asynchrony, between gestures and speech (for overview, osf.io/dj5vr/; Data). Since the amount of semantic content differences for each participant went down to a minimum and then went up again, we decided that the delay with the least amount of semantic content-differences was the correct delay. If there were more delays with least amount of semantic content-differences, we picked the delay with the lowest average asynchrony between gestures and speech. The data files with the maximally aligned gestures and speech can be found here: osf.io/dj5vr/; Data > For analyses.

We calculated the difference between *amplitude* peaks (not *velocity* peaks) in the aligned time series to create a duration-time series for speech, and we used this time series to analyze temporal alignment between gestures and speech. The amplitude peak of the amplitude envelope corresponds to the stressed syllable in a word (see Figure 5). In each of the three words that the participants said, the first syllable of the word is stressed ("links", "mid-den",

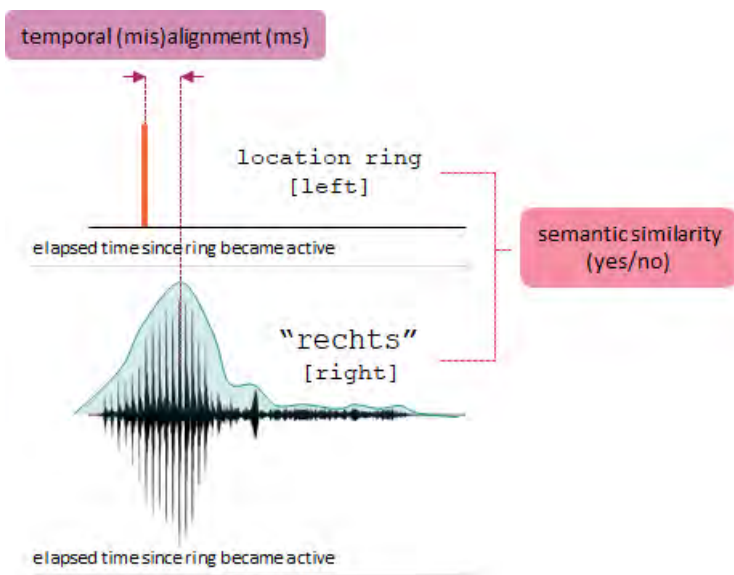


Figure 5. Illustration of how we calculated temporal alignment and semantic similarity within a trial. The orange vertical line indicates the moment the participant's finger touched the screen when the participant pointed at the ring. The peak of the blue curve corresponds to the amplitude peak of the word that the participant said.

“rechts”). The amplitude peak therefore yielded a similar time point for each of the three words. Furthermore, to analyze semantic similarity, we used the semantic content-time series of speech. We divided the duration-time series of speech with the Index of Difficulty for that particular movement between ring and bar or vice versa to create a MT/ID time series for speech. We used this time series for speech to analyze complexity matching between gestures and speech.

Analysis

Calculating temporal alignment

For each trial, from ring to bar or from bar to ring, we know the time between the moment the ring or bar became activated, and a) the moment that participants pointed to and touched a bar or ring, and b) the amplitude peak of the word the participant said to indicate the ring's or bar's location. We compared these durations between the moment of pointing and the moment of the amplitude peak. For each participant, we calculated the average absolute difference between moments of pointing and amplitude peak, and used this as our measure of temporal alignment. Please note that higher values correspond to lower temporal alignment. Figure 5 displays how we estimated temporal alignment and semantic similarity within a trial. To check whether participant's temporal alignment was significantly higher than chance level, for each participant we compared the empirical temporal alignment with the temporal alignment between their repeatedly shuffled durations of gestures and speech.

Calculating semantic similarity

For each trial, from ring to bar or from bar to ring, we know whether participants' pointed to the left, middle, or right object, and which location they mentioned in speech. We compared the location in gestures and in speech location and identified whether they did or did not match. We calculated the sum of mismatches in location, and used this as our measure of semantic similarity. Please note that higher values correspond to lower semantic similarity. To check whether participant's semantic similarity was significantly higher than chance level, for each participant we compared the empirical semantic similarity with the semantic similarity between their repeatedly shuffled (mentioned) location of gestures and speech.

Calculating complexity matching

We applied Multifractal Detrended Fluctuation Analysis (Ihlen, 2012; Ihlen & Vereijken, 2010; Kantelhardt et al., 2002; Wallot et al., 2014) to the time series of gestures and speech. MFDFA is a method to reliably approximate a time series *temporal multifractality*. MFDFA is an extension of Detrended Fluctuation Analysis (DFA), which is a method to reliably approximate

Task difficulty's influence on gesture-speech synchronization

a time series' *temporal fractality*. An accessible explanation of MFDFA can be found in Appendix B.

In short, performing MFDFA on a timeseries yields a so-called multifractal spectrum (see Figure 6; the details of going from timeseries to multifractal spectrum can be found in Appendix B).

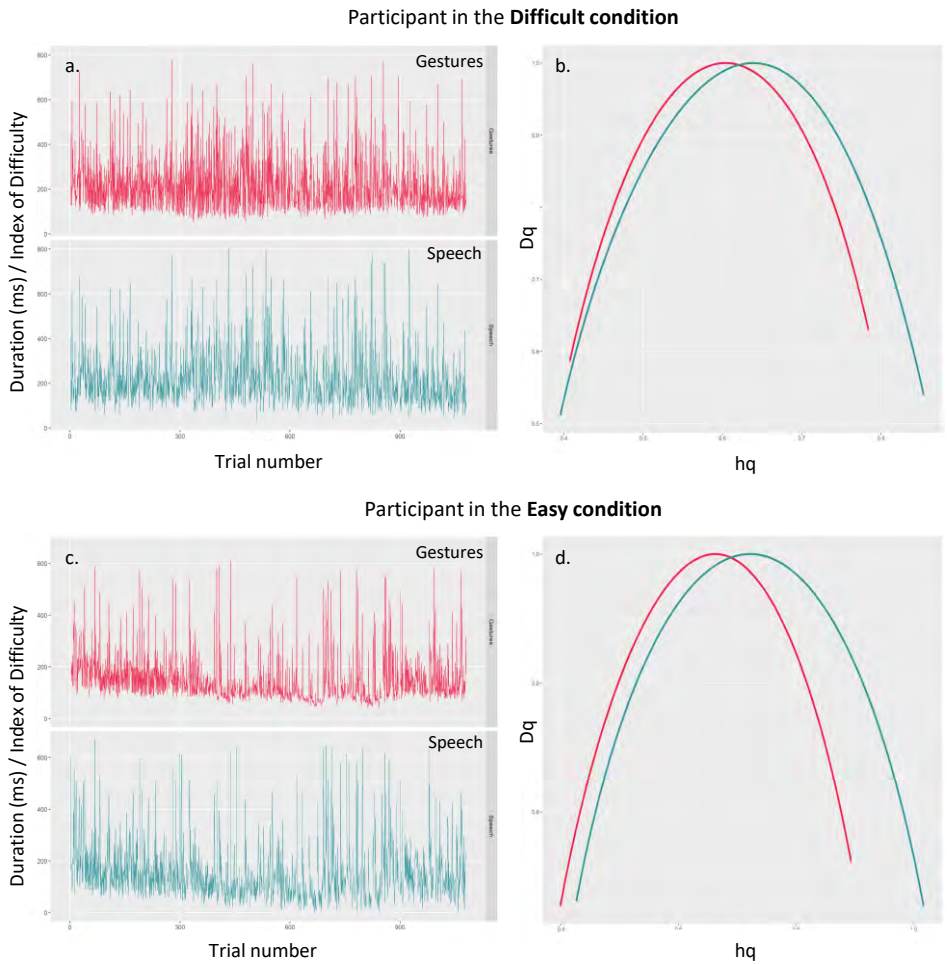


Figure 6. Timeseries of duration (ms) divided by Index of Difficulty (panel a and c), and corresponding multifractal spectrums (panel b and d, resp.), for gestures (red) and speech (blue). Panel a and b illustrate the MT/ID of timeseries gestures and speech and corresponding multifractal spectrums of a participant in the Difficult condition, and panel c and d of a participant in the Easy condition. The difference in multifractal spectrum width is 0.081 for the participant in the difficult condition and 0.096 for the participant in the easy condition. We interpret this as more complexity matching between gestures and speech for the participant in the difficult condition, compared to the participant in the easy condition.

The width of this multifractal spectrum indicates the degree of temporal multifractality of the timeseries, and is a measure of the multifractal structure of the timeseries' variability. In short, a higher degree of multifractal structure leads to a wider multifractal spectrum, while a lower degree of multifractal structure (or higher degree of monofractal structure) leads to a narrower multifractal spectrum. As previously described, complexity matching requires that the fractal structure of variability of the behavior of two complex systems matches. To investigate the degree of complexity matching between gestures and speech, we therefore calculated the difference in gestures' and speech's multifractal spectrum width. To check whether complexity matching between gestures and speech was significant, for each participant we compared the actual difference in multifractal spectrum width with the difference in repeatedly sampled, random pairs of gestures' and speech's multifractal spectrum width.

Monte Carlo permutation testing

We calculated all p -values using Monte Carlo (MC) Permutation tests (Ninness et al., 2002; Todman & Dugard, 2001), because MC permutations tests do not require a specific underlying distribution of the data. By drawing 10,000 random samples from the original data, the probability that differences are caused by chance was measured. We used custom made R scripts to calculate p -values using MC permutation tests (osf.io/dj5vr/; Scripts & Materials).

Results

Descriptives

Participants in the difficult condition performed the task on average within 987 sec. ($SD = 138$ sec.). While they always pointed to the correct location of the bar and ring, they said the incorrect location on average 119.8 out of 1080 trials ($SD = 29.6$), i.e. 11%. A semantic dissimilarity was thus always a combination of a correct gesture and an incorrect utterance. In the difficult condition, gestures' width of the MFDFA-spectrum was on average .473 ($SD = .203$), and speech's width of the MFDFA-spectrum was on average .432 ($SD = .178$).

Participants in the easy condition performed the task on average within 749 sec. ($SD = 151$ sec.). Similar to the difficult condition, they always pointed to the correct location of the bar and ring, but they said the incorrect location on average 45.8 out of 1080 trials ($SD = 47.3$), i.e. 4%. Gestures' width of the MFDFA-spectrum was on average .618 ($SD = .169$), and speech's width of the MFDFA-spectrum was on average .496 ($SD = .104$), in the easy condition.

Task difficulty's influence on gesture-speech synchronization

RQ1: Task difficulty's influence on temporal alignment, semantic similarity, and complexity matching

With regard to temporal alignment, we found significantly less temporal alignment between participants' gestures and speech in the difficult condition ($M = 218.538$ ms, $SD = 43.652$) than in the easy condition ($M = 167.182$ ms, $SD = 62.322$), $p = .009$ ($\Delta_M = 51.356$, 95% $CI_{\Delta-MC} = -34.598, 35.322$), with a large effect size, $d = .955$ (see Figure 7, left panel). This finding is opposite from our hypothesis 1A, as we expected that gestures and speech would be more temporally aligned in the difficult than in the easy condition. For all participants, the empirically observed temporal alignment between gestures and speech throughout the task was significantly higher than the temporal alignment between random pairs of their gestures' and speech's duration ($p < .001$).

For semantic similarity, we found significantly less semantic similarity between participants' gestures and speech in the difficult condition ($M_{mismatches} = 119.750$, $SD = 47.301$) than in the easy condition ($M_{mismatches} = 45.769$, $SD = 29.601$), $p < .001$ ($\Delta_M = 73.981$, 95% $CI_{\Delta-MC} = -32.661, 32.506$), with a very large effect size, $d = 1.875$ (see Figure 7, center panel). This finding is in line with our hypothesis 1B, as we expected less semantic similarity between gestures and speech in the difficult than in the easy condition. For all participants, the empirically observed semantic similarity between gestures and speech throughout the task was significantly higher than the semantic similarity between random pairs of their gestures' and speech's semantic content ($p < .001$).

With regard to complexity matching, we found more complexity matching between gestures and speech for participants in the difficult condition ($M_{diff. MFDDFA-spectrum width} = 0.065$, $SD = 0.049$) than in the easy condition ($M_{diff. MFDDFA-spectrum width} = 0.123$, $SD = 0.102$), $p = 0.026$ ($\Delta_M = -.058$, 95% $CI_{\Delta-MC} = -0.049, 0.047$), with a medium to large effect size, $d = .726$ (see Figure 7, right panel). When we visually inspected the density plot, participants in the difficult condition showed a

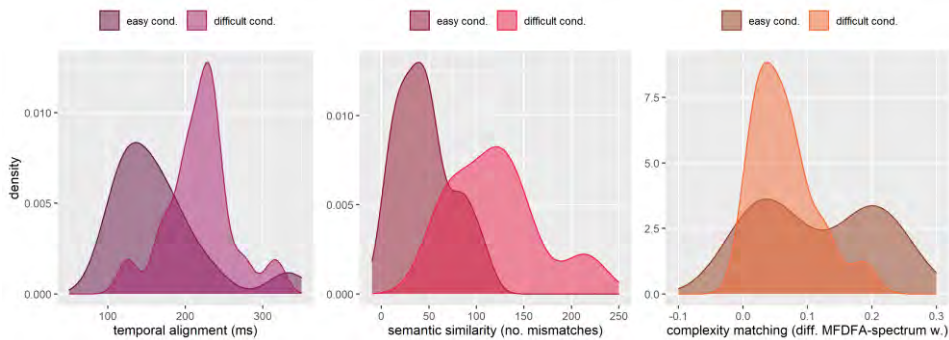


Figure 7. Density plots of temporal alignment, semantic similarity, and complexity matching in the difficult and easy condition.

striking peak around 0.0 and 0.1 in difference of MFDFA-spectrum width. However, participants in the easy condition showed no clear peak in difference in MFDFA-spectrum width, but instead showed a wide range of values. In line with this, for 15 out of 16 participants in the difficult condition we found the difference in MFDFA-spectrum width to be significantly smaller ($p < .05$) than the difference in MDFA-spectrum between random pairs of participants' gestures and speech, while we found this to be true for only 8 out of 14 participants in the easy condition. Note that we did not make a prediction about the difference in complexity matching between the two conditions.

RQ2: Relations between temporal alignment, semantic similarity, and complexity matching

In the difficult condition, we found a significant, moderate, positive correlation between average temporal alignment (ms) and semantic similarity (no. of gesture-speech mismatches), $r = .555$, $p = .014$ (95% $CI_{r_{MC}} = -.422, .433$; see Figure 8, panel a). This finding is opposite from our hypothesis 2B, as we expected a negative relation between temporal alignment and semantic similarity in the difficult condition. We found a significant, moderate, negative correlation

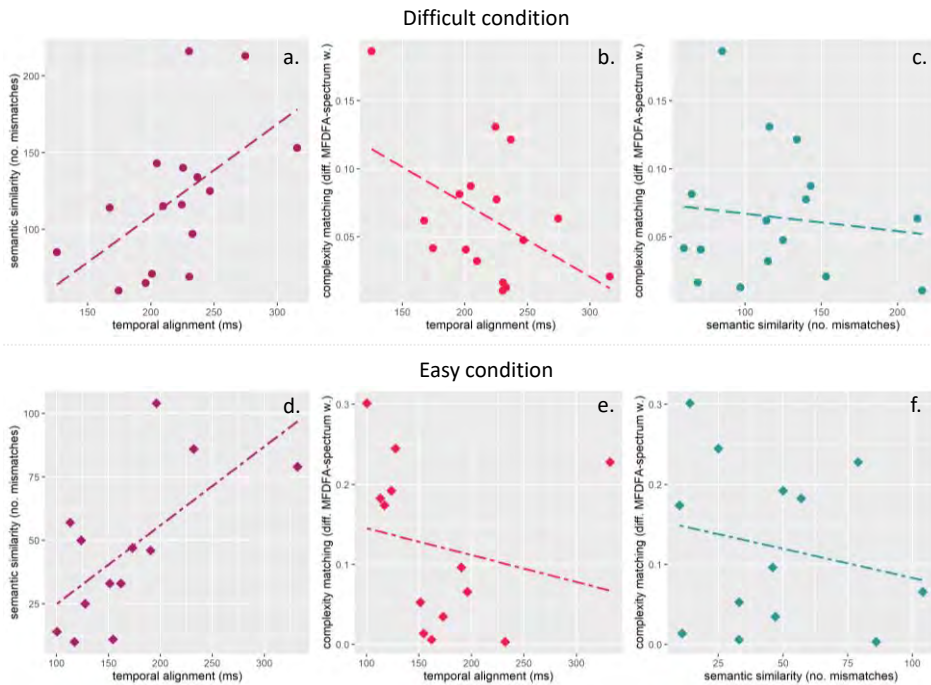


Figure 8. Scatterplots of relations between the variables temporal alignment (ms), semantic similarity (no. of mismatches), and complexity matching (difference in MFDFA-spectrum width). Panels a, b, and c display the relations in the difficult condition; panels d, e, and f display the relations in the easy condition.

Task difficulty's influence on gesture-speech synchronization

between average temporal alignment (ms) and complexity matching (difference in MFDFA-spectrum width), $r = -.481, p = .031$ (95% $CI_{r_{MC}} = -.430, .433$; see Figure 8, panel b). We did not state a hypothesis about the relation between temporal alignment and complexity matching. We found a non-significant, low, negative correlation between semantic similarity (no. of gesture-speech mismatches) and complexity matching (difference in MFDFA-spectrum width), $r = -.125, p = .336$ (95% $CI_{r_{MC}} = -.414, .448$; see Figure 8, panel c). We did not state a hypothesis about the relation between semantic similarity and complexity matching. An overview of our findings with regards to research question 2 can be found in Figure 9.

In the easy condition, we found a significant, moderate, positive correlation between average temporal alignment (ms) and semantic similarity (no. of gesture-speech mismatches), $r = .653, p = .013$ (95% $CI_{r_{MC}} = -.438, .511$; see Figure 8, panel d). This finding is in line with our hypothesis 2A, as we expected a positive relation between temporal alignment and semantic similarity in the easy condition. We found a non-significant, low, negative correlation between average temporal alignment (ms) and complexity matching (difference in MFDFA-spectrum width), $r = -.205, p = .269$ (95% $CI_{r_{MC}} = -.444, .489$; see Figure 8, panel e). This finding is not in line with our hypothesis 2C, as we expected a positive relation between temporal alignment and complexity matching. We found a non-significant, low, negative correlation between semantic similarity (no. of gesture-speech mismatches) and complexity matching (difference in MFDFA-spectrum width), $r = -.211, p = .253$ (95% $CI_{r_{MC}} = -.475, .477$; see Figure 8, panel f). This finding is not in line

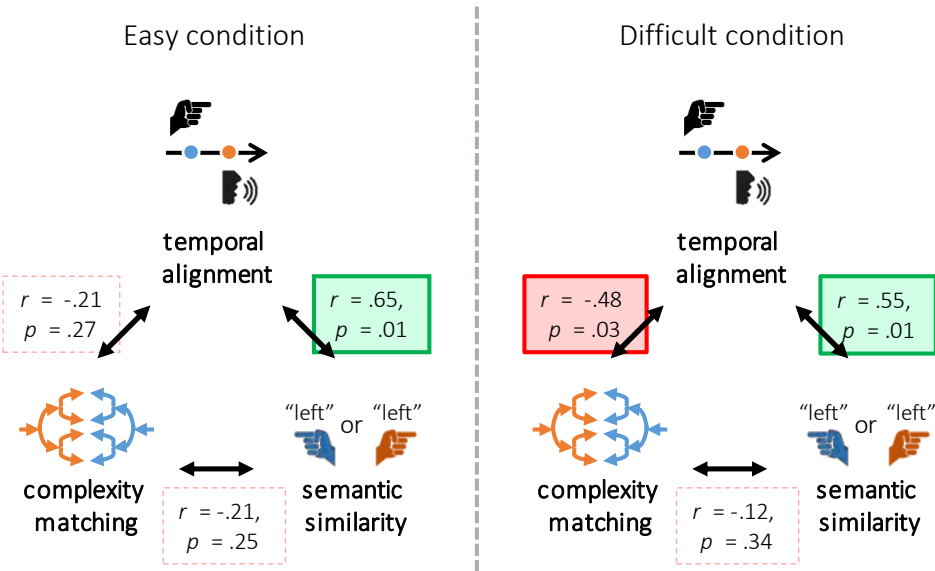


Figure 9. Overview of the empirical relations between temporal alignment, semantic similarity and complexity matching, in the easy and difficult condition.

with our hypothesis 2D, as we expected a positive relation between semantic similarity and complexity matching.

RQ3: Predict task performance with temporal alignment, semantic similarity, and complexity matching

We performed a multiple linear regression to predict task performance (total time), based on temporal alignment, semantic similarity, and complexity matching.

With regard to the individual variables, greater temporal alignment significantly predicted better (i.e. a more speedy) task performance than condition alone, with R^2 increasing from .423 to .616, $p < .001$ ($\Delta_{R^2} = .192$, 95% $CI_{\Delta-MC} = .000, .082$). Less semantic similarity did not significantly predict better task performance than condition alone, with R^2 increasing from .423 to .425, $p = .764$ ($\Delta_{R^2} = .002$, 95% $CI_{\Delta-MC} = .000, .082$). Less complexity matching did not significantly predict better task performance than condition alone, with R^2 increasing from .423 to .456, $p = .214$ ($\Delta_{R^2} = .033$, 95% $CI_{\Delta-MC} = .000, .079$).

Given that temporal alignment was a predictor of performance with only condition in the model, we asked whether semantic similarity and complexity matching would contribute additional unique variance. When semantic similarity was included in the model with condition and temporal alignment, we obtained a significant increase in R^2 from .616 to .734, $p = .003$ ($\Delta_{R^2} = .118$, 95% $CI_{\Delta-MC} = .000, .057$), whereby greater temporal alignment and less semantic similarity significantly predicted task performance. When we added complexity matching to the model containing condition and temporal alignment, we obtained a non-significant increase in R^2 from .616 to .619, $p = .628$ ($\Delta_{R^2} = .004$, 95% $CI_{\Delta-MC} = .000, .057$). Furthermore, when we added complexity matching to the model containing condition, temporal alignment, and semantic similarity, we obtained a non-significant increase in R^2 from .734 to .737, $p = .601$ ($\Delta_{R^2} = .003$, 95% $CI_{\Delta-MC} = .000, .040$).

Discussion

In this study, we investigated how a difference in task difficulty influences the synchronization between participant's gestures and speech, in terms of temporal alignment, semantic similarity, and complexity matching.

Summary of results

Our first research question was: How does task difficulty influence temporal alignment, semantic similarity, and complexity matching between participant's gestures and speech? We found significantly less temporal alignment, less semantic similarity and more complexity matching in the difficult condition than in the easy condition. With regard to complexity

Task difficulty's influence on gesture-speech synchronization

matching, we additionally observed a more peaked distribution of differences in MFDFA-spectrum widths in the difficult condition, while the distribution was clearly flatter in the easy condition. This suggests that, for participants in the difficult condition, the fractal structure of variability of gestures' and speech' matches to a similar degree, which also points to complexity matching. Participants in the easy condition show a more variable degree of this matching, so no clear complexity matching.

Our second research question was: How are temporal alignment, semantic similarity, and complexity matching between gestures and speech related in the easy and difficult condition? In the difficult condition, we found (a) a moderate and significant positive relation between temporal alignment and semantic similarity, (b) a moderate and significant negative relation between temporal alignment and complexity matching, and (c) a low and nonsignificant negative relation between complexity matching and semantic similarity. In the easy condition, we found (A) a moderate and significant positive relation between temporal alignment and semantic similarity, (B) a low and nonsignificant negative relation between temporal alignment and complexity matching, and (C) a low and nonsignificant negative relation between complexity matching and semantic similarity.

Our third research question was: How do temporal alignment, semantic similarity, and complexity matching between gestures and speech predict task performance? With regard to individual variables, we found that temporal alignment significantly predicted task performance, whereby more temporal alignment went together with better (i.e. a more speedy) task performance. Neither semantic similarity nor complexity matching significantly predicted task performance. With regard to combinations of variables, we found that temporal alignment and semantic similarity together predicted task performance better than temporal alignment alone, whereby more temporal alignment and less semantic similarity went together with better task performance. Adding complexity matching to the model did not significantly increase the model's exploratory power.

Phase synchronization

When two (weakly) coupled *oscillating* systems interact, their rhythm adjusts and their frequency entrains. This phenomenon is called *phase synchronization* (e.g. Pikovsky et al., 2003; Warren, 2006), and results in temporal alignment. We have viewed gestures and speech as two coupled systems throughout this paper. Akin to oscillating systems, we observed that participants in the easy condition rapidly got into a regular rhythm of gesturing and speaking. However, participants in the difficult condition struggled to get into and maintain a rhythm. In line with the higher temporal alignment that we found in the easy condition, we believe that participant's gestures and speech also exhibited phase synchronization in the easy condition.

Similarly, Pouw, Harrison et al. (2019) found that rhythmical arm beating, but not wrist beating, entrained the amplitude envelope of speech. Although less pronounced than beating, participants in the easy condition of the current study also rhythmically moved their arm.

Pouw and Dixon (2019b) investigated temporal alignment between gestures and speech while participants told a story. As previously described, Pouw and Dixon (2019b) found an increase in temporal alignment between participants' gestures and speech under Delayed Auditory Feedback. Delayed Auditory Feedback is a delayed stimulus that entrains both gestures and speech, and gestures and speech become more synchronized to each other because they are entrained together. Pouw and Dixon (2019b) reasoned that Delayed Auditory Feedback perturbs hand movements and speech, and that the increase in gesture-speech synchrony is a way to stabilize rhythmic activity (such as gestures and speech) under disrupting circumstances (also see Pikovsky et al., 2001), i.e. "stability through synergy" (Pouw & Dixon, 2019b, p. 28).

While the difficult task in our study did *disrupt* gestures' and speech's rhythm, task difficulty did not *entrain* gestures and speech. The nature of our perturbation was different from Pouw and Dixon (2019b), and indeed we did not find more temporal alignment in the difficult condition than in the easy condition. However, we did find more complexity matching in the difficult condition than in the easy condition. Extending Pouw and Dixon's (2019b) notion of "stability through synergy", in the difficult condition, gestures and speech may have stabilized together by means of complexity matching, which entails coordination at multiple timescales, instead of entrainment, that is, coordination at a single timescale. Metaphorically speaking, the difficult condition might elicit a form of gesture-speech coordination which shares similarities with the coordination between a jazz-saxophonist and a jazz-pianist while improvising together, which is characterized by "...a multitude of simple and complex rhythms, all interwoven extemporaneously into one cohesive sound" (i.e. complexity matching; Herby Hancock Institute of jazz, <https://bit.ly/2FlypCm>; also see Walton et al., 2015, 2018). The easy condition might elicit a form of gesture-speech coordination similar to clapping one's hands in a regular, monotonous rhythm (i.e. entrainment). Furthermore, in the easy condition, entrainment may overrule complexity matching. This might suggest a trade-off between phase synchronization and complexity matching, which could be reflected in the negative relation between temporal alignment and complexity matching in the difficult condition that we found (also see Marmelat & Delignieres, 2012). In terms of our metaphor, if either the saxophonist or the pianist start playing a regular, monotonous rhythm, the other musician will be drawn to that regular and monotonous rhythm and will have a very hard time to maintain improvisation in all its complexity. We will discuss our findings' implications for the concept of complexity matching in the next paragraphs.

Task difficulty's influence on gesture-speech synchronization

Complexity matching

While a body of research has shown that complexity matching exists between different human systems and under different circumstances (e.g. Abney, 2016; Abney, Paxton et al., 2014; Almurad et al., 2017; Coey, 2016, 2018; Den Hartigh et al., 2018; Fusaroli et al., 2013; Marmelat & Delignieres, 2012; Ramirez-Aristizabal et al., 2018; Schloesser et al., 2019; Schneider et al., 2019), we are still grappling with what complexity matching *actually does* for people. In our study, we found more complexity matching between gestures and speech in the difficult condition than in the easy condition, and we interpreted this as a way for gestures and speech to stabilize together when entrainment is difficult to impossible. However, complexity matching did not predict participant's task performance in terms of time to finish the task, and complexity matching was also not related to semantic content-alignment (i.e. number of speech errors). Apart from gestures and speech potentially being more stable, as we proposed, it is unclear whether and how participants benefited from more complexity matching.

Different studies about complexity matching during dyadic tasks do show that participants who demonstrated complexity matching benefited from this, in terms of reaching a collaborative goal (Abney, Paxton, et al., 2014; Fusaroli et al., 2013; see also Schloesser et al., 2019). Important to note is that the performance measures in the studies by Abney, Paxton, et al. (2014) and Fusaroli et al. (2013) are more sophisticated and captured higher-order goals, than our simple performance measure of total time to perform the task did. In line with our findings, Schloesser et al. (2019) also found a weak and slightly negative relation between complexity matching - both within and between participants - and performance in terms of total time.

From a theoretical point of view, West et al. (2008) showed that complexity matching increases the information exchange between complex networks. However, as argued before by Abney (2016), we know little about what this *information* actually is, and how to operationalize it. We could speculate that complexity matching only increases performance on tasks that involve the (re)organization of components to a higher-order structure. This higher-order structure could be the *common ground* that interacting people needed to establish during a conversation involving many different utterances (Abney, Paxton et al., 2014), or the *joint decision* that people needed to converge to during a series of joint decision making (Fusaroli et al., 2013). If it is true that complexity matching only increases performance on tasks that involve the (re)organization of components to a higher-order structure, this could hint that the information as proposed by West entails *interactions between components that form a synergy*.

An interesting study by Rigoli et al. (2014; also see Schloesser et al., 2019) similarly suggests that information in complexity matching entails interactions between components that form a synergy. Rigoli et al. (2014) investigated participants who were asked to tap to a visual

metronome, by pressing a key. Rigoli et al. (2014) found complexity matching between the time series of participants' key press times and durations [*key press synergy*], and they found complexity matching between the time series of participants' pupil dilation and heart rate [*anatomic synergy*]. However, Rigoli et al. (2014) did not find complexity matching between key press time series and the anatomic time series. Rigoli et al. (2014) therefore concluded that the key press network and anatomic network did not exchange information during the simple and relaxed task of tapping to the visual metronome. Similarly, in the easy (simple and relaxed) condition of the current study we did not find complexity matching between gestures and speech, which suggests that these systems did not exchange information either. We did find complexity matching in the difficult condition however, which suggests that the gestures and speech exchanged information and (re)organized as a synergy under these difficult task constraints. Future studies could investigate whether difficult tasks, involving higher-order goals, indeed elicit more complexity matching between systems than simple tasks. With regard to difficult tasks involving higher order-goals for children, one example are Piagetian conversation tasks, which have been used to study the interplay between gestures and speech before (e.g. Alibali et al., 2000; Church & Goldin-Meadow, 1986; De Jonge-Hoekstra et al., 2020; Pine et al., 2004, 2007).

Gesture-speech mismatches

As previously described, Goldin-Meadow and colleagues (e.g. Church & Goldin-Meadow, 1986; Goldin-Meadow et al., 1993; Goldin-Meadow, 2003) found that children make gesture-speech mismatches (i.e. semantic *dissimilarities*) when they are on the verge of learning something new. Moreover, during these gesture-speech mismatches, children show a more advanced understanding in gestures than in speech. In the current study, we found more gesture-speech mismatches (i.e. less semantic similarity) in the difficult than in the easy condition, and these gesture-speech mismatches were always due to speech errors in semantic content. With our findings, we thus extend the phenomenon of gesture-speech mismatches from tasks in which people acquire understanding about cognitive problems, to difficult, cognitive tasks in general. Since a transition between “old” understanding and “new” understanding was impossible in our experiment, participants' gesture-speech mismatches were due to something different than competing cognitive understanding.

First, both in the current study and in previous studies, gestures had a clear spatial component that was directly linked to the physical properties of the task material (e.g. Bergmann & Kopp, 2010; De Jonge-Hoekstra et al., 2020; Hostetter & Alibali, 2008; Yeo & Alibali, 2018). This is not true for speech, however, and Smith and Gasser (2005) even propose that a too close resemblance between the physical structure of the environment and the structure of speech

Task difficulty's influence on gesture-speech synchronization

would limit speech's functionality. Maybe difficult, cognitive tasks amplify this difference between gestures and speech in how they are coupled to the physical properties of the spatial environment, which could result in gesture-speech mismatches. Furthermore, we could question the extent to which speech actually needed to be functional in the current study. Participants performed the task individually and their speech did not have to be understandable for someone else (also see Fowler, 2010). Future studies could investigate how task constraints related to spatial structure and social context influence the occurrence of gesture-speech mismatches.

Second, participants had to verbally discriminate left from right in our experiment, which is known to be notoriously difficult for children and adults alike (e.g. Fisher & Camenzuli, 1987; McKinley et al., 2015; Vingerhoets & Sarrechia, 2009). To our knowledge, no studies have investigated whether people find it difficult to discriminate between left and right using gestures as well. However, Abarbanell and (2020) recently found that instructing children to use gestures to discriminate between left and right benefits their performance on a rotation task more than instructing children to say the (Spanish) words (for) "left" and "right". The authors explain this effect by gestures being directly linked to the spatial properties of a task, similar to our reasoning in the previous paragraph. This direct link between gestures and spatial properties of a task is particularly evident for deictic gestures, like the pointing of participants in our study. Therefore discriminating between left and right using gestures was probably easier for the participants than using speech. Furthermore, while participants in the easy condition could just repeat the same sequence of words without much thought about their meaning, participants in the difficult condition needed to think about the words' meaning constantly. Participants in the difficult condition were therefore more prone to confuse the words "left" and "right", while they could correctly differentiate between left and right by means of pointing. This could explain why we found more gesture-speech mismatches in the difficult condition than in the easy condition. Future studies need to investigate whether this phenomenon is more evident in tasks which require left-right discrimination, as compared to spatial temporal tasks in general, as we argued in the previous paragraph.

Third, in line with Bergmann et al. (2011), we found a positive relation between temporal alignment and semantic similarity in both the difficult and easy condition, which suggests that more temporal alignment goes together with less gesture-speech mismatches. However, it is yet unclear whether temporal alignment is causally related to gesture-speech mismatches and what the direction of this potential relation would be. According to the Information Packaging Hypothesis (Kita, 2000; also see Kita et al., 2017), gestures help to organize and "package" spatial information to both enable verbalization about this spatial information, and to ensure that the spatial information "fits" within the structure of speaking. When verbalization is challenging,

speakers take more time to “package” information by means of gesturing. This would result in low temporal alignment between gestures and speech in the during gesture-speech mismatches, as well as low temporal alignment in the difficult condition. This is in line with the positive relation between temporal alignment and semantic similarity and less temporal alignment, and also with less temporal alignment in the difficult condition, that we found. Follow-up studies could research the relation between gestures, speech, and gesture-speech mismatches in more detail, using methods to quantify the temporal direction of gesture-speech coupling, such as Cross Recurrence Quantification Analysis (see also De Jonge-Hoekstra et al., 2016). Moreover, in previous studies, temporal information usually has been disregarded when coding gesture-speech mismatches (e.g. Alibali et al., 2000).

Limitations

Our study has a number of limitations. We will address the limitations that we deem most important.

First, participant's utterances during the experiment were very limited in scope and syntactic complexity (i.e. “left”, “middle”, “right”), which leaves open the question of how our findings will correspond to more typical, fluent, and syntactically complex speaking and gesturing. Previous studies have found complexity matching between participant's fluent speech (Abney et al., 2014; Fusaroli et al., 2013). Furthermore, Abney et al. (2018) created spike trains of participant's language and gesture events during fluent conversations and subsequently calculated the *burstiness* of both language and gesture events. Bursty processes are typical for complex dynamical systems (Barabási, 2005; Karsai et al., 2012), and in this sense, burstiness shares similarities with multifractality (albeit the scope of burstiness analysis is not multi-scaled). The methods used by Abney and colleagues (Abney, Paxton et al., 2014, 2018; Fusaroli et al., 2013) provide viable directions for investigating complexity matching between gestures and speech in more typical and fluent speaking and gesturing.

Second, instead of changing the physical lay-out and order of the task, we could have increased task difficulty in a way that is closer to cognitive problem solving. For instance, we could have asked participants to follow sets of rules about when to put which color ring on which color bar, and investigate how rules of varying difficulty influence gesture-speech coupling. However, such a manipulation would not have perturbed participants continuously as participants get used to rules, while the random order that we used in the current study did continuously perturb them.

Third, while we treated the trials from ring to bar and from bar to ring equally, the instruction for these trials differed. For the trials from ring to bar, participants were instructed to select the bar which has the same color as the ring. For the trials from bar to ring, participants were instructed to select the enlarged ring. This difference in instruction could potentially lead to a

Task difficulty's influence on gesture-speech synchronization

different pattern of multifractal scaling for the trials from ring to bar and for the trials from bar to ring. In an interesting study, Kello et al. (2007) investigated a task whereby participants needed to press a key on a keyboard when they saw a stimulus on screen, thereby responding as fast as they could. Participants were allocated to either an easy, predictable condition, whereby the time between the stimuli was constant, or to a difficult, unpredictable condition, whereby the time between stimuli was random within a certain range. Kello et al. (2007) analyzed two time series: 1) A time series of the time between the appearance of the stimuli and pressing the key (reaction time), and 2) a time series of the time between pressing the key and releasing the key again (key-contact duration). The authors argue that participants only received an instruction about reaction time (responding as fast as possible), while they received no instruction about key-contact duration. Kello et al. (2007) found the reaction times and key-contact durations in both conditions to be not or only weakly correlated. Furthermore, they found fractal scaling in both the reaction time series and the key-contact duration time series and in both conditions. The fractal scaling of the reaction time series of the difficult, predictable condition was lower than the fractal scaling of the other three time series. Although the study by Kello et al. (2007) shares some similarities with our study, there are notable differences as well. While pressing down a key as fast as possible and releasing a key correspond to a simple instruction vs. no instruction, respectively (Kello et al., 2007), selecting a bar with the same color and selecting an enlarged ring correspond to a more complicated instruction vs. a simple instruction, respectively (current study). Furthermore, while pressing down and releasing a key are two different motions, involving the contraction of different muscles (Kello et al., 2007), trials from bar to ring and from ring to bar *both* involved pointing to a target and saying the location of that target (current study). A follow-up study could investigate whether the ring to bar trials differ from the bar to ring trials with regard to duration and multifractal scaling.

Fourth, our sample size is relatively small, which is largely due to failed audio-recordings. However, we do have many datapoints per participant. Fifth, the number of measurements per participant (1024) was on the small side for performing MFDFA (Ihlen & Vereijken, 2010), yet sufficient. Albeit challenging, we need to come up with ways to increase the number of measurements per participants while still keeping the task feasible for participants to do. Furthermore, Almurad and Delignières (2016) propose an alternative way of performing DFA (the monofractal variant of MFDFA) which allows for timeseries which are even shorter than 1024 datapoints.

Conclusions

We aimed to investigate how task difficulty affects the synchronization between gestures and speech, thereby empirically addressing De Jonge-Hoekstra et al.'s (2016) proposal. By doing so,

we brought together different perspectives on and ways of investigating gesture-speech synchronization. We found that task difficulty indeed influences gesture-speech synchronization in terms of temporal alignment, semantic similarity, and complexity matching. With our findings of less semantic similarity in the difficult condition, we extended the phenomenon of gesture-speech mismatches to difficult, cognitive tasks. Furthermore, we found more temporal alignment in the easy condition, which we related to phase synchronization between gestures and speech. We found more complexity matching between gestures and speech in the difficult condition, which we related to gestures and speech forming a more stable synergy under the influence of more difficult task constraints. Our findings add another piece to the puzzle of why complexity matching between occurs in complex dynamical systems.

In sum, our study demonstrates how this perspective can be used to study the relation between gestures and speech, and gesture-speech mismatches – subjects that primarily have been studied from within cognitive psychology. While the body of research that tries to bridge between complex dynamical systems and coordination research, and cognitive psychology is steadily growing, we acknowledge that many gaps between the two perspectives still remain. We look forward to future work that continues to build connections between the two fields, and we hope that these future studies can build on our study.

Acknowledgements

We would like to thank Dr. Mark Span for constructing the tablet task in OpenSesame. Furthermore, we would like to thank all the students who helped with collecting data, and coding words. In addition, we would like to thank Dr. Wim Pouw for his advice on how to automatically calculate temporal alignment between gestures and speech. Lastly, we would like to thank Dr. Alexandra Paxton for the discussion we had about applying complexity matching to our gesture and speech data.



4

Movers and shakers of cognition: Hand movements, speech, task properties, and variability

This chapter is based on:

De Jonge-Hoekstra, L., Van Der Steen, S., & Cox, R. F. A. (2020). Movers and shakers of cognition: Hand movements, speech, task properties, and variability. *Acta Psychologica*, 211, 103187. doi:10.1016/j.actpsy.2020.103187

Movers and shakers of cognition: Hand movements, speech, task properties, and variability

Children explore, learn, and communicate with their hands. This is especially evident during so-called *hands-on learning* activities. Hands-on learning implies that children are encouraged to actively engage with the task material, initiate different actions and thereby circumstances, and find out what happens when they do so (Kuhn et al., 2009; Zhang, 2019). Asking children to verbally explain why and how these events happen further increases their understanding of the task (Van der Steen, Steenbeek, Van Dijk, & Van Geert, 2014; Van der Steen, Steenbeek, Den Hartigh, & Van Geert, 2019). During these explanations children show a variety of hand movements, such as pointing, simulating, and demonstrating what has happened (Novack & Goldin-Meadow, 2015). Similar to manipulating task material, these hand movements are characterized by recruiting the environment. For instance, pointing is usually directed at a specific object or location (Delgado et al., 2011), while simulating and demonstrating involves taking on spatial and temporal properties (i.e. shape, movement) of the task (Boncoddio et al., 2010; Hostetter & Alibali, 2008; Yeo & Alibali, 2018). Speaking, on the other hand, is not characterized by a direct correspondence to these spatiotemporal task properties (see also Fowler, 2010; Smith & Gasser, 2005). In the next sections, we will describe in more detail: 1) how hand movements and speaking are related to spatiotemporal properties and cognitive development, 2) how spatiotemporal properties affect behavior's diversity, complexity, and development, and 3) how we explore and combine the above topics in the current study. With this study, we aim to understand whether and how hand movements' leading role in cognitive development is related to its ability to correspond to spatiotemporal task properties, while speech is unable to do so.

Spatiotemporal properties and cognitive development

How hand movements and speaking differ in their correspondence to spatiotemporal properties is particularly interesting in light of hand movement's leading role in cognitive development. When a child explores a new object, they use their hands to touch, feel, and manipulate the object, and to bring it to their eyes, ears, nose and mouth (Adolph & Franchak, 2017; Adolph & Kretch, 2015). This exploratory learning is also typical for hands-on learning activities (Fischer & Bidell, 2006; Roth, 2002). Another strand of research is devoted to children's (hand) gestures when they learn (Adolph et al., 2015). Goldin-Meadow and colleagues found that children are able to display cognitive understanding in gestures, before they are able to put this into words (Alibali & Goldin-Meadow, 1993a; Church & Goldin-Meadow, 1986; Goldin-Meadow, Wein, & Chang, 1992; Goldin-Meadow, Alibali, & Church, 1993; Perry, Church, & Goldin-Meadow, 1992). In these studies, this understanding in gesture usually takes the form

Hand movements, speech, task properties, and variability

of a shape of an object (Church & Goldin-Meadow, 1986; Hilliard & Cook, 2017) or simulation of an action (Alibali & Goldin-Meadow, 1993a; Hostetter & Alibali, 2008; Yeo & Alibali, 2018). In other words, children naturally move their hands in correspondence to relevant spatiotemporal properties of the task when they gesture, which seems to precede verbal explanations involving these properties. In addition, more recent studies found that also *guiding* children's hands to move in correspondence to these relevant spatiotemporal properties fosters later verbal explaining of new concepts (Broaders, Cook, Mitchell, & Goldin-Meadow, 2007; Brooks & Goldin-Meadow, 2016; Novack, Congdon, Hemani-Lopez, & Goldin-Meadow, 2014). These studies suggest that hand movements' leading role in cognitive development may originate from their correspondence to spatiotemporal task properties.

The saliency of relevant spatiotemporal task properties not only influences children's hand movements, but also their verbal explanations (Kloos et al., 2010; Meindertsma, 2014). Still, it is unclear how those children's hand movements are affected by this saliency of spatiotemporal task properties, and how this is related to the change in their verbal explanations. Furthermore, an explanation for how children, who engage with different task properties, thereby cognitively develop is lacking. However, studies onto children's *motor development* have long recognized the importance of different task properties, and how the consequential *variability* is essential for developing new skills.

Behavioral variability: Diversity and Complexity

The influence of (saliency of) different task (or more broadly: environmental) properties is widely known in the area of motor development (Adolph et al., 2015, 2018; E. J. Gibson & Pick, 2000). Children constantly have to adapt their movements to the different environments that they are in (not to mention the constant changes in their own, growing body). This implies that their behavior needs to be variable and diverse, in order to be functional and adaptive to different task demands (Adolph et al., 2015). A similar diversity of behavior has also been found in cognitive development, where it is indicative of learning something new (Siegler, 2007; Van der Steen et al., in press).

Next to diversity of behavior, Adolph et al. (2015) describe a second feature of variability that is important in (motor) development: Its structure (see also Abney, Warlaumont, Haussman, Ross, & Wallot, 2014; Cox & Van Dijk, 2013; Kello, Beltz, Holden, & Van Orden, 2007; Van der Steen, Steenbeek, & Van Geert, 2012; Van Dijk & Van Geert, 2014; Van Orden, Holden, & Turvey, 2003; Wijnants, Hasselman, Cox, Bosman, & Van Orden, 2012). Behavior never happens in a vacuum, but is instead nested in sequences of previous and future behavioral events (time series). However, the degree to which previous behavior determines next behavior can differ.

When behavior is relatively independent from previous behavior it leans more toward randomness. An example about hands-on learning would be sequences of hand movements or speech that are highly unstructured with regard to duration, type, and order (i.e. doing things at random). On the other side of the spectrum are behaviors that are almost completely determined by previous behavior. For instance, a child could repeat a sequence of hand movements or speech over and over again (i.e. remaining stable, not getting any further). In between these two extremes lies complex behavior, which depends on previous behavior, but also flexibly deviates from what has happened before. This flexibility is related to handling changes in task demands. In complex systems' terms, handling changes in task demands can be thought of as a *system* of interrelated components changing from one stable *state* to another, potentially novel, stable state (e.g. Smith & Thelen, 2003; Stephen, Boncoddio, et al., 2009; Stephen, Dixon, et al., 2009; Thelen & Smith, 1994; Thelen & Smith, 2007; Van Geert, 2008; Van Geert, 2011). Changing from one state to another entails a *reorganization* of a system's components and their relations, which is only possible when the coupling between components *loosens* and the system becomes *more flexible*. Metaphorically, one could think if this reorganization as building a new LEGO-structure from an old structure – this is only possible when you break the old structure (loosen the coupling, increase flexibility) and use the bricks to create a new structure. An example of this in our study would be the emergence of novel hand movements and speech, which build upon previous behavior (i.e. flexibility, complexity).

Current study

In the current study, we combined (a) hand movements' leading role in cognitive development by corresponding to spatiotemporal task properties with (b) Diversity of behavior as functional adjustment to new task demands, and (c) Complexity of behavior as functional flexibility when changes in task properties demand it.

We systematically manipulated the salience of spatiotemporal properties relevant to a hands-on task. We specifically investigated children's (4-7 years) hand movements and speech while they were asked to predict and explain about balance scale problems. In accordance with Siegler (1976), two dimensions are important when solving balance scale problems: The *mass* of the weights and the *distance* from the fulcrum of the balance scale. We therefore manipulated the salience of the distance-dimension and the weight-dimension in two experiments, which we will further explain below. Children as young as 4 years old have been found to consider the weight-dimension when they solve balance scale problems (Schrauf et al., 2011). However, taking account of the distance-dimension in predicting about balance scale problems only rarely happens at age 5 to 6 (Siegler, 1976). Children in our sample thus reflected

Hand movements, speech, task properties, and variability

the age group that uses the weight-dimension in balance scale problems, while they still have to learn about the distance-dimension. Furthermore, Pine, Lufkin, Kirk and Messer (2007) found that specifically gestures' leading role in cognitive development is also evident when children reason about balance. Lastly, Messer, Pine and Butler (2008) found that being able to physically manipulate either the distance- or the weight-dimension affects the probability of explaining about the distance-dimension.

In the current study, four-to-seven-year-olds were asked to predict, describe and explain what happens when different weights are hanged at different positions of a balance scale. We manipulated the distance-dimension and the weight-dimension of a balance scale task in two experiments (see Figure 1), each consisting of eight trials. In Experiment 1, children were first presented a long balance scale and then a short balance scale, or vice versa. By manipulating the length of the balance scale, we manipulated a task property that is related to the perceptual *salience of the distance-dimension* (Van De Langenberg et al., 2006). To clarify this, with a longer balance scale, the distance of the balance scale stands out more, both visually and haptic. To hang weights at the more distant hooks of a longer balance scale participants have to stretch

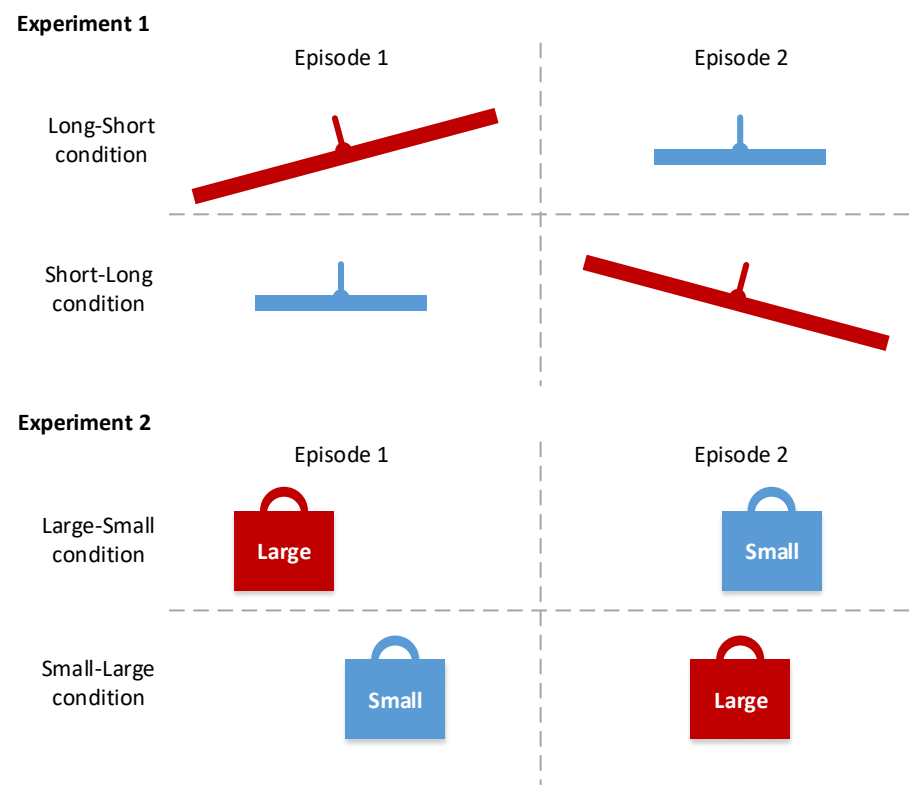


Figure 1. Design of Experiment 1 and Experiment 2.

the arms further and apply more force. In Experiment 2, children first received weights with a large mass or with a large difference in mass, and then weights with a small mass or with a small difference in mass, or vice versa. Hereafter, we will simply use *large mass* to refer to the episodes in which participants worked with a large mass or large difference in mass, and use *small mass* to refer to the episodes in which participants worked with a small mass or small difference in mass. By manipulating the mass of the weights, we manipulated a task property that is related to the perceptual *salience of the weight-dimension*. With a larger mass participants have to exert more force to resist gravity's pull on the weights when they hold them or attach them to the balance scale.

As pointed out before, the children in our sample were of an age (4-7 years) at which they generally use the weight-dimension in balance scale problems, while they still have to learn about the distance dimension (Schrauf et al., 2011; Siegler, 1976). However, in the balance scale problems that we presented, we not only varied the weight-dimension, but also the distance-dimension. This implies that children needed to adapt to a new task demand (i.e. learn) - taking account of the distance-dimension - to perform the task correctly. According to Adolph et al., (2015), Harbourne and Stergiou (2009), Smith and Thelen (2003), Van Dijk and Van Geert (2014), Van Orden et al. (2003), and Wijnants et al. (2012), adapting to a new task demand goes together with an increase in behavior's diversity and complexity. Furthermore, the change in salience of the distance- and the weight-dimension is a change in the spatiotemporal properties of the task. Following Adolph and Franchak (2017), Alibali and Goldin-Meadow (1993a), Church and Goldin-Meadow (1986), Hilliard and Cook (2017), Hostetter and Alibali (2008), and Yeo and Alibali (2018), children's hand movements correspond to this change in spatiotemporal task properties, while this is not the case for speech. Possibly due to this correspondence with spatiotemporal task properties, hand movements are leading in cognitive development, ahead of speech (Alibali & Goldin-Meadow, 1993a; Church & Goldin-Meadow, 1986; Goldin-Meadow et al., 1992; Goldin-Meadow et al., 1993; Perry et al., 1992). Tying all this together, we explored¹ the following research question in both experiments: How does a change in task property affect diversity and complexity in children's hand movements and

¹ We submitted a manuscript about the same video data to another journal (preprint: <https://osf.io/t2dkr/>) in 2018, where it was rejected. The objections of the reviewers were valid and their feedback was constructive, and we used their suggestions to improve our codings of the hand movements (which were called "gestures" in the previous submission) and we rewrote most of the manuscript. Furthermore, we improved our variability analyses. First, concerning our variability measure of complexity, Leonardi (2018) published a new and superior variability measure for complexity of categorical time series, based on Recurrence Quantification Analysis, which we used for our analyses. Second, we also improved our variability measure for diversity, by taking the duration of behaviors into account. These changes have led to different and more robust results. Because we changed the analyses after we knew the outcomes of the previous analyses, the hypotheses in this study are explorative.

Hand movements, speech, task properties, and variability

speech when they are asked to predict, describe and explain about an unfamiliar dimension of balance scale problems (the distance-dimension)?

Note that in Experiment 1, the change in salience of task property (i.e. length of the balance scale) is congruent with the new task demand to consider the distance-dimension. Our exploratory hypothesis for Experiment 1 is therefore that we find an increase in diversity and complexity for hand movements, but not for speech (hypothesis A). In Experiment 2, however, the change in salience of the task property (i.e. mass or different in mass) is not congruent with this new task demand. Instead, changing the salience of the weight-dimension converges with the “old” task demand to consider the weight-dimension, at which children generally are skilled already. For Experiment 2, our exploratory hypothesis therefore is that we find no difference in diversity and complexity, nor for hand movements, nor speech (hypothesis B).

This is one of the first studies that incorporates multiple measures of behavioral variability, thereby contributing to understanding how these types of variability are related. Moreover, to our knowledge, this is the first study that investigates how spatiotemporal properties are related to diversity and complexity of hand movements and speech in a hands-on learning task. The outcomes of this study shed light on how hand movements and speech are related to changes in spatiotemporal task properties and changes in task demands. This study thereby adds to the growing field devoted to how children learn by interacting with their environment.

Experiment 1

Method

Participants

A total of 20 children from Kindergarten ($n = 15$) and first grade ($n = 5$), age 4 to 7 years ($M = 5.18$; $SD = 0.92$) participated in this experiment. We recruited all participants at their schools located in the north of the Netherlands, and asked parents of the participants to give written consent. We informed the parents that their children would work on science and technology tasks with different task properties, but not about the specific nature of the tasks. The ethics committee of the host institution approved the study.

Materials

We used two balance scales: A long and a short balance scale (scale 2:1). The long balance scale measured 84 cm, had six hooks on each side of the center of the balance scale, which were spaced 7 cm apart. The short balance scale measured 42 cm, and had six hooks on each side of the center, which were spaced 3.5 cm apart (see Figure 2 for an illustration). For both balance

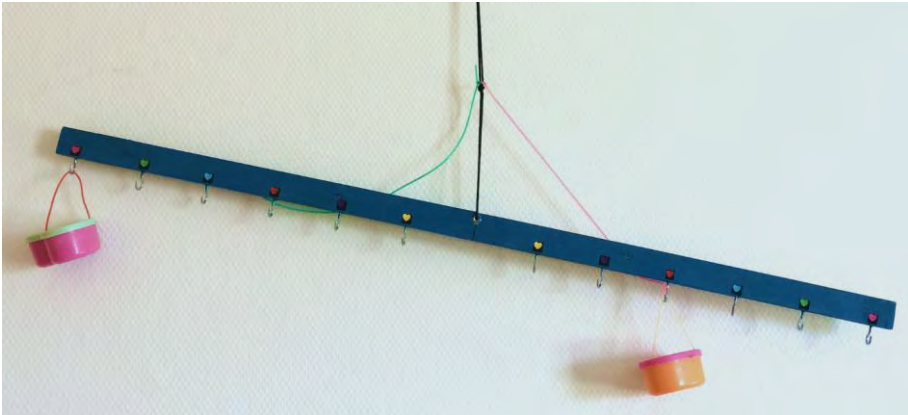


Figure 2. The long balance scale with two weights attached to it.

scales, we tied a small rope to the center, in order for the balance scales to tilt to the left or the right when weights were attached. We used eight weights for administering the balance task, with a mass of either 50 g, 75 g, 100 g or 150 g (two weights of each mass). Besides color, there were no other differential features of the weights.

To enable detailed analysis of the behavior of the participants during the task, we recorded the task administration on video. We placed two video-recorders on tripods and positioned them in two different angles, in order to fully record the hand movements of the participants. After we collected the video-data, we manually coded the hand movements and speech of the participants using the program MediaCoder (Bos & Steenbeek, 2006). With MediaCoder, video recordings can be played and codes can be added to specific points in time, yielding an overview of the course of the behavior under investigation. We used R [3.6.1] and RStudio [1.1.456] to analyze the data.

Procedure

The children were randomly assigned to one of the following conditions: in one condition, we presented them with a long balance scale in the first half of the task and with a short balance scale in the second half. We reversed the order of presenting this task property in the other condition. The participants engaged in a hands-on balance task, guided by an experimenter. The experimenter followed a structured protocol when administering the task, which allowed for asking follow-up questions to encourage reasoning (i.e., “Why do you think so?”, “How would that work?”) and for clarification. The task was set-up with the balance scale attached to a table, so that it could tilt, and the weights arranged at the floor. The experimenter first asked if the participant had ever seen something similar. After answering this question, the participant was asked to explore the balance scale and weights. Next, the experimenter explained the

Hand movements, speech, task properties, and variability

procedure of the task and emphasized that the participant was free to say what he/she thought, and that there were no wrong answers. After this introductory-phase, the trials commenced.

The participants were asked questions about balance problems during eight trials. In each trial, the experimenter first asked the participant to feel two specific weights. Then the participant was asked to predict what would happen when the weights were attached at hooks on either side of the balance scale, at a specific distance from the center. After predicting and performing this task, the participant was asked to describe and explain what happened. Following the completion of eight of these trials, the participants were thanked and received a small reward for their participation.

Although the general procedure of the trials was the same for all participants, there were differences in the configuration (i.e. position and mass of the weights) and properties of the task (i.e. length balance scale), depending on the condition the participants were assigned to (see Table 1). In the Long-Short condition, the participants worked with the long balance scale during the first four trials (Long-balance episode), and then with the short balance scale for the last four trials (Short-balance episode). Conversely, in the Short-Long condition, participants first worked with the short balance scale (Short-balance episode), and then with the long one (Long-balance episode).

Coding procedure

We coded both participants' hand movements and speech, using the computer program MediaCoder (Bos & Steenbeek, 2006). First, we coded hand movements, while we muted the

Table 1
Weights and their position on the balance scale for the eight trials in the two experiments

Trial	Experiment 1				Experiment 2			
	Long-Short-condition		Short-Long-condition		Large-Small-condition		Small-Large-condition	
	Weights	Position	Weights	Position	Weights	Position	Weights	Position
1	50; 50	3; 3	50; 50	3; 3	50; 50	3; 3	50; 50	3; 3
2	50; 75	5; 5	50; 75	5; 5	50; 150	4; 4	50; 75	5; 5
3	50; 50	5; 3	50; 50	5; 3	150; 150	5; 4	50; 50	3; 5
4	50; 75	6; 4	50; 75	6; 4	25; 75	3; 1	50; 75	3; 1
5	100; 150	2; 2	100; 150	2; 2	50; 75	5; 5	150; 150	5; 4
6	75; 75	1; 3	75; 75	1; 3	50; 50	3; 5	25; 75	3; 1
7	100; 150	3; 2	100; 150	3; 2	50; 75	3; 2	50; 150	4; 4
8	50; 75	4; 2	50; 75	4; 2	100; 150	4; 2	25; 100	6; 1

Note. The mass of the weights is in grams. Position ranges from 1 to 6, which corresponds to the two hooks closest to the center (position 1) to the two hooks closest to both ends (position 6) of the balance scale.

sound of the video-recordings in order to forestall interpretation of the hand movements based on what participants said. Movements of the left- and right hand were coded in two subsequent rounds, to be able to focus on the movements of each individual hand, which could be different from the other hand. While coding, the behavioral categories *no hand movements*, *attaching* (of weights on the balance scale), *gesturing*, *hand movements with task materials*, *hand movements without task materials* were differentiated. *Attaching* corresponded to the moment of attaching weights on the balance scale, *gesturing* corresponded to all deictic and representational gestures, *hand movements with task materials* corresponded to hand movements in which participants' hands made contact with task materials, and *hand movements without task materials* corresponded to all other hand movements that did not fall under the previous categories. When a hand movement started, we coded the corresponding behavioral category, and when a hand movement stopped, we coded the category *no hand movements*.

After we coded the hand movements of the left- and right hand, the sound of the video-recordings was put on and speech was coded. For speech, the behavioral categories of *no speech*, *predicting*, *explaining*, and *other speech* were differentiated. *Predicting* corresponded to all task related utterances that happened before the balance scale was released, while *explaining* applied to all task related utterances that happened after the balance scale was released in each trial. In the same manner as for coding hand movements, when a speech utterance started, we coded the corresponding behavioral category, and when a speech utterance stopped, we coded the category *no speech*.

The video-recordings were coded by students, using a standardized codebook. Before coding the video-recordings, the students received a training in which they had to code several video-fragments to familiarize themselves with the codebook. When the students thought they were ready, they coded movements of both hands and speech of an 11-minute video recording which was previously coded by the first author. The coding of the students was compared with the coding of the first author, and if a student reached a proportion of inter-rater agreement of .75 or more, they were allowed to code the video-recordings. Each video recording was then coded by two students, and their coding was compared, leading to proportions of inter-rater agreement. The proportion of inter-rater agreement for the coded hand movements was on average .96 ($SD = .02$), and .91 ($SD = .01$) for speech. Based on the high levels of inter-rater agreement, we used the coding with the highest detail for analysis.

Analysis

To analyze the data, we transformed the codes of the video recordings to a time series of hand movements (Figure 3, panel a) and a time series of speech (Figure 3, panel b), with a sample

Hand movements, speech, task properties, and variability

rate of 2 Hz. For hand movements, we combined the codes for the left- and right hand into one time series, which preserved the possible different actions of both hands. For example, if the left hand made a gesture while the right hand did nothing, this appeared as “g_0” in the time series. Subsequently, we split the time series of hand movements and speech and investigated two parts: One part which contained the first four trials and a second part which contained the last four trials (i.e., after the switch in task property). The first exploratory hypothesis was that changes in the distance-dimension of the balance scale would yield an increase in Diversity and Complexity for hand movements, while not for speech. An overview of our analysis procedure can be found in Figure 3.

Diversity. We operationalized Diversity by calculating Shannon entropy (Shannon, 1948) on the frequency distribution of the duration and occurrence of behavioral categories in the two parts of each of the time series. Shannon entropy has been used in a broad range of fields, such as ecology (Jost, 2006), evolutionary genetics (Sherwin et al., 2017), and linguistics (Jarvis, 2013), and captures the unpredictability of a system’s state (i.e., behavioral category). We calculated Shannon entropy by means of the following formula: $H(X) = -\sum_{i=1}^n p(x_i) \log_2 p(x_i)$, where $p(x_i)$ is the frequency of a behavioral category of a certain duration (see Figure 3, panel c). Our calculations yielded four Shannon entropy-values for each participant: Two for each part (i.e., before and after the task property-switch) of the gestures-time series and two for each part of the speech-time series. The Diversity values indicate the amount of variability of the participants’ gestures and speech without taking into account the temporal structure of the behavioral sequence. To calculate Diversity, we wrote a custom R script (link to script: <https://osf.io/2sy5u/>).

Complexity. We derived a measure of Complexity by performing Recurrence Quantification Analysis (Marwan et al., 2007; C. L. Webber Jr. & Zbilut, 2005) on the time series of gestures and of speech. RQA is a nonlinear method to analyze time series, which is based on the notion of recurrence. Recurrence - the re-occurrence of states over time - is a central property of complex dynamical systems, such as the weather, mechanical engines, and also humans (Abney, Warlaumont et al., 2014; Cox et al., 2016; Riley & Turvey, 2002; Wijnants et al., 2012). These recurrences are represented in a Recurrence Plot (RP), which, for categorical time series, is created by plotting that time series against itself in a plane and marking all instances that pertain to the same state in x and y with a dot (see Figure 3, panel d).

The distribution of dots in a RP reveals the temporal dynamics of a system by means of the *line structures* that they form. Subsequent recurrences create diagonal lines, whereby their line length is related to stability of the system (Webber Jr. & Zbilut, 2005). RQA on a perfect periodic function like a sine wave yields long diagonal lines, whereas less regular and unpredictable

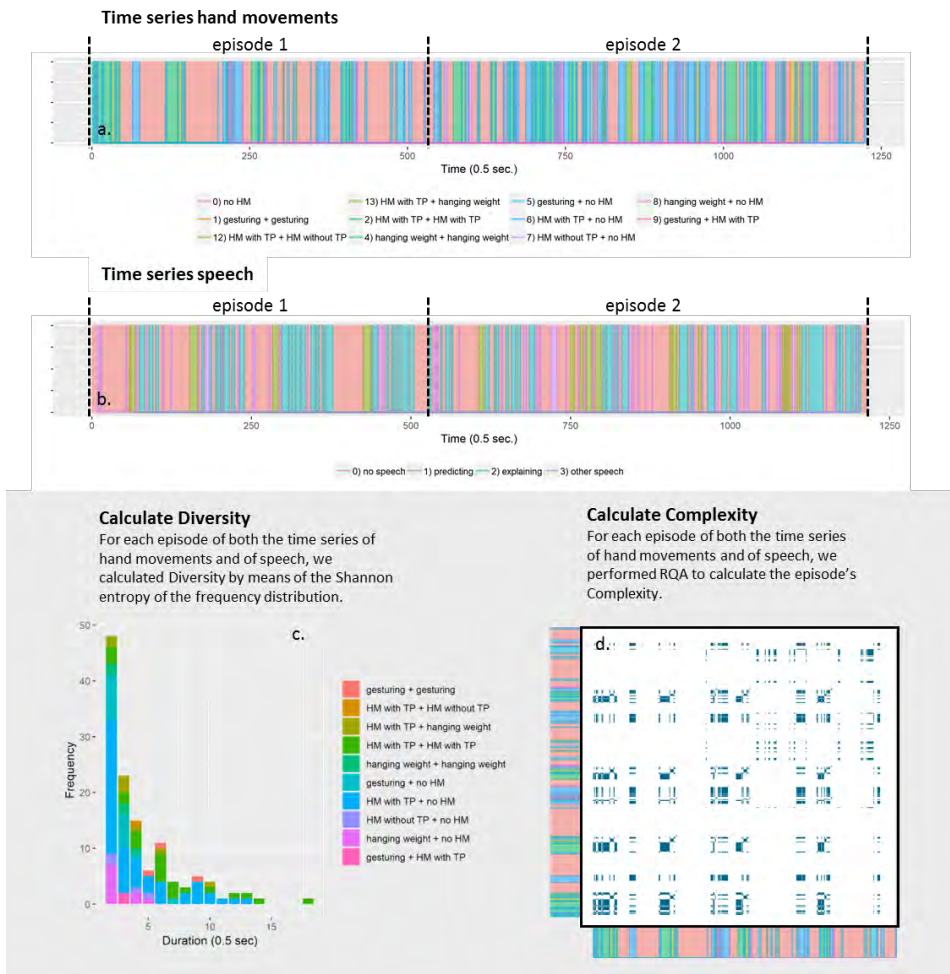


Figure 3. Overview of data and analyses. *Panel a* shows the categories of hand movements (HM = Hand movements, TP = Task Property) over time for one participant in the Long-Short condition, and *panel b* shows the categories of speech over time for the same participant. The dotted line in the middle indicates the switch in task properties, and thereby the start of episode 2. *Panel c* shows the frequency distribution of the categories and durations of the time series of episode 1 of hand movements in panel a. Diversity is calculated by the Shannon entropy of this frequency distribution, and captures how diverse the frequency distribution is. *Panel d* shows the recurrence plot of the time series of episode 1 of hand movements in panel a, whereby the time series is plotted against itself. When a behavior is the same in both x and y, this appears as a dot in the recurrence plot. The dots form block structures with different sizes, and Complexity is calculated by the Shannon entropy of the frequency distribution of the block structures' sizes.

Hand movements, speech, task properties, and variability

systems (such as humans) yield diagonal lines with a wide variety of different line lengths. The Shannon entropy of the frequency distribution of the *diagonal* line lengths gives a measure of complexity of the system (Pellecchia & Shockley, 2005; Webber Jr. & Zbilut, 2005). However, in categorical RQA, *vertical* and *horizontal lines* (Cox et al., 2016; Xu et al., 2020), and the rectangular *block structures* (Leonardi, 2018), are much more informative about a system's dynamics, instead of diagonal lines (also see Figure 3, panel d). Therefore, the Shannon entropy of the frequency distribution of the size of the block structures in the RP is better measure of a system's Complexity, specifically suited for categorical data (Leonardi, 2018). In terms of measuring changes between stable states, previous studies have linked *stable states* and the corresponding *strong and tight coupling* to a *low* Shannon entropy of line structures and block structures in the Recurrence Plot (Leonardi, 2018; Lichtwarck-Aschoff et al., 2012; Pellecchia & Shockley, 2005; Stephen, Boncoddio, et al., 2009; Stephen, Dixon, et al., 2009; C. L. Webber Jr. & Zbilut, 2005). Vice versa, *reorganization* and the corresponding *loose and flexible coupling* has been related to a *high* Shannon entropy of line structures and block structures in the Recurrence Plot. We used the crqa-package by Coco and Dale (2014) to perform RQA and create the RP, and we edited their script to calculate the Shannon entropy of the frequency distribution of the size of the block structures in the RP (link to script: <https://osf.io/2sy5u/>).

Please note that, although Diversity and Complexity are both based on Shannon entropy measures, they apply it to different distributions, thereby quantifying different types of variability. Diversity is based on the frequencies of the different behavioral categories of hand movements and speech and their duration, whereas Complexity is based on the block structures in the RP, which reflects the dynamic, temporal organization of hand movements and speech.

To investigate if a change in task property affects Diversity and Complexity in children's hand movements and speech, we calculated the Diversity and Complexity of each episode, for hand movements and for speech (see Figure 3, panel d, and Figure 4). We subsequently performed a within-subjects comparison between either Diversity or Complexity of gestures or speech in the two episodes. Because the a-priori chance of number of categories of children's hand movements and speech differs between children, and this influences the a-priori value of Diversity and Complexity, we calculated the standardized difference between the episodes as $(M_{Long} - M_{Short}) / (M_{Long} + M_{Short})$, to measure children's relative change in Diversity and Complexity. We calculated *p*-values using Monte Carlo (MC) Permutation tests (Ninness et al., 2002; Todman & Dugard, 2001), because these require no specific underlying distribution of the data. By drawing 10,000 random samples from the original data, the probability that differences are caused by chance was measured. We used custom-made R scripts to calculate *p*-values using MC permutation tests (link to scripts: <https://osf.io/2sy5u/>).

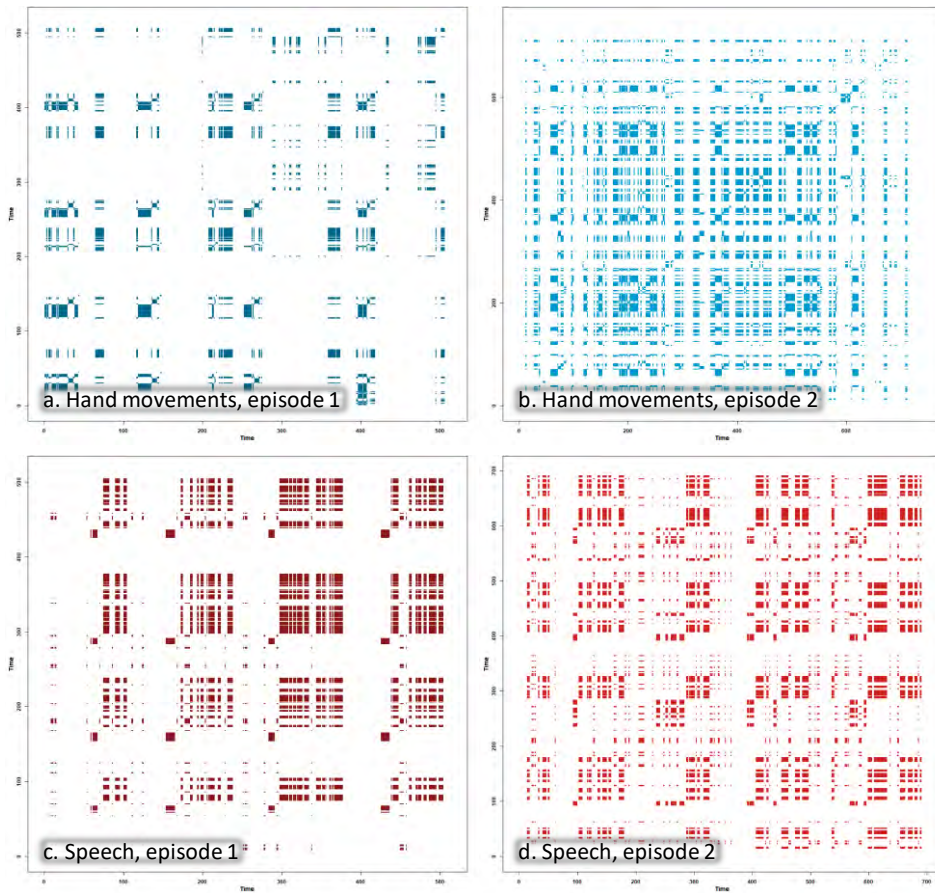


Figure 4. Recurrence plots for the first episode (left) and the second episode (right) for hand movements (a and b) and speech (c and d) of one participant in the Long-Short condition.

Results

In the Long-Short condition, we found no significant differences in Diversity between the first and the second episode, for neither hand movements ($M_{st. diff.} = 0.01$; $SD_{st. diff.} = 0.06$; $p = .37$; 95% $CI_{MC} = -.04, .04$) nor speech ($M_{st. diff.} = 0.02$; $SD_{st. diff.} = 0.05$; $p = .24$; 95% $CI_{MC} = -.04, .04$). We also found no significant differences in Complexity between the first and second episode, for neither hand movements ($M_{st. diff.} = 0.00$; $SD_{st. diff.} = 0.04$; $p = .45$; 95% $CI_{MC} = -.03, .03$) nor speech ($M_{st. diff.} = -0.00$; $SD_{st. diff.} = 0.03$; $p = .36$; 95% $CI_{MC} = -.02, .02$). Figure 4 shows the recurrence plots for both episodes of hand movements and speech for one participant in the Long-Short condition. The recurrence plots of the other participants can be found at <https://osf.io/2sy5u/>.

In the Short-Long condition, we did not find significant differences in Diversity between the first and second episode, not for hand movements ($M_{st. diff.} = -0.01$; $SD_{st. diff.} = 0.05$; $p = .35$; 95% $CI_{MC} =$

Hand movements, speech, task properties, and variability

= -.04, .03), nor for speech ($M_{st. diff.} = -0.00$; $SD_{st. diff.} = 0.05$; $p = .48$; 95% $CI_{MC} = -.04, .04$). Lastly, we did find significant differences in Complexity between the both episodes, for both hand movements ($M_{st. diff.} = -0.04$; $SD_{st. diff.} = 0.02$; $p = .01$; 95% $CI_{MC} = -.02, 0.2$) and speech ($M_{st. diff.} = -0.03$; $SD_{st. diff.} = 0.06$; $p = .04$; 95% $CI_{MC} = -.03, .03$).

These results are not in line with our first exploratory hypothesis (1A) that we would find an increase in Diversity and Complexity for hand movements, but not for speech. Instead, we found no significant differences in neither Diversity nor Complexity for both modalities in the Long-Short condition. In the Short-Long condition however, we found a decrease of Complexity for both modalities, but no significant differences in Diversity.

Since the results for hand movements and speech were similar, we additionally analyzed whether the standardized differences between episodes of hand movements and speech were related (see Figure 5). In the Long-Short condition we found a moderate and insignificant negative correlation for Diversity ($r = -.46$; $p = .06$; 95% $CI_{MC} = -.52, .48$), and a negligible and insignificant correlation for Complexity ($r = -.04$; $p = .47$; 95% $CI_{MC} = -.54, .48$). In the Short-Long condition we found a negligible and insignificant correlation for both Diversity ($r = -.01$; $p = .50$; 95% $CI_{MC} = -.45, .44$) and Complexity ($r = .07$; $p = .40$; 95% $CI_{MC} = -.42, .46$). These results show that the standardized differences between episodes of hand movements and speech are unrelated.

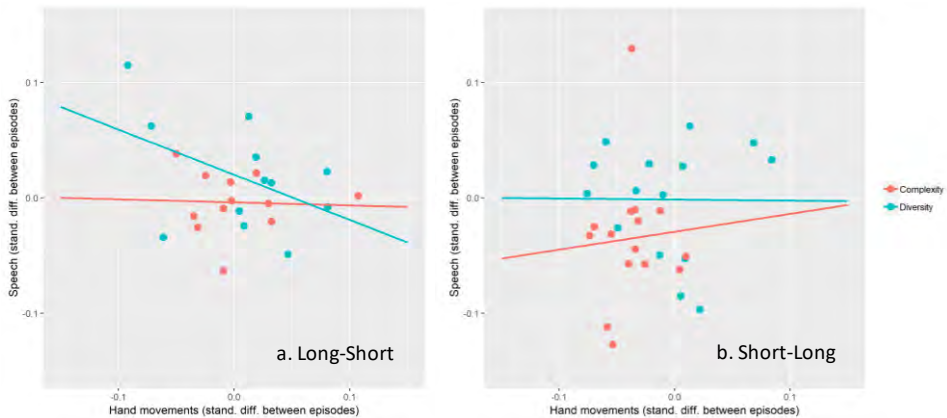


Figure 5. Relation between hand movements and speech with regard to standardized difference between episodes of Diversity (blue) and Complexity (red) in the Long-Short condition (panel a, $r_{Diversity} = -.46$, $p_{Diversity} = .06$, $r_{Complexity} = -.04$, $p_{Complexity} = .47$) and the Short-Long condition (panel b, $r_{Diversity} = -.01$, $p_{Diversity} = .50$, $r_{Complexity} = .07$, $p_{Complexity} = .40$).

Discussion

Our first hypothesis, for Experiment 1, was that we would find an increase in Diversity and Complexity for hand movements, but not for speech (hypothesis A). However, our results are not in line with this. We found different results for the two conditions, which differed in order of presenting the task properties: In the Long-Short condition we found no differences in Diversity and Complexity between episodes, neither for hand movements nor speech, while we found a decrease in Complexity but not in Diversity for both hand movements and speech in the Short-Long condition. Such an influence of order of presenting stimuli has been found before (Schöner & Thelen, 2006), and is in line with a widely known phenomenon of a system's current state being dependent on what happened before, i.e. on its history (e.g. Kelso, 1995). A possible explanation for our findings that involves history-dependence is that a salient distance-dimension influences hand movements' and speech's Diversity and Complexity, but a non-salient distance-dimension does not. This would mean that in the Long-episodes in both conditions, Diversity and Complexity of hand movements and speech changed when the participants started with the salient distance-dimension. However, in the Long-Short condition Diversity and Complexity did not change back to the previous state when being presented with the non-salient distance-dimension, hence we did not find a difference. Since we did not measure participants' Diversity and Complexity of hand movements and speech before and after the task, this explanation for the different findings in both conditions, based on the influence of a salient distance-dimension, remains speculative.

Furthermore, we found a difference in Complexity between episodes in the Short-Long condition, but not in Diversity. This means that the temporal organization of participants' hand movements and speech (Complexity) differed, while the frequency distribution of type and duration of hand movements and speech (Diversity) did not differ. Shockley, Butwill, Zbilut, and Webber Jr. (2002) found RQA-measures to pick up subtle changes in coupling characteristics that were missed by traditional linear measures. It could be that Complexity, also a RQA-measure, is more sensitive to changes in variability than Diversity, which would explain why we only found a difference in Complexity, but not in Diversity.

However, the direction of the difference in Complexity is opposite from what we expected. Instead of a decrease, we expected an increase in Complexity (and Diversity), because children were expected to adapt to the new task demand of considering the distance-dimension while working with a salient distance-dimension. Stephen, Dixon, and Isenhower (2009) found a peak in complexity, followed by a decrease in complexity of hand movements just before participants reported the discovery of a cognitive strategy. Perhaps participants in our study discovered the importance of the distance-dimension during the balance scale task. In line with Stephen et al.

Hand movements, speech, task properties, and variability

(2009), this might have led to an increase in Complexity of hand movements and speech in the Short-episode and a decrease in the Long-episode, which would have become evident as a decrease in Complexity between the two episodes. Because we did not measure whether participants discovered the importance of the distance-dimension, this argument remains speculative. Therefore, it is equally likely that we did not find an increase in Complexity because participants in the Short-Long condition did not gain new cognitive insights. Yet the difference in Complexity between the two episodes that we found does indicate that something happened around the switch from a non-salient to a salient distance-dimension. A follow-up study with a qualitative approach to analyzing the video-data might shed more light on what happened around that switch.

Lastly, contra to what we expected, we found a difference in Complexity for both hand movements and speech in the Short-Long condition. We expected an increase in Complexity (and Diversity) in hand movements only, because we expected that the change in spatiotemporal characteristics of the balance scale would influence hand movements more directly than speech, thereby leading cognitive development. When the change in spatiotemporal characteristics would equally influence speech and hand movements, we would expect the difference in Complexity between episodes of hand movements and speech to be related, but our additional analysis showed that this was not the case. Instead, as can be seen in Figure 5, participants varied in how a change in spatiotemporal characteristics of the balance scale simultaneously influenced the Complexity of their hand movements and speech before and after the switch. Follow-up research could investigate whether differences in the influence of task properties on the relation between hand movements' and speech's variability is related to different learning outcomes. Similarly, gesture-speech mismatches could also be viewed as changes in the relation between hand movements and speech (also see De Jonge-Hoekstra, Van der Steen, Van Geert, & Cox, 2016), and are indicative of learning. In addition, the apparent discrepancy between what we found on a group level in the Short-Long condition (i.e. a difference in Complexity for both hand movements and speech) and what individual participants showed (i.e. no relation between differences in Complexity of hand movements and speech) might illustrate a typical case of non-ergodicity (e.g. Molenaar & Campbell, 2009). A non-ergodic relation means that connections between variables on a group level are different from the connections between variables within people. Research with larger samples is needed to confirm or reject the existence of this non-ergodic relation.

Experiment 2

Method

Participants

A second and separate sample of 27 children from Kindergarten ($n = 18$) and first grade ($n = 9$), age 4 to 7 years ($M = 5.46$; $SD = 0.70$) participated in this experiment. The procedure of recruiting participants and ethical approval of the study was the same as in Experiment 1. The participants were randomly assigned to two conditions, in which the weights differed in mass (i.e. large vs. small mass, resp.; large vs. small difference in mass, resp.) and order of presenting this task property.

Materials

The materials used in Experiment 2 were the same as in Experiment 1, with two exceptions. Children in this experiment only worked with the long balance scale and they also worked with an additional pair of weights of 25 g (see Table 1).

Procedure

The general procedure in Experiment 2 was similar to Experiment 1. In the Large-Small condition participants worked with weights with a relatively large mass during the first four trials, while they worked with weights with a relatively small mass during the last four trials (see Table 1). Participants in the Small-Large condition first worked with the weights with a relatively small mass, followed by the four trials with weights with a relatively large mass.

Analysis

The coding procedure and analysis in Experiment 2 were similar to Experiment 1 (see Figure 3). As a brief reminder, we expected to find no difference in diversity and complexity, nor for hand movements, nor speech (hypothesis B). Regarding the analysis, the Large-episodes were compared to the Small-episodes in a similar manner to Experiment 1.

Results

In the Large-Small condition, we found no significant differences in Diversity between the first and the second episode, for neither hand movements ($M_{st. diff.} = 0.01$; $SD_{st. diff.} = 0.03$; $p = .10$; 95% $CI_{MC} = -.02, .02$) nor for speech ($M_{st. diff.} = 0.00$; $SD_{st. diff.} = 0.04$; $p = .46$; 95% $CI_{MC} = -.05, .05$). We did find significant differences in Complexity between the first and second episode, both for hand movements ($M_{st. diff.} = -0.05$; $SD_{st. diff.} = 0.05$; $p = .01$; 95% $CI_{MC} = -.03, .03$) and speech ($M_{st. diff.} = -0.09$; $SD_{st. diff.} = 0.04$; $p < .01$; 95% $CI_{MC} = -.04, .04$).

Hand movements, speech, task properties, and variability

In the Small-Large condition, we did not find significant differences in Diversity between the first and second episode, not for hand movement ($M_{st. diff.} = -0.01$; $SD_{st. diff.} = 0.03$; $p = .31$; 95% $CI_{MC} = -.04, .04$) nor for speech ($M_{st. diff.} = 0.02$; $SD_{st. diff.} = 0.05$; $p = .12$; 95% $CI_{MC} = -.02, .02$). Lastly, we found significant differences in Complexity between both episodes, both for hand movements ($M_{st. diff.} = -0.03$; $SD_{st. diff.} = 0.02$; $p = .02$; 95% $CI_{MC} = -.02, .02$) and speech ($M_{st. diff.} = -0.03$; $SD_{st. diff.} = 0.02$; $p = .01$; 95% $CI_{MC} = -.02, .02$).

These results are not in line with our first hypothesis (1B) that we would find no significant differences between episodes in Diversity and Complexity for both modalities. Instead, in both conditions we found a significant decrease in Complexity between episodes for both modalities, while we found no significant differences in Diversity for both modalities.

Similar to Experiment 1, we additionally analyzed whether the standardized differences between episodes of hand movements and speech were related (see Figure 6). In the Large-Small condition, we found a moderate and significant positive correlation for Diversity ($r = .49$, $p = .05$, $CI_{MC} = -.55, .48$) as well as for Complexity ($r = .58$, $p = .04$, $CI_{MC} = -.52, .53$). In the Small-Large condition, we found a low and insignificant positive correlation for Diversity ($r = .32$, $p = .20$, $CI_{MC} = -.64, .53$) and a negligible and insignificant negative correlation for Complexity ($r = -.03$, $p = .47$, $CI_{MC} = -.58, .59$). These results show that the standardized differences between episodes of Diversity and Complexity of hand movements and speech were related in the Large-Small condition, but unrelated in the Small-Large condition.

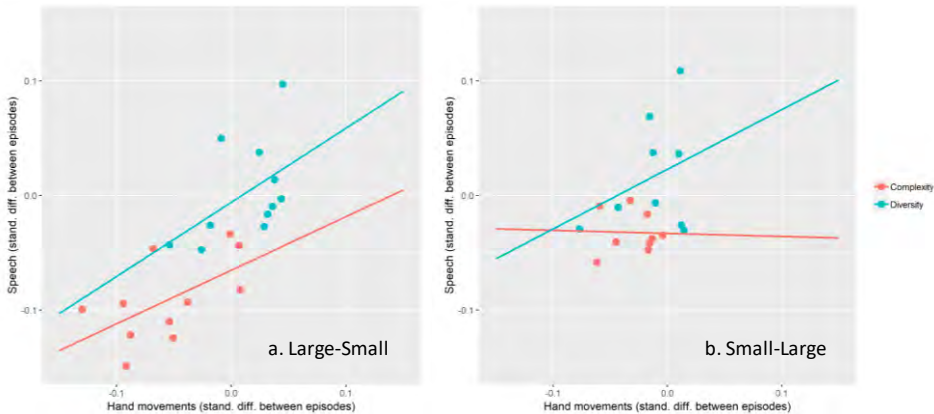


Figure 6. Relation between hand movements and speech with regard to standardized difference between episodes of Diversity (blue) and Complexity (red) in the Large-Small condition (panel a, $r_{Diversity} = .49$, $p_{Diversity} = .05$, $r_{Complexity} = .58$, $p_{Complexity} = .04$) and the Small-Large condition (panel b, $r_{Diversity} = .32$, $p_{Diversity} = .20$, $r_{Complexity} = -.03$, $p_{Complexity} = .47$).

Discussion

For Experiment 2 our second hypothesis was that we would find no difference in Diversity and Complexity, nor for hand movements or speech (hypothesis B). Contrary to our hypothesis, in both the Large-Small and Small-Large conditions we found a significant decrease of Complexity, but not of Diversity, between episodes for both hand movements and speech. Similar to Experiment 1, we attribute the found difference in Complexity but not Diversity to RQA measures' higher sensitivity to changes in variability (Shockley et al., 2002). Dissimilar to Experiment 1, we did not find different results for the two conditions. This implies that the change in Complexity might not be related to the direction of the change in saliency of the weight-dimension, but to things that both conditions had in common.

First, participants in both the Large-Small and Small-Large condition worked with a long balance scale throughout the whole task. We expected no difference in Complexity (and Diversity) because changing the salience of the weight-dimension converges with the "old" task demand to consider the weight-dimension, at which children between 4 to 7 years (as in our sample) are skilled already. However, the new task demand to consider the distance-dimension may have been introduced by presenting children with the long balance scale. Again in line with Stephen et al. (2009), the discovery of the importance of the distance-dimension might have led to an increase in Complexity of hand movements and speech in the first episode and a decrease in the second episode. Again, since we did not measure whether participants discovered the importance of the distance-dimension, this argument remains speculative.

Second, participants in both conditions experienced a change in the salience of weight. Maybe the task property -i.e. small (difference in) mass vs. large (difference in) mass- itself does not influence children's hand movements and speech, but instead the change in saliency of the weight dimension, regardless of direction of change, does. Moreover, if considering the weight-dimension in balance scale problems is a new task demand for participants, the decrease in Complexity between the two episodes might reflect their adaptation to this new task demand. Because we did not measure participants' initial understanding of the weight-dimension in balance scale problems before they participated, this argument also remains speculative. A counterargument against the unimportance of the direction of change in the saliency of the weight dimension comes from Fitzpatrick, Bui, and Garry (2018). Fitzpatrick et al. found that children less easily uncovered relevant weight-information in a hammering task when the weight-dimension was less salient. Furthermore, Beilock and Goldin-Meadow (2010) found that switching the weights of the disks in a Tower of Hanoi-task for adults, who gestured while they explained their solution, disrupted -and thus not benefitted- their learning process.

Hand movements, speech, task properties, and variability

While participants in both the Large-Small and Small-Large condition showed a decrease in Complexity and no difference in Diversity, only in the Large-Small condition we found the difference in Complexity and Diversity between episodes of hand movements and speech to be related. This suggests that a change from salient to non-salient weight-dimension in the Large-Small condition affected the change in variability of hand movements and speech to a similar degree within participants, and that this influence even is evident for the less sensitive variability measure of Diversity. Perhaps the combination of the long balance scale and heavy weights in the Large-episode resulted in a strong increase of force (i.e. long arm stretch, large mass) that was needed to hang weight at the balance scale. This task demand of exerting a strong force could be a new task demand in itself to which participants needed to adapt, and which would go together with an increase in variability. In the Small-episode, with weights with a small mass, children no longer needed to adapt to the task demand of exerting a large force, which would result in a decrease of variability again. Because hand movements and speech are tightly coupled, this perturbation of hand movements would also extend to speech. In line with this, Pouw, Harrison, and Dixon (2019) found that forcefully moving one's arms directly and physically affects speech.

Although the account above does explain why the change in variability of hand movements and speech between episodes of the Large-Small condition is related, it does not explain why we found a decrease in Complexity for hand movements and speech between the two episodes of the Small-Large condition. Maybe participant's experience with the task in the Small-episode guards them from the perturbation of the large force that they need to exert in the subsequent Large-episode. A follow-up experiment using only the small balance scale could show whether a smaller force would lead to different patterns of changes in hand movements' and speech's variability.

General discussion

With this study, we aimed to understand whether and how hand movements' leading role in cognitive development is related to its ability to correspond to spatiotemporal task properties, while speech is unable to do so. We therefore investigated how a change in the salience of the distance- or weight-dimension influenced hand movements' and speech's Diversity and Complexity. As a brief reminder, Diversity of behavior reflects functional adjustment to new task demands, and Complexity of behavior reflects functional flexibility when changes in task properties demand it.

A nuanced picture emerged from our findings. In Experiment 1, where we changed the salience of the distance-dimension, we found no significant differences in Diversity and Complexity in the Long-Short condition, while we found a significant decrease in Complexity for both hand

movements and speech in the Short-Long condition. We tentatively suggested 1) that the different findings in the two conditions fall under the larger phenomenon of history-dependence (or hysteresis), and 2) that the decrease might actually follow upon an increase in the previous episode. Furthermore, we found no significant relation between hand movements' and speech's change in Diversity and Complexity for both conditions. We proposed follow-up studies to investigate whether participants' relation between hand movements' and speech's change in Diversity and Complexity is related to learning outcomes, because gesture-speech mismatches could also be viewed as changes in the relation between hand movements and speech (De Jonge-Hoekstra et al., 2016).

In Experiment 2, where we changed the salience of the weight-dimension, a nuanced picture also emerged from our findings. We found a significant decrease in Complexity but not in Diversity for both hand movements and speech in the Large-Small as well as the Small-Large condition. We speculated that the similar findings in both conditions might have originated from similarities between the conditions, such as a long balance scale and a change in salience of the weight-dimension. In addition, we found a significant correlation between hand movements' and speech's change in Diversity and Complexity for the Large-Small condition only. We tentatively proposed that the force needed to hang heavy weights at distant hooks perturbs hand movements considerably, which in turn influences speech (Pouw, Harrison et al., 2019), but only when participants just started with the task, and thus are less experienced.

In regards to the aim of this study, most changes in spatiotemporal task properties seem to influence and decrease both hand movements' and speech's functional flexibility (Complexity). We found no differences in *whether* spatiotemporal task properties influence hand movements' and speech's variability. Our findings therefore do not suggest that hand movements' leading role in cognitive development stems from its ability to correspond to spatiotemporal task properties, while speech is unable to do so. However, our findings seem to indicate that there are differences in *how* spatiotemporal task properties influence hand movement's and speech's variability, except when participants start the task with a salient distance- and weight-dimension.

We might explain these differences from the perspective of affordances. Affordances are an agent's possibilities for action in their (current) environment (Chemero, 2003; J. J. Gibson, 1966; Stoffregen, 2003; see also Adolph & Kretch, 2015). An example of such possibilities for action are the different ways in which a baby descends slopes with different angles, such as stepping, sliding, or going backwards (Adolph et al., 2015). Most, if not all of our movements, show this dependency on spatiotemporal properties of the environment, whereby we need to adapt our movements to the environment in order for them to be functional. On the other hand, we do

Hand movements, speech, task properties, and variability

not have to adapt our speech to the spatiotemporal properties of the environment, but to our social environment instead. Speech is functional when it is clearly identifiable for a listener (Fowler, 2010). Smith and Gasser (2005) even propose that speech's functionality would be limited by a too close resemblance of physical structure in the structure of speech. In regards to our findings, it might be that changes in the spatiotemporal affordances influenced hand movement's variability, while changes in the social affordances influenced speech's variability. An example of such a change in social affordances might be trying to explain something clearly, while not being sure from time to time whether one understands how it works, or switching between refraining or not refraining from an explanation. Future research could investigate the circumstances under which changes in hand movement's and speech's variability do and do not occur together, and whether this is meaningful in terms of learning (i.e. when both the spatiotemporal and social affordances change).

An alternative explanation for our findings is connected to the pattern of a decrease in Complexity between the two episodes that we found in three of the four conditions. Maybe this decrease does not result from the change in task property, but reflects an order-effect. For all participants, the experimental setting and task is new, which might require them to adapt and might have caused an increase in Complexity during the first episode. In the second episode, participants are more used to the experimental setting and task, which would go together with a decrease in Complexity. Interestingly, Stephen et al. (2009) found that random changes in task properties induced variability in hand movements, and actually increased the likelihood of finding a new cognitive strategy. Future studies could try to disentangle how different types of changes in task properties (e.g. magnitude, newness, random) influence variability and cognitive change, and whether their influence is mutual. Furthermore, if the decrease in Complexity stems from getting used to the experimental setting, it is unclear why we did not find this order-effect in one condition, and why the changes in variability of hand movements and speech were related in another condition.

A first limitation of our study is that we did not measure participants' understanding of the weight and distance dimension before and after the task. Therefore, any relation between changes in spatiotemporal task properties, variability of hand movements and speech, and cognitive change remains unsubstantiated. While we believe that our study provides valuable insight into the influence of changes in spatiotemporal properties on changes in hand movements' and speech's variability, more research is needed to establish a link to learning.

A second, potential, limitation is the age range (4-7 years) of participants in our study. Children's cognitive skills develop tremendously between 4 and 7 years of age, and this might influence whether they understand the influence of the weight- or distance-dimension in balance skill

problems, thus possibly confounding the influence of our manipulation of spatiotemporal task properties. Accordingly, while children as young as 4 years old have been found to consider the weight-dimension when they solve balance scale problems (Schrauf et al., 2011), also many 4-year-olds do not. Since we did not measure participants' understanding of the weight and distance dimension before and after the task, we cannot formally analyze this potential relation between age and understanding. However, careful (post-hoc) inspection of the video recordings did not provide evidence for age-related differences in children's performance. Therefore, we speculate that age is not a relevant factor in explaining the results we found. For example, a number of 4-year-olds already seem to grasp the importance of distance from the fulcrum for balance scale problems, while a number of 6-year-olds have difficulty to understand the importance of mass of the weights on some of the trials. Instead, verbal reasoning skills and previous experience seem more important than age with respect to children's (ability to acquire) understanding about balance scale problems. Future studies could investigate whether a change in spatiotemporal task properties is related to individual differences between children, such as age, verbal reasoning skills, and previous experience (see also De Jonge-Hoekstra et al., 2016).

A third limitation is the crude coding system that we used to categorize hand movements and speech, with only four categories for each modality. More fine-grained measures are able to capture changes in hand movements and speech, and therefore in their variability, in more detail. Pouw and Dixon (2019b) for instance used very dense (240 Hz, ~ 0.13 mm spatial resolution) continuous measurements of hand movements and speech to investigate how changes in intensity of the two modalities are related. Nevertheless, because we coded hand movements and speech at 2 Hz, even these four categories per modality can capture part of the complex temporal organization, as can be seen in the time series examples in Figure 3. Future research could investigate how variability on these different measurement and time scales is related.

Our study has several methodological implications. To our knowledge, this study is the first to use the entropy-measure for categorical RQA, as proposed by Leonardi (2018). We highlighted how this measure can be used to investigate empirical time series, and showed that the entropy-measure is sensitive to experimentally manipulated changes. This entropy-measure could be extended to Cross RQA, to investigate whether the shared Complexity of two interacting systems, coded with similar coding systems (e.g. De Jonge-Hoekstra et al., 2016), informs about changes in the systems' coupling and shared state.

Furthermore, we believe that this study is the first to investigate the relation between the variability-measures Diversity and Complexity under different spatiotemporal task properties.

Hand movements, speech, task properties, and variability

We only found differences between episodes in Complexity, and never in Diversity, which made us think that Complexity is more sensitive to changes in variability than Diversity. Complexity's higher sensitivity is in line with our interpretation of Diversity as functional adaptation and Complexity as flexibility to adapt. In other words, whereas Diversity indicates *adapting itself*, i.e. *reorganizing*, Complexity indicates the *process by which adapting comes about*, i.e. *increased flexibility of a system that is about to reorganize* (potentially in a more adaptive state). In addition, Adolph et al. (2015) use examples of qualitatively different strategies (e.g. descending a slope by sliding, stepping, or going backwards) to explain why diversity of behavior is important for development. Our task manipulation did not require children to use qualitatively different strategies to perform the task, which might explain why we did not find any differences in Diversity. Adolph et al.'s (2015) examples for changes in the structure of behavior (i.e. clumsy and rigid steps of a new-walker vs. smooth and flexible steps of an adult walker) seem to be closer to the behavioral changes that children were required to make between episodes. This might also explain why we indeed found differences in Complexity. Future studies could investigate whether changes between qualitatively different strategies will influence only Diversity, or both Diversity and Complexity, which would be in line with Complexity being a more sensitive variability measure. Previous studies about changes in Complexity when people discovered qualitatively different cognitive strategies (e.g. Anastas, Stephen, & Dixon, 2011; Stephen, Boncoddio, Magnuson, & Dixon, 2009; Stephen, Dixon, et al., 2009) suggest the latter.

Our study adds to the field of hands-on learning. From previous studies, we know that children use their hands to learn (Kuhn et al., 2009; Zhang, 2019), and that asking children to explain what they are doing further increases their understanding (Van Der Steen et al., 2014, 2019). Based on our findings, changes in the saliency of spatiotemporal task properties seem to influence hand movements' and speech's variability in a nuanced way, but only when certain circumstances, such as the order and magnitude of the changes, are met. Furthermore, the changes in variability between hand movements and speech seem to be unrelated, most of the time. Abney, Paxton, Dale, and Kello (2015) investigated participants in a dyadic task and found that weak coupling and role structure is functional for dyadic problem solving. Perhaps certain hands-on learning activities elicit a similar weak coupling and role structure (e.g. the spatiotemporal vs. social affordances) between hand movements and speech as well, which might explain why we found no relation in changes of variability between the two modalities. De Jonge-Hoekstra et al. (2016) indeed found that differences in gesture-speech coupling during a science and technology task are related to performance on past tasks and to standardized math scores. Future research could investigate under which circumstances a stronger or weaker coupling between hand movements and speech is functional for learning.

Conclusion

In this study, we explored whether and how hand movements' leading role in cognitive development is related to its ability to correspond and adapt to spatiotemporal task properties, while speech is unable to do so. We used new analysis methods to investigate changes in hand movements' and speech's Diversity and Complexity. In short, we found that differences in how hand movements and speech correspond to spatiotemporal task properties do not simply explain hand movement's leading role in cognitive development. Instead, we found that both hand movements' and speech's Complexity changes with changing spatiotemporal task properties most of the time, but that these changes are only mutually related in one out of four conditions. This study generates more questions than it answers, and we aimed to address these follow-up questions and provided multiple directions for future research in the extensive Discussion sections of this paper. Our study follows theoretical accounts that explain cognition as intertwined with all levels of human behavior, and as inseparable from perception and action of persons in their environment (e.g. Chemero, 2011; Smith & Thelen, 2003; Smith, 2005). To conclude, we hope that our study serves as a starting point to investigate how these theoretical accounts of cognition can explain how actual children learn and reason about how the world works.

Acknowledgements

The authors would like to thank the primary schools that facilitated the data collection and the students who participated in coding children's gestures and speech in the video-recordings. Furthermore, we would like to thank the reviewers of the previous manuscript for their constructive feedback.



5

Get it together: How do school-aged children interact multimodally and at multiple time scales when they solve dyadic balance scale problems?

A revised version of this chapter will be submitted as:

De Jonge-Hoekstra, L., Pouw, W., Van der Steen, S., Dixon, J.A., & Cox, R.F.A. (2021). Get it together: How do school-aged children interact multimodally and at multiple time scales when they solve dyadic balance scale problems?

Get it together: How do school-aged children coordinate their multimodal behavior at multiple time scales when they solve dyadic balance scale problems?

Interpersonal coordination refers to how people adjust their behavior to each other when they do things together (Dumas & Fairhurst, 2019), and therefore applies to many of our social interactions. While the number of studies about interpersonal coordination between adults is steadily increasing (e.g. Abney, Paxton et al., 2014; Fujiwara et al., 2019b; Paxton & Dale, 2017), we know relatively little about interpersonal coordination between school aged children (cf. Rauchbauer, 2020; Vink et al., 2017; Xavier et al., 2018). In this study, we researched how school aged children (age: 6-10 years) coordinate their speech, hand movements, and head movements at multiple timescales when they solve balance scale problems together. Furthermore, we investigated whether these measures of interpersonal coordination predict children's task performance.

Interpersonal coordination

During social interactions, people change their behavior at multiple levels. On a physiological level, they tend to synchronize their heart rate and respiration (Konvalinka et al., 2011; Tschacher & Meier, 2020). On a physical level, people tend to align their body movements, such as rocking in rocking chairs (Richardson et al., 2007), walking (Almurad et al., 2017), or postural sway (Shockley et al., 2003). In addition, people imitate and mirror each other's body movements when they do something together (Tomasello, 2008). Furthermore, in many social interactions people respond to each other, and they adapt their behavior to the previous and upcoming behavior of their interaction partner, sometimes even across many timescales. In adult behavior, research shows that, among other things, people adapt their speech, hand movements, and head movements to each other when they interact (Holler & Levinson, 2019).

Speech

With regard to interpersonal coordination of speech, people align their utterances in terms of timing, semantics, syntax, and prosody (e.g. Fowler et al., 2007; Fusaroli et al., 2012, 2014; Pickering & Garrod, 2004; Reed, 2010; see also Rasenberget al., 2020 for a recent review). Moreover, research has shown that speech alignment spans many time scales. For instance, Abney, Paxton et al. (2014) found that people who engaged in a friendly conversation coordinated their speech timing not only on a behavioral level (i.e. turn taking), but on a whole range of timescales, such as the timescales of phonemes, words, sentences, semantics, and turn taking. In other words, these conversations showed multiscale coordination. On the other hand, people who engaged in an argumentative conversation coordinated their speech timing

Multiscale multimodal coordination during dyadic problem solving

on a behavioral level only. Fusaroli et al. (2013) showed that dyads' multiscale coordination of speech timing increased during the course of a perceptual decision-making task. Furthermore, this multiscale coordination of speech timing - but not similarity in speech diversity, turn taking, or initial coordination - was related to dyadic performance. Interestingly, multiscale coordination of speech timing between infants and caregivers is increasingly evident during the first two years of life already (Abney et al., 2017). According to Fusaroli et al., (2014; see also Fowler et al., 2007), different forms of linguistic alignment naturally emerge and self-organize when people engage in a dialog, and become an interpersonal synergy (i.e. a social unit).

Hand movements

Regarding interpersonal coordination of hand movements, two different strands of research are relevant (however, see Furuyama, 2002, for a combination). The first strand of research addresses the alignment of people's hand gestures during conversations (for a recent review, see Rasenberg et al., 2020). For example, interacting people use hand gestures that are similar, with respect to a number of form features (Bergmann & Kopp, 2012). Holler and Wilkin (2011) found that people even mimic each other's hand gestures during dialog. Furthermore, this gesture-mimicry is thought to enhance shared understanding between people (Holler & Wilkin, 2011). In addition, people mimic and extend each other's hand gestures, to further increase shared understanding (Chui, 2014).

Moving beyond gestures and mimicry, the second strand of research centers around coordination of people's hand movements during interpersonal tasks. For instance, people synchronize their hand movements while oscillating hand-held pendulums, when they are able to see each other's hand movements (Schmidt & Turvey, 1994). Richardson et al. (2005) extended these findings to a task in which people were asked to solve a puzzle together. They found that seeing each other's movements automatically leads to interpersonal hand movement synchronization, while merely conversing, without seeing each other, does not. Furthermore, interpersonal hand movement synchronization has also been found in handclapping games (Schmidt et al., 2011). In line with Fusaroli et al. (2014), Schmidt and Richardson (2008) state that (hand) movement coordination naturally emerges when people interact and become an interpersonal synergy.

Head movements

With regard to interpersonal coordination of head movements, people tend to move their heads ubiquitously during conversations (for a comprehensive review, see Wagner et al., 2014). Particular head movements patterns seem to be related to conversational roles. For example, head nodding is more evident in *listeners*, while *speakers* tend to move their heads according to the prosodic and deictic properties of their speech (Esteve-Gibert et al., 2017; Wagner et al.,

2014). Furthermore, Louwerse et al. (2012) investigated dyads in a directions giving task, and found that the instruction *giver's* head nodding was leading the instruction *follower's* head nodding. In addition, overall matching between dyads' head movements increased over the course of the task, i.e. when the dyad had interacted more. This increase in overall matching is similar to Fusaroli et al.'s (2013) results regarding speech timing. In a recent study, Hale et al. (2019) found that people tend to mimic low frequency (0.2 – 1.1 Hz) head movements with a delay of 600ms, but they also found a decoupling of people's high frequency (2.6 – 6.5 Hz) head movements. These high frequency head movements stem from fast nodding by the listener, and Hale et al. (2019) propose that this pattern of decoupling functions as a social signal to coordinate the structure of a conversation. Besides this, head movements in general are thought to be central to establishing common ground between conversing people (Wagner et al., 2014). Interestingly, also for interacting musicians, head movements are driving synchronized behavior, and for structuring their joint music playing (Bishop & Goebel, 2018).

Multimodality at different timescales

The previous studies described interpersonal coordination of speech, hand movements, and head movements in isolation. However, social interactions are typically multimodal (e.g. Holler & Levinson, 2019; Macuch Silva et al., 2020; Wagner et al., 2014). During these multimodal social interactions, coordination of speech, hand movements, and head movements seems to center around different time scales (Hale et al., 2019; see also Pouw & Dixon, 2019b).

Based on previous studies, Pouw and Dixon (2019b) distinguished three timescales to investigate gesture-speech synchronization: A fast, a middle, and a slow timescale. The fast timescale ranged from 0.125 to 0.5 s (8 – 2 Hz), and captures the average length of a syllable (Pouw & Dixon, 2019b). Furthermore, a recent review shows that the coordination of speech at the fast timescale is remarkably robust across speakers, languages, and situations (Poeppel & Assaneo, 2020). Besides syllables in speech, the fast time scale also captures listeners' high frequency head movements (2.6 – 6.5 Hz; Hale et al., 2019). The middle timescale ranged from 0.5 to 2 s (2 – 0.5 Hz), and captures the timescale at which gestures occur (Pouw & Dixon, 2019b). The slow timescale ranged from 2 to 4 s (0.5 – 0.25 Hz), and corresponds to the timescale of clauses and sentences (~2.00 – 6.00 s, or 0.5 – 0.16 Hz; Pouw & Dixon, 2019b). In addition, coherence between head movements at a time scale from 0.2 – 1.1 Hz corresponds to mimicry (Hale et al., 2019), and thus is captured by both the middle and slow timescale.

In the current study, we will distinguish these same fast, middle, and slow timescales to investigate coordination of school aged children's speech, hand movements, and head movements when they solve balance scale problems together. In the next paragraphs, we will discuss how children collaborate, and how they communicate multimodally while collaborating.

Multiscale multimodal coordination during dyadic problem solving

Collaboration in children

Most of the time, young children do not spontaneously collaborate when they are asked to do a task together. Instead, children between 4 and 6 years of age prefer to work in parallel (i.e. 10-20% of the interactions), and collaborate for only 0-5% of the interactions (Guevara Guerrero, 2015). In addition, Guevarra-Guerrero (2015) found no relation between interaction type (i.e. in parallel, collaboration) and children's task performance. However, Fawcett and Garton (2005) found that children between 6 and 7 years of age performed better when they collaborated with a higher performing peer. These mixed findings could be due to age differences, as another study found that first graders ($M_{age} = 7y$) improved their performance when they jointly solved causal interventions together with an adult, while kindergartner's ($M_{age} = 6y$) performance decreased (Young et al., 2019).

The literature about collaborative learning emphasizes the importance of verbal interactions for collaboration (also see Rieber & Carton, 1988; Rowe & Wertsch, 2002). In a key study, Teasley (1995) found that children around 10 years of age (range: 8.9 to 11.8 y), who worked and talked together in a dyad on a scientific reasoning task, outperformed children who worked alone and either did or did not talk (to themselves), or children who worked together but did not talk. Similarly, in a study with adult participants, dyads with nearly equal information outperformed the better participant in a perceptual decision-making task, but only when a dyad was allowed to communicate freely (Bahrami et al., 2010). Although children's collaborative performance seems to benefit from verbal interactions with sophisticated linguistic alignment, these high quality verbal interactions actually rarely occur when children work together (Molenaar et al., 2014; Storch, 2001; Weinberger & Fischer, 2006).

Children not only collaborate by means of verbal interactions, but they also show non-verbal behaviors, such as gestures, nodding, changes in posture, and manipulating task materials. For example, Yliveronen, Marjanen, and Seitamaa-Hakkarainen (2016) investigated how six-year-olds designed animal shelters together, and found both verbal and non-verbal behaviors to be important for collaborating. A similar central role for verbal and non-verbal behaviors has been found for children between 9 and 10 (Taylor, 2014), between 10 and 13 (Granott et al., 2002), and between 15 and 16 (Roth, 2001) years of age, who collaborated and co-constructed meaning in the classroom. The studies above suggest that coordination of speech, hand movements, and head movements might be important for collaboration in (school-aged) children, and the task performance which stems from it. In addition, we know that children interpersonally coordinate and synchronize their (whole) body movements with others (i.e. adults and peers) during social interactions, both at early ages (Cirelli, 2018; Endedijk et al.,

2015; Meyer & Hunnius, 2020) and beyond (Rauchbauer, 2020; Satta, 2017; Vink et al., 2017; Xavier et al., 2018).

Current study

As described above, much is known about how adults coordinate their speech, hand movements and head movements during social interactions, and at multiple timescales. To our knowledge, no studies have investigated whether school aged children coordinate their multimodal behavior at these same timescales, however. As school aged children spend most of their time together with peers, learning what school aged children's interpersonal multimodal coordination looks like will yield important understanding about a major and influential component of children's everyday life. From the studies described above, we do know that children communicate both verbally and non-verbally when they collaborate. In addition, these verbal and non-verbal behaviors seem to be important for task performance during collaborative tasks. However, as far as we know, no studies have investigated whether measures of interpersonal coordination of school aged children's multimodal behavior at certain timescales are related to their collaborative task performance. Given the importance of successful collaboration in both child- and adulthood, it is vital to understand whether and how children's interpersonal multimodal coordination contributes to collaborative task performance. We researched how school aged children (age: 6-10 years) coordinate their speech, hand movements, and head movements at multiple timescales when they solve balance scale problems together, and how these measures of interpersonal coordination predict children's task performance, thereby addressing the above stated gaps in our knowledge.

In the current study, we investigated school aged children's interpersonal coordination while they engaged in a dyadic scientific reasoning task at a science center. During the dyadic scientific reasoning task, we asked each child to predict about balance scale problems individually, and to discuss their predictions within their dyad if one or both of the children made an incorrect prediction about a balance scale problem (for a similar procedure, see Bahrami et al., 2010; Fusaroli et al., 2012). After each bout of discussion, we asked children to predict the outcome of the balance scale problem individually again.

To investigate how dyads of school aged children coordinate their speech, hand movements, and head movements during discussion moments, on multiple timescales, we used *cross wavelet analysis* (Grinsted et al., 2004; also see Rösch & Schmidbauer, 2016). In short, cross-wavelet analysis is a method to compare two time series' similarity over time, in terms of oscillations, at a range of frequencies (i.e. timescales), that reside in both time series (see Figure 3; a detailed explanation will follow in the Method-section). This similarity of co-occurrent time

Multiscale multimodal coordination during dyadic problem solving

series' oscillations is captured by the *degree of coherence* (see Figure 1), which ranges from 0 (no coherence) to 1 (perfect coherence). Previous studies have shown that significant (i.e. above chance) coherence at different timescales is evident between whole body movements of joke tellers and listeners (Schmidt et al., 2014), and autistic children and clinicians (Romero et al., 2018), and between head movements of speakers and listeners (Hale et al., 2019). Furthermore, Pouw and Dixon (2019b) used cross wavelet analysis to investigate the coherence between gestures' and speech's oscillations within participants.

In addition to similarity of oscillations in time series, cross wavelet analysis yields *relative phase angles* (see Figure 1; and Figure 4b). The relative phase angle characterizes whether the oscillations are more in-phase (0° ; in synchrony) or anti-phase (180° ; alternating), and which oscillations are leading. Previous findings are mixed about the presence of a distinct relative phase angle relation between oscillations across social interactions (Hale et al., 2019; Romero et al., 2018; Schmidt et al., 2014).

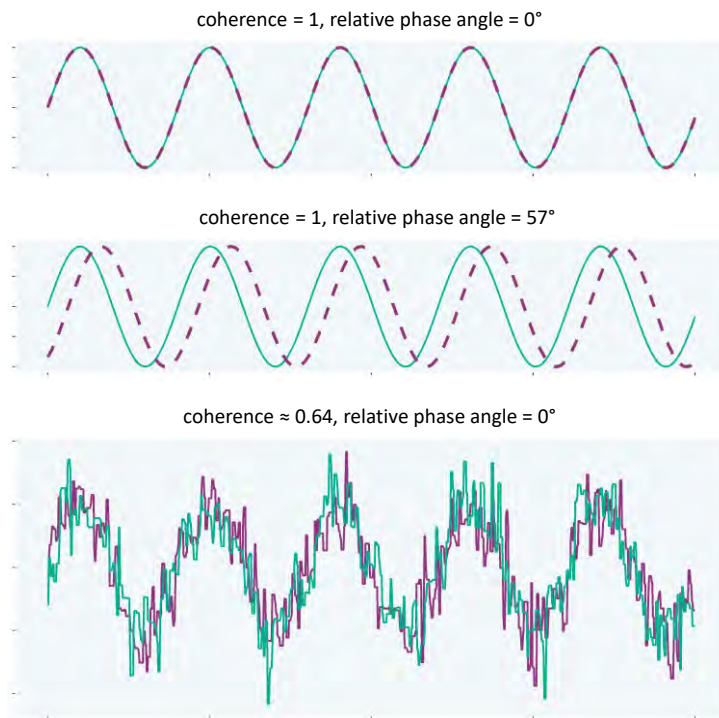


Figure 1. The top and middle panels show oscillating signals with perfect coherence (coherence = 1), while the bottom panel shows oscillating signals with imperfect coherence (coherence ≈ 0.64). Coherence can range from 1 (perfect coherence) to 0 (no coherence). Furthermore, the top and bottom panels show oscillating signals with no relative phase difference (relative phase angle = 0°), while the middle panel shows oscillating signals with a relative phase angle of 57° . The relative phase angle can range from 0° (in-phase; in synchrony) to 180° (anti-phase; alternating, also see Figure 3b).

Research questions

Our first research question is: What are the differences in coherence between the fast (8 – 2 Hz), middle (2 – 0.5 Hz), and slow timescales (0.5 – 0.25 Hz) for the oscillations of school aged children's speech, hand movements, and head movements across discussion moments? Previous studies have shown that for adults, speech is mostly coordinated at the fast timescale (Poeppel & Assaneo, 2020; Pouw & Dixon, 2019b), hand movements at the middle timescale (Pouw & Dixon, 2019b), and head movements at both the middle and slow timescale (Hale et al., 2019). Because this is the first study to investigate the coherence of oscillations of school aged children's multimodal behavior, we felt that we lacked the sufficient empirical and theoretical grounds to formulate hypotheses.

Our second research question is: How do the measures of coherence relate to dyadic task performance? As indices of task performance, we included: a) change in number of correct predictions from before to after the discussion moments, b) agreement of predictions after discussion moments, and c) adopting the other child's pre-discussion prediction for one's own later prediction (which we will further explain in the Method-section). For the same reasons as for the first research question, we did not formulate hypotheses for the second research question.

In contrast to the findings about the presence of significant coherence between people during conversations, previous findings about relative phase angle relations are equivocal (Hale et al., 2019; Romero et al., 2018; Schmidt et al., 2014). Thus, in addition to our two main research questions, we also included two exploratory research questions regarding relative phase relations: a) What are the differences in relative phase angle between the fast (8 – 2 Hz), middle (2 – 0.5 Hz), and slow timescales (0.5 – 0.25 Hz) for the oscillations of school aged children's speech, hand movements, and head movements across discussion moments?; b) How do the measures of coherence affect our three indices (see previous paragraph) of dyadic task performance?

Method

Participants

A total of 50 children (25 dyads) between 5.7 and 10.7 years ($M = 7.9$, $SD = 1.5$) of age participated in our study. In 9 out of 25 dyads the children had never met before. Of the other 16 dyads, 12 dyads were siblings (of which 6 dyads were twins), 2 dyads were cousins, and 2 dyads were friends. We recruited children between 6 and 10 years at the Connecticut Science Center in Hartford, as part of the so-called *Living Lab* concept. The Living Lab is a collaboration between the Connecticut Science Center and neighboring universities. Researchers from these

Multiscale multimodal coordination during dyadic problem solving

universities are invited to combine science communication with actual data collection on site, at the Connecticut Science Center. We asked both the children and their legal guardians whether they would like to participate in our experiment, and obtained written informed consent from the children's legal guardians. The Institutional Review Board of the University of Connecticut approved our study (protocol no.: H18-195).

Materials

Balance scale game

Participants played a game in dyads (see Figure 2 for the set-up of the experiment) in which they were asked to predict to which side a balance scale would tilt downwards. The balance scale had variously sized weights at specified distances from the fulcrum (see Procedure). Both the pegs on the balance scale and the weights did not include any numbers. We used photos of unreleased balance scales to elicit participants' predictions, and videos of releasing balance scales to later show the outcome (for a similar set-up, see Castillo et al., 2015; 2017). A real balance scale was used in the task introduction to the task only. We programmed the dyadic

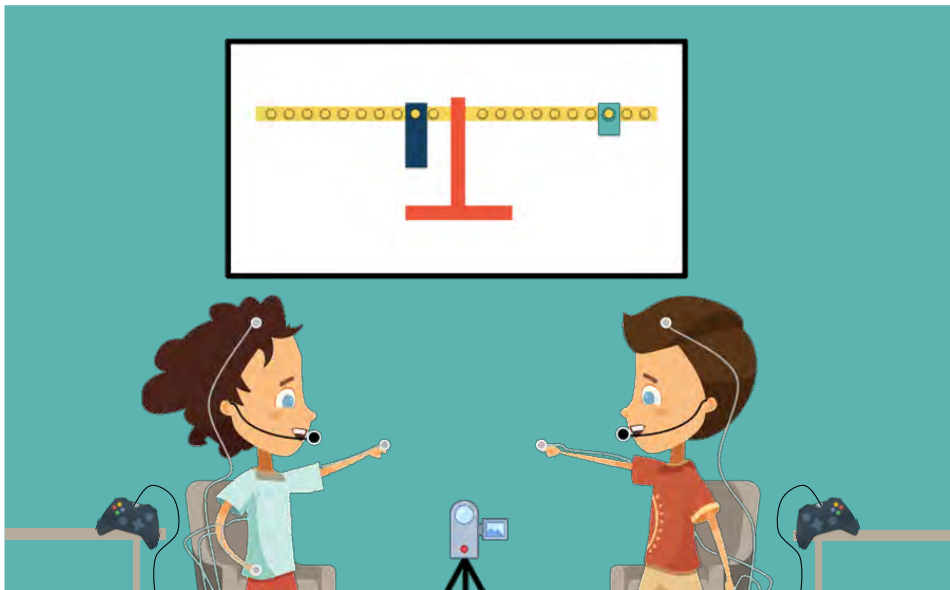


Figure 2. Set-up of the experiment. Children both sat on a chair. They had movements sensors (the grey dots) attached to the index fingers of both their hands, and on a tiara on their head (tiara not shown in the Figure). They wore head-worn condenser microphones (the black dots). The game controllers were attached to the tables, which stood next to the participants. Photos of balance scale problems and videos of the outcome of these problems were projected on a large tv-screen on the wall. Lastly, children's behavior was recorded using a video recorder on a tripod, which was positioned underneath the tv-screen.

game in OpenSesame [version 3.0.0] (Mathôt et al., 2012), which is an open-source, Python-based program to build (social science) experiments. The game was displayed on a large screen on the wall, in front of participants. We asked participants to predict by pressing one of three buttons on a Logitech F310 game controller (one for each participant). With OpenSesame, we recorded which button the participants pressed as well as the timing of their button presses.

Recording

We recorded participants' speech and the movements of both their hands and head. We used AKG C 520 head-worn condenser microphones to record participants' speech. These microphones are typically used to record a drummer's voice, which makes them particularly suited to record speech in a noisy environment such as the Connecticut Science Center. We used the Polhemus Liberty with 6 sensors to track participants' hands and head positions in 3D, which has a temporal resolution of 240 Hz and a spatial accuracy of ~ 0.13 mm. For each participant, we used medical tape to attach one sensor to both their right and left index fingers, and we attached an additional sensor to a tiara (not shown in Figure 2) which we then placed on their head. Furthermore, we video recorded participants while they played the game, using a Sony Digital HD Camera HDR-XR5504. We used ELAN (Max Planck Institute for Psycholinguistics, The Language Archive, 2020; Wittenburg et al., 2006) to identify the discussion moments in the videos.

We recorded several data streams, which brings specific challenges for the data collection with it. We needed two computers to save the output of the different data streams: One to synchronously save the OpenSesame output and audio of one participant, and another one to synchronously save the Polhemus motion data and audio of the other participant (for details, see Pouw & Dixon, 2019b; also see Richardson, 2009). To subsequently synchronize the different data streams that were saved at the two computers, and the video recordings, we used Adobe Premiere Pro (see Procedure for the details). We used Adobe Audition to remove background noise from the audio recordings.

Analysis

To analyze the data, we used R-Studio. Based on scripts by Pouw and Dixon (2019b), we wrote custom R-scripts to aggregate the motion data, speech data, OpenSesame output, and ELAN data. To investigate multiscale coordination between children's movements and between children's speech, we used the WaveletComp-package (Rösch & Schmidbauer, 2018). Furthermore, we used the circular-package (Lund et al., 2017) to calculate the circular mean of the relative phase angles, the mgcv-package (Wood, 2020) to perform general additive modelling, the lme4-package (Bates et al., 2020) to carry out multilevel analyses, the emmeans-package (Lenth et al., 2020) to perform posthoc analyses, and the MuMin-package (Barton,

Multiscale multimodal coordination during dyadic problem solving

2020) to calculate Nakagawa's and Schielzeth's (2013) marginal and conditional R^2 for generalized linear mixed-effects models. Lastly, we used ggplot2 (Wickham, 2016), and the ggplot2-extensions cowplot (Wilke, 2020b), ggtext (Wilke, 2020a) and ggforce (Pedersen, 2020) to create plots of our data, following many of the tips from Chase (2020).

Procedure

Before the participants engaged in the game, we first asked their legal guardians to fill out a small questionnaire and thereby indicate participants year and month of birth, and their relation to the other child in the dyad. After we asked participants to take a seat in one of two chairs, we attached the sensors to the participants' hands and head, and attached the microphone. We subsequently started video recording, motion tracking and audio recording.

We then introduced the game. We showed a real balance scale and attached equal weights at an equal distance from the fulcrum to the balance scale, and we asked participants whether they had ever seen something similar before. We then explained how participants should use the game controller's buttons to individually predict the outcome of balance scale problems. We instructed them to press the *blue* (left) button when they thought the balance scale would tilt to the *left*, press the *yellow* (middle) button when they thought the balance scale would stay *even*, and press the *red* (right) button when they thought the balance scale would tilt to the *right*. Furthermore, we told them that they could earn points during the game, and the better they would work together and try to explain things to each other, the more points they could earn.

After the introduction, the game began. The game consisted of a series of problem solving trials (loosely based on Fusaroli et al., 2012). As can be seen in Figure 3, each trial started with a photo of a balance scale problem on the large screen (t1 in Figure 3). We then asked participants to individually predict what would happen with the balance scale after it would be released, by means of pressing a button on the game controller (t2). To ensure that children's predictions were independent from each other, during each trial we explicitly urged them to not look at or talk to each other while making their initial prediction. Based on the individual predictions, the participants received feedback about their predictions: 1) [both] correct; 2) 1

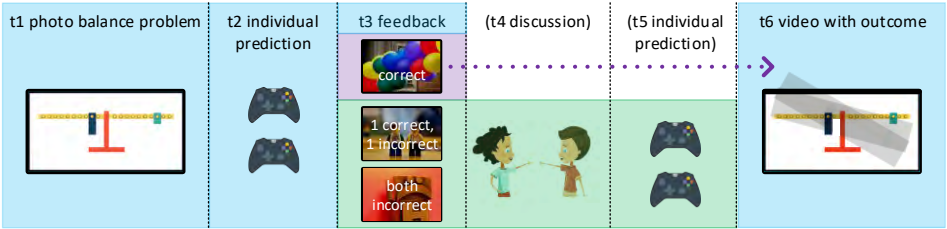


Figure 3. Time line of a trial.

correct, 1 incorrect; or 3) both incorrect (t3 in Figure 3). If both participants correctly predicted what happened, they received three points, and were shown a video of the outcome (release of the balance scale; t6). The dyad then moved to the next trial. If neither or only one of the participants made a correct prediction, they received either zero or one point(s), respectively, and were asked to explain to each other what they predicted and why they did so (t4). Some dyads ($n = 5$; see also Results) found it difficult to start the first discussion and remained silent, typically because they seemed shy. When this happened, we tried to help them along by asking each member of the dyad what they pressed and why. Since these trials were not dyadic, but triadic, we later removed these discussion moments from our analyses. After the discussion moment, we again asked participants to individually predict what would happen with the balance scale, by pressing a button (t5). The dyad's reconsideration of the problem affords the opportunity to assess the potential relationship between the interaction during the discussion and each participants' subsequent prediction. After this second prediction, participants were shown a video of the outcome (t6), and then proceeded to the next trial.

To increase the chances that participants would disagree or predict incorrectly, and thereby maximizing the opportunity and need for children to discuss, we incorporated difficulty levels in the game. We based the seven difficulty levels, which can be found in Table 1, on the work of Siegler (1976; see also Boom et al., 2001; Jansen & van der Maas, 2002; van Rijn et al., 2003). Participants always started with balance scale problems at Difficulty level 1 (see Table 1). The game's difficulty increased by one level if both participants made three correct, individual predictions in a row. We did not tell participants about the different difficulty levels.

Table 1
Overview of difficulty levels in the game

Difficulty level	Description	Outcome
1 - Balance	Two weights of the <i>same mass</i> at the <i>same distance</i> from the fulcrum	Balance scale remains <i>even</i>
2 - Weight	Two weights of a <i>different mass</i> at the <i>same distance</i> from the fulcrum	Balance scale tilts to the side with the <i>heaviest weight</i>
3 - Distance	Two weights of the <i>same mass</i> at a <i>different distance</i> from the fulcrum	Balance scale tilts to the side with the <i>most distant weight</i>
4 - Conflict weight	Two weights of a <i>different mass</i> at a <i>different distance</i> from the fulcrum	Balance scale tilts to the side with the <i>heaviest weight</i>
5 - Conflict distance	Two weights of a <i>different mass</i> at a <i>different distance</i> from the fulcrum	Balance scale tilts to the side with the <i>most distant weight</i>
6 - Conflict balance	Two weights of a <i>different mass</i> at a <i>different distance</i> from the fulcrum	Balance scale remains <i>even</i>
7 - Unsolvable with addition rule	Two weights of a <i>different mass</i> at a <i>different distance</i> from the fulcrum	Balance scale tilts to the side with the <i>largest product</i> of the [weight X distance] from fulcrum

Multiscale multimodal coordination during dyadic problem solving

The game automatically stopped after 10 minutes, after which participants' final score appeared on the screen. We then detached the sensors and microphones, and gave each participant a certificate with their name and score.

Data preparation

We recorded several data streams, which we needed to synchronize afterwards. We used Adobe Premiere Pro's "Synchronize based on audio" feature for this (see also Pouw, Trujillo et al., 2019). First, we synchronized the video recording with the first audio stream, which was already synchronized with the motion tracking data streams. Second, we synchronized the (synchronized) video recording with the second audio stream, which was already time-locked with the OpenSesame output. Third, we could use the completely silent period at the start of the second audio stream after synchronization with the (synchronized) video to automatically infer the delay between the OpenSesame output and the other data streams, using a custom R-script. After this third step, all data streams were synchronized, and ready for further processing.

We measured the intensity of participants' speech, their head movements, and their hand movements. Our procedure to convert the raw speech and motion data into timeseries data that we can analyze, largely follows Pouw and Dixon (2019b) and Pouw, Trujillo et al. (2019). With regard to speech, we first removed background noise from the audio recordings, using Adobe Audition. We then calculated the smoothed (5Hz Hanning filter) amplitude envelope of the speech signal as proposed by He and Dellwo (2016), using an R-script from Pouw and Trujillo (2019). The amplitude envelope is a smoothed trace of the audio waveforms amplitude fluctuations (He & Dellwo, 2016). The amplitude envelope crucially defines quasi-rhythmic aspects of speech (Tilsen & Arvaniti, 2013), and is highly correlated with lip aperture kinematics during speaking (Chandrasekaran et al., 2009; He & Dellwo, 2017). With regard to movements, we used a custom R-script based on Pouw, Trujillo et al. (2019) to calculate the velocity of participant's head and hands. We applied a low-pass first-order Butterworth filter to the velocity time series with a cut-off of 33 Hz. We subsequently aggregated the motion and speech data of each participant at 240 Hz, again using a custom R-script based on Pouw, Trujillo et al. (2019).

We were interested in participants' speech and movements during bouts of discussion (i.e. during the t4 period in Figure 3). We used ELAN to identify the discussion bouts in the video-recordings. We then used the bout's start time and end time to extract the data points within each episode, and applied cross wavelet analysis (see below for a detailed explanation) on the motion and speech data within each episode, using a custom R-script. This same R-script enabled us to subsequently link the cross wavelet analysis outcomes to the trials that were saved as OpenSesame output.

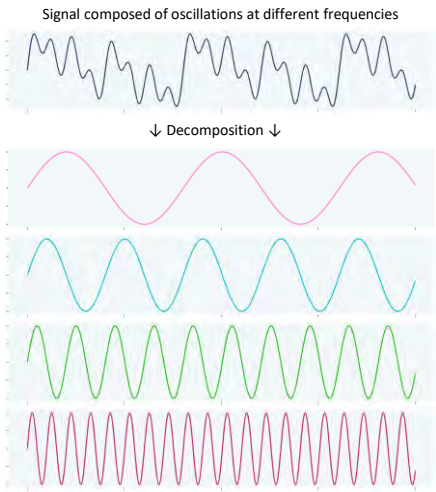
Cross wavelet analysis

Cross wavelet analysis (see Figure 4) detects stable oscillations at different frequencies (i.e. a range of timescales) that co-occur in two signals, as time unfolds (e.g., over the course of the discussion). Figure 4 (panel a) shows this decomposition of oscillations at different frequencies from one, hypothetical, signal. Subsequently, with cross wavelet analysis we calculated when, and at which frequency, the rhythmic waveforms of the two signals *cohere* during a discussion bout. Coherence ranges from 0 to 1, with 1 indicating perfect coupling in time between the signal's oscillations, at certain frequencies and certain moments of a discussion bout. Furthermore, cross wavelet analysis estimates the so-called *relative phase angles* of both signal's cohering oscillations. The relative phase angle indicates the degree of one signal leading the other signal, and whether the signals are *in-phase* or *anti-phase* (see Figure 4, panel b). Cross wavelet plots (see Figure 4, panel c) visualize the coherence and relative phase angles at specific frequencies (y-axis) across time (x-axis), whereby the color corresponds to the degree of coherence (red ≈ 1 ; blue ≈ 0) and the arrows correspond to the relative phase angle. Arrows are only drawn when coherence is significant ($p < .05$). Cross wavelet analysis calculates p -values by repeatedly (in our study: 50 times) simulating surrogate time series, and comparing the empirically found coherence with the coherence derived from the simulations. We only used the significant coherence values and associated relative phase angles for our further analysis.

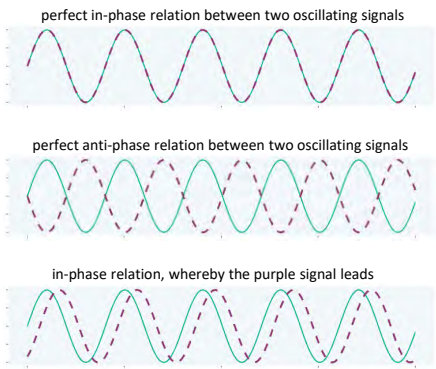
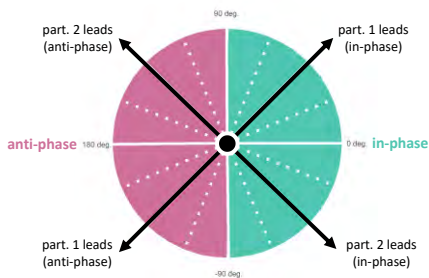
After obtaining the significant coherence values and associated relative phase angles from cross wavelet analysis, we transformed the data in a couple of steps, to further prepare the data for our subsequent analyses. First, we detrended the coherence values across the range of frequencies that we incorporated, by subtracting a positive trend from the coherence values. We detrended the coherence values, because the coherence values are higher for slower timescales¹ (see Figure 5). We calculated this trend using general additive modelling. Previous studies have interpreted such detrended coherence values as a measure for the degree of

¹ This can be explained by the range of periods that we incorporated in the cross wavelet analysis (0.125 - 4 s, or 8 - 0.25 Hz), and the range of durations of our discussion moments (8 - 64.12 s, $M = 22.36$ s). With discussion moments this short, slow oscillations can occur only a couple of times across a discussion moment, while fast oscillations can occur many times across a discussion moment. As a consequence, the coherence between slow oscillations needs to be very high in order to be significant, while this is not the case for fast oscillations (which have more repetitions within a discussion moments). Detrending is a common practice to deal with the issue of relatively high coherence at slow timescales as compared to the faster timescales (e.g. Alviar et al., 2020; Brookshire et al., 2017; Poeppel & Assaneo, 2020).

a. Decomposition of oscillations



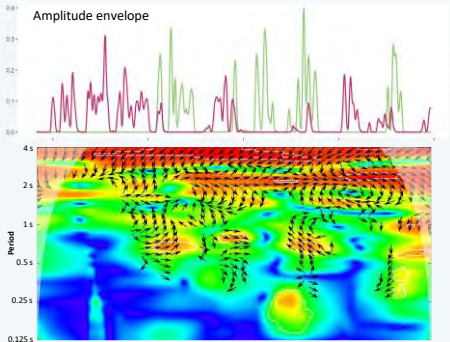
b. Phase relations between oscillating signals



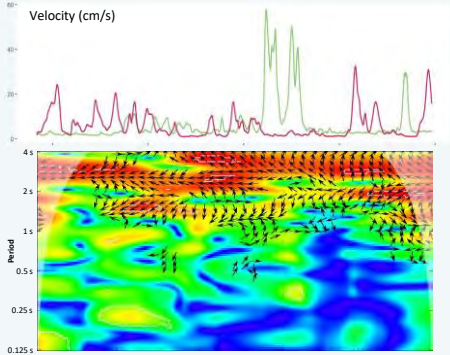
c. Timeseries and cross wavelet plots

These are the timeseries and cross wavelet plots of speech, hand movements, and head movements of 1 dyad during 1 discussion moment (duration = 20 s). The dyad declined 1 correct prediction, agreed about their second prediction, and the pink child adopted the first prediction of the green child for their second prediction.

Speech



Hand movements



Head movements

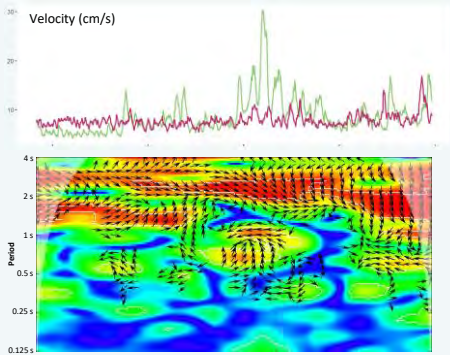


Figure 4. (previous page). Panel a displays how oscillations with a range of frequencies can be detected in a combined waveform. Panel b shows examples of possible phase relations between two oscillating signals. The protractor in panel b shows how arrows in different directions correspond to phase relations between oscillating signals with various relative phase angles (also see Figure 1). Panel c displays the timeseries and cross wavelet plots of speech, hand movements, and head movements of one dyad in our study, across one discussion moment. Within the coherence plot, blue regions indicate that the coherence was low, at a particular timepoint (x-axis) and for oscillations at a particular frequency (y-axis), while red regions indicate that coherence was high. When coherence is significant (i.e. higher than chance level), an arrow is drawn in the plot. The direction of the arrow informs about the phase relations between the oscillations of children's speech, hand movements, or head movements.

coordination at certain timescales (e.g. Alviar et al., 2020; Brookshire et al., 2017; Poeppel & Assaneo, 2020).

Second, we standardized the variance around the detrended mean of each frequency to a standard deviation of 1. Without standardizing, we would not have been able to compare the coherence values across timescales, as the range of coherence values becomes increasingly smaller with lower frequencies (also see Figure 5).

Third, we calculated the coherence values and relative phase angles for evenly spaced frequencies on a linear scale, instead of the default logarithmic scale. We needed to perform this step in order to subsequently calculate the average coherence and circular mean of the relative phase angle for each timescale within each bout of discussion. We took the interval between the longest period and the second longest period ($4 - 3.864 \text{ s} = 0.136 \text{ s}$) as the interval between all of our evenly spaced frequencies (new range: $4 - 0.185 \text{ s}$).

As a last transformation, we transformed the circular mean of the relative phase angles, which spanned 360° , to a linear scale from 0° (in-phase) to 180° (anti-phase). From this transformed circular mean of the relative phase angles, we could infer whether participants' speech, hand movements, and head movements were more in-phase (towards 0°) or anti-phase (towards 180°), and whether one of the participants was leading (towards 90°)². The detrended and standardized average coherence (hereafter: transformed coherence) per modality, timescale, and discussion moment were then used to answer our research questions, using linear mixed-effects regression for dependent variables on an interval scale, and logistic mixed-effects regression for dependent variables on a nominal scale. Similarly, we used the transformed

² Due to this last transformation, we lost the information about who of the two participants was leading. We needed to transform the circular mean of the relative phase angles to a linear scale for our subsequent multilevel linear and logistic regression, however (for an accessible explanation and Bayesian alternative, see Cremers & Klugkist, 2018). Furthermore, it is important to note that, while a relative phase angle of 90° in itself indicates that one signal (e.g. speech) is maximally leading, a random combination of relative phase angles between 0 and 180° will also average around 90° .

Multiscale multimodal coordination during dyadic problem solving

circular mean of the relative phase angles (hereafter: transformed relative phase angle) for carrying out linear and logistic mixed-effects regression.

Mixed-effects regression analysis

We carried out linear and logistic mixed-effects regression analysis (also see Winter, 2013). We performed stepwise mixed-effects regression, whereby we added variables to the mixed-effects regression model, and compared the new model to the previous model for each variable that we added. Following Winter (2013), we obtained p -values of these comparisons between models by likelihood ratio tests. In addition, for each model we calculated Nakagawa's and Schielzeth's (2013) marginal and conditional R^2 for generalized linear mixed-effects models. The marginal R^2 indicates the proportion of total variance which is explained by the fixed effects only. The conditional R^2 indicates the proportion of total variance which is explained by both

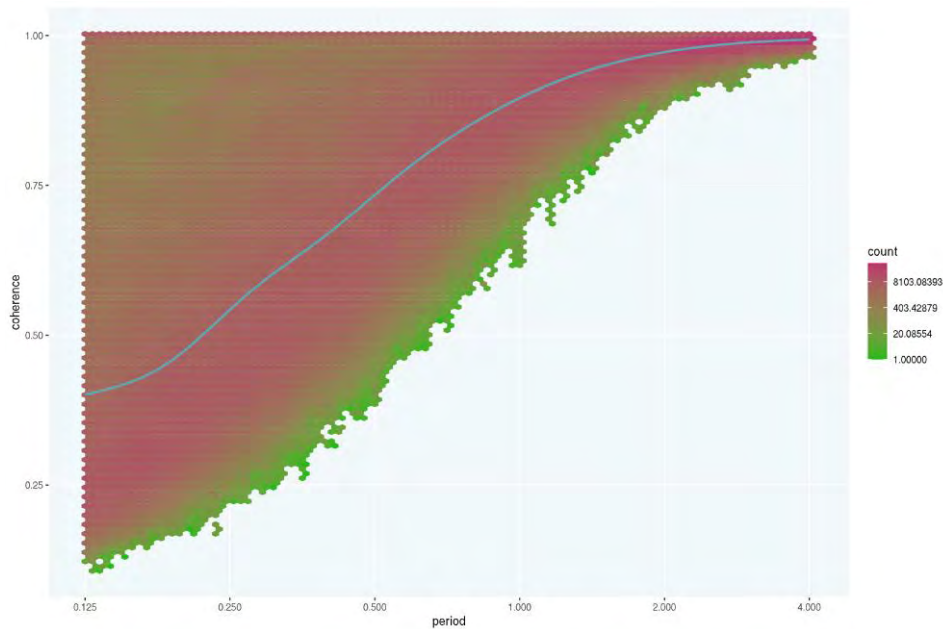


Figure 5. Amount of significant values for each coherence value (y-axis), and each frequency with a specific period in s (x-axis), across all the bouts of discussion, of both speech, hand movements, and head movements of all 25 dyads. Areas which are more intense pink hexagons show that there were more significant values of a particular coherence for a particular frequency, while more intense green hexagons show that there were less significant values of a particular coherence for a particular frequency. If no hexagon is drawn, this means that there were no significant values of a particular coherence for a particular frequency. We therefore can see that the bandwidth of significant coherence values becomes increasingly smaller and closer to 1 for frequencies with a higher period (i.e. closer to a period of 4 s). The blue line is the line we used for detrending, and was derived by applying general additive modeling.

the fixed and random effects. We always included random intercepts for dyads in the mixed-effect regression models, and random slopes when possible. This means that the marginal R^2 indicates the degree to which an effect is similar across dyads, while the difference between the marginal and conditional R^2 indicates the degree to which an effect differs between dyads. For research question 2 and exploratory research question B we standardized the independent variables before adding them to the mixed-effects models.

Results

Descriptives

The 25 dyads each engaged in 4 to 11 discussion bouts ($M = 7.0$, $SD = 1.7$), with a total of 177 bouts. One dyad did not follow the task instructions during the entire experiment, as they talked throughout the whole task and did not make any individual predictions. Another dyad never engaged in dyadic interaction, as one of the dyad members only answered the experiment leader's questions. The 17 discussion bouts of these two dyads ($n = 11 + 6$) were therefore removed from the analyses. We removed another 15 discussion bouts from the analyses due to major interference because the experiment leader needed to reattach the microphone or one of the movement sensors ($n = 4$), the dyad did not follow task instructions and therefore did not make individual predictions ($n = 5$), or the dyad only answered the experiment leader's questions ($n = 6$). This left us with 145 discussion bouts of 23 dyads. Of these 145 bouts, the duration ranged from 4.18 to 64.12 s ($M = 22.36$, $SD = 11.02$). Three of these 145 bouts lasted shorter than 8 s and were therefore also removed from the analyses. In total, the analyses included 142 discussion bouts.

Regarding task performance, for 18 discussion moments the number of correct predictions increased by 2 (13.2%), for 59 discussion moments the number of correct predictions increased by 1 (43.4%), for 39 discussion moments there was no increase or decrease in the number of correct predictions (28.7%), and for 20 discussion moments the number of correct predictions decreased by 1 (14.7%). On average, the number of correct predictions increased by 0.55 ($SD = 0.90$). With regard to agreement about the post-discussion prediction, dyads agreed 85 times (62.5%) and they disagreed 51 times (37.5%). Regarding adopting the other member's pre-discussion prediction, 71 times (52.2%) one member of the dyad adopted the other member's pre-discussion prediction for their post-discussion prediction, while this was not the case for the other 65 post-discussion predictions (47.8%).

Multiscale multimodal coordination during dyadic problem solving

Research question 1: What are the differences in (transformed) coherence between the fast, middle, and slow timescales for the oscillations of children’s speech, hand movements, and head movements across discussion moments?

Both modality, timescale, and the interaction effect between modality (i.e. speech, hand movements, and head movements) and timescale are significantly ($p < .001$) related to transformed coherence (see Table 2, and Figure 6). However, the small values of $R^2_{marginalized}$ and $R^2_{conditional}$ indicate that both the fixed and random effects, respectively, do not explain a large proportion of the variance in transformed coherence. Post hoc analyses (see Table 3) showed that for speech the transformed coherence at the fast timescale was higher than at the middle timescale ($p = .198$), and significantly higher than at the slow timescale ($p < .001$). For both head movements and hand movements, transformed coherence at the middle and slow timescale was significantly higher than transformed coherence at the fast timescale ($p < .001$).

Research question 2: How do the measures of (transformed) coherence relate to dyadic task performance?

As a brief reminder, we analyzed three indices of task performance: a) change in number of correct predictions from before to after the discussion moments, b) agreement of predictions after discussion moments, and c) adopting the other child’s pre-discussion prediction for post-discussion predicting. An overview of the results can be found in Figure 7 and Table A1 (Appendix C). We found that none of the fixed effects significantly ($p > .05$) affected either of the three indices of task performance. The differences between the relatively small values of $R^2_{marginalized}$ and the relatively large values of $R^2_{conditional}$ suggest that the effect of transformed coherence on the three indices of task performance mostly differs between dyads.

Table 2
Results stepwise mixed effects regression on transformed coherence

Dep. variable	Random effects		Fixed effects	R^2_{marg}	R^2_{cond}	Comparison with prev. model		
	Slope	Intercept				df	χ^2	p
transformed coherence	-	dyad + discussion moment within dyad	-	-	.020	-	-	-
			modality	.016	.038	2	18.714	< .001
			modality + timescale	.033	.057	2	21.053	< .001
			modality * timescale	.093	.125	4	77.229	< .001

Note. The model with the interaction effect also contains the individual fixed effects of modality and timescale.

Across dyad transformed coherence (x-axis) for speech, head movements and hand movements at the fast, middle, and slow time scale

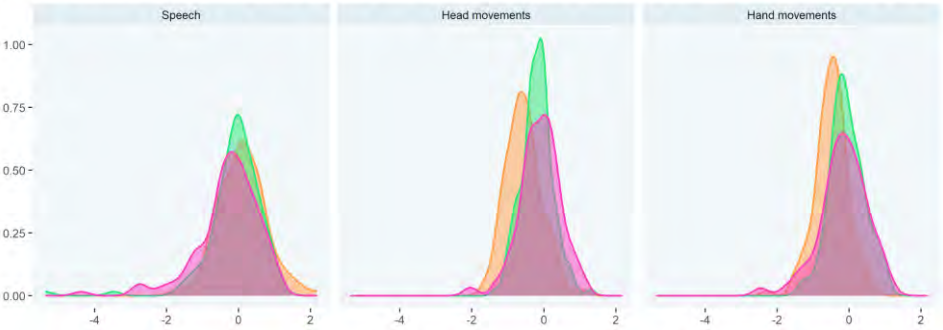


Figure 6. Density plots of the transformed coherence for speech (left panel), head movements (middle panel) and hand movements (right panel), for the fast (0.125 – 0.5s; orange), middle (0.5 – 2s; green), and slow (2 – 4s; pink) timescale.

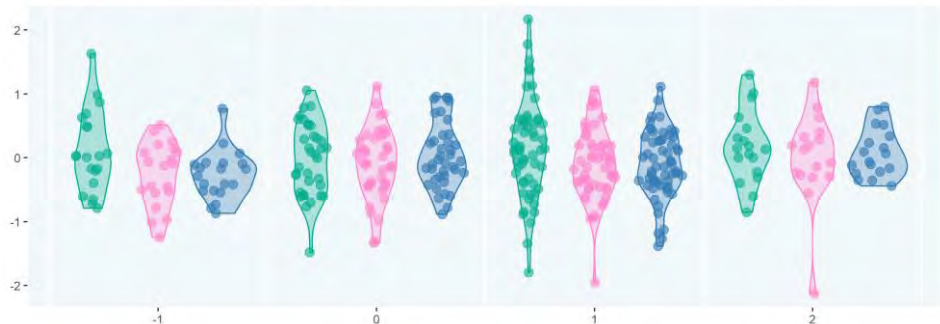
Table 3

Post hoc analyses of differences in transformed coherence between timescales per modality

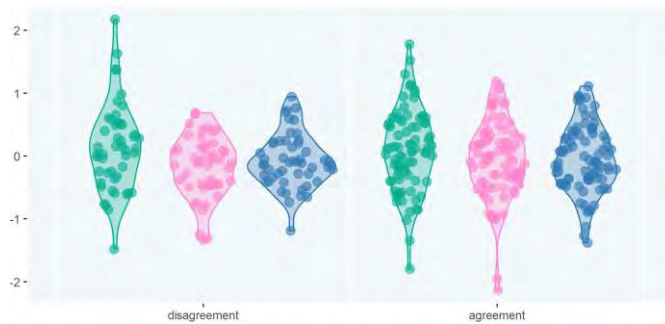
			Modality		
		Statistic	Speech	Head mov.	Hand mov.
Timescale	Fast	<i>M</i>	0.097	-0.563	-0.496
		<i>SE</i>	0.056	0.057	0.056
	Middle	<i>M</i>	-0.098	-0.190	-0.056
		<i>SE</i>	0.056	0.055	0.055
	Slow	<i>M</i>	-0.288	-0.106	-0.160
		<i>SE</i>	0.058	0.056	0.055
Post hoc comparison	Fast vs. middle	<i>df</i>	1045	1044	1041
		<i>t</i>	2.576	4.936	5.838
		<i>p</i>	.198	< .001	< .001
	Fast vs. slow	<i>df</i>	1052	1050	1041
		<i>t</i>	4.976	5.973	5.838
		<i>p</i>	< .001	< .001	< .001
	Middle vs. slow	<i>df</i>	1053	1041	1039
		<i>t</i>	2.442	1.128	1.352
		<i>p</i>	.263	.970	.915

Multiscale multimodal coordination during dyadic problem solving

Across dyad relation between change in number of correct predictions (x-axis), and transformed coherence (y-axis) between dyad's **speech**, **head movements**, and **hand movements**



Across dyad relation between alignment of predictions (x-axis), and transformed coherence (y-axis) between dyad's **speech**, **head movements**, and **hand movements**



Across dyad relation between adopting the other child's pre-discussion prediction (x-axis), and transformed coherence (y-axis) between dyad's **speech**, **head movements**, and **hand movements**

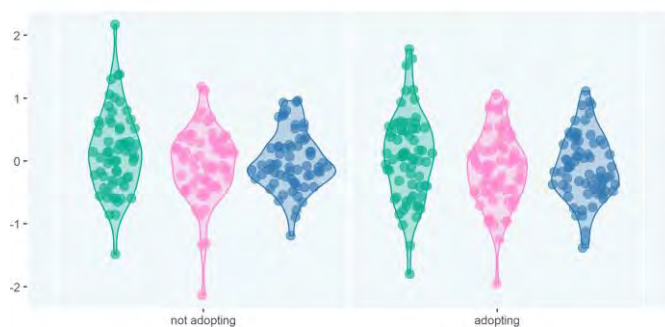


Figure 7. Across dyad relation between transformed coherence of speech (green), head movements (pink), and hand movements (blue) and indices of task performance. The top panel shows the relation between transformed coherence and change in *number of correct predictions*, the middle panel shows the relation between transformed coherence and *agreement of (post-discussion) predictions*, and the bottom panel shows the relation between transformed coherence and *adopting the other child's pre-discussion prediction for post-discussion predicting*.

Exploratory research question A: What are the differences in (transformed) relative phase angle between the fast, middle, and slow timescales for the oscillations of children's speech, hand movements, and head movements across discussion moments?

Both modality, timescale, and the interaction effect between modality and timescale significantly ($p < .01$) affect average transformed relative phase angle (see Table 4, and Figure 8). However, the small values of $R^2_{marginalized}$ and $R^2_{conditional}$ indicate that both the fixed and random effects, respectively, do not explain a large proportion of the variance in average transformed relative phase angle. Post hoc analyses (see Table 5) showed that for speech the average transformed relative phase angle at the middle and fast timescale was significantly different from the average transformed relative phase angle at the slow timescale ($p < .01$).

Table 4

Results stepwise mixed effects regression on transformed relative phase angle

Dep. variable	Random effects		Fixed effects	R^2_{marg}	R^2_{cond}	Comparison with prev. model		
	Slope	Intercept				df	χ^2	p
transformed relative phase angle	-	dyad + discussion moment within dyad	-	-	.002	-	-	-
			modality	.017	.021	2	20.559	< .001
			modality + timescale	.033	.040	2	19.055	< .001
			modality * timescale	.046	.054	4	14.914	.005

Note. The model with the interaction effect also contains the individual fixed effects of modality and timescale.

Across dyad transformed relative phase angle (degrees) for speech, head movements and hand movements at the fast, middle, and slow time scale

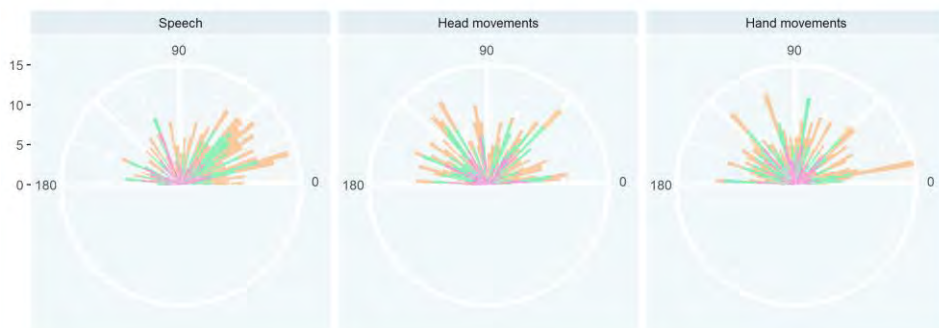


Figure 8. Frequency plots (binwidth = 3°) of the relative phase angle for speech (left panel), head movements (middle panel) and hand movements (right panel), for the fast (0.125 – 0.5 s; orange), middle (0.5 – 2 s; green), and slow (2 – 4 s; pink) timescale.

Multiscale multimodal coordination during dyadic problem solving

There were no significant ($p > .05$) differences in average transformed relative phase angle between timescales for either head movements or hand movements.

Exploratory research question B: How do the measures of average (transformed) relative phase angle affect dyadic task performance?

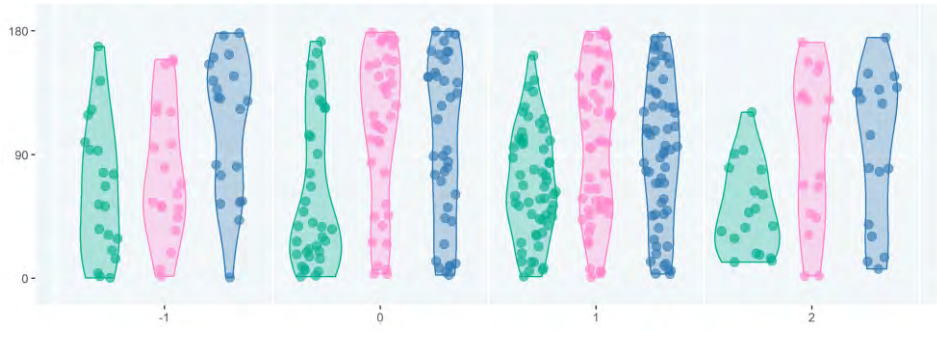
An overview of the results can be found in Figure 9 (next page) and Table A2 (Appendix C). Only average transformed relative phase angle of speech (at the fast timescale) as a fixed effect significantly affected c) adopting the other child's pre-discussion prediction for post-discussion predicting ($p = .007$), whereby the model explained a relatively large proportion of the variance ($R^2_{\text{marginalized}} = .138$; $R^2_{\text{conditional}} = .389$). In addition, the model with both the transformed relative phase angles of speech, hand movements, and head movements as fixed effects and their interaction effect explained a particularly large proportion of the variance of adopting the other child's prediction ($R^2_{\text{marginalized}} = .329$; $R^2_{\text{conditional}} = .559$), albeit this model was not significantly ($p = .080$) better than the model with the transformed relative phase angles of speech and hand movements as fixed effects only ($R^2_{\text{marginalized}} = .171$; $R^2_{\text{conditional}} = .420$). As can be seen in the bottom panel of Figure 7, not adopting the other child's prediction generally goes together with an in-phase relation between children's speech, and often with an anti-phase relation between children's head movements and hand movements. None of the other fixed effects significantly

Table 5

Post hoc analyses of differences in transformed relative phase angle between timescales per modality

			Modality		
		Statistic	Speech	Head mov.	Hand mov.
Timescale	Fast	<i>M</i>	63.1°	82.8°	82.9°
		<i>SE</i>	4.58°	4.71°	4.60°
	Middle	<i>M</i>	68.5°	89.4°	96.4°
		<i>SE</i>	4.64°	4.50°	4.52°
	Slow	<i>M</i>	94.0°	97.3°	88.1°
		<i>SE</i>	4.83°	4.62°	4.57°
Post hoc comparison	Fast vs. middle	<i>df</i>	1047	1046	1042
		<i>t</i>	0.845	1.016	2.110
		<i>p</i>	.995	.984	.467
	Fast vs. slow	<i>df</i>	1056	1052	1044
		<i>t</i>	4.702	2.223	0.809
		<i>p</i>	< .001	.391	.997
	Middle vs. slow	<i>df</i>	1057	1042	1040
		<i>t</i>	3.853	1.249	1.302
		<i>p</i>	.004	.945	.931

Across dyad relation between change in number of correct predictions (x-axis), and transformed relative phase angle (y-axis) between dyad's **speech**, **head movements**, and **hand movements**



Across dyad relation between alignment of predictions (x-axis), and transformed relative phase angle (y-axis) between dyad's **speech**, **head movements**, and **hand movements**



Across dyad relation between adopting the other child's pre-discussion prediction (x-axis), and transformed relative phase angle (y-axis) between dyad's **speech**, **head movements**, and **hand movements**



Figure 9. Across dyad relation between transformed relative phase angle of speech (green), head movements (pink), and hand movements (blue) and indices of task performance. Values between 0° and 90° indicate in-phase and values between 90° and 180° indicate anti-phase. The top panel shows the relation between transformed relative phase angle and change in number of correct predictions, the middle panel shows the relation between transformed relative phase angle and agreement of (post-discussion) predictions, and the bottom panel shows the relation between transformed relative phase angle and adopting the other child's pre-discussion prediction for post-discussion predicting.

Multiscale multimodal coordination during dyadic problem solving

($p > .05$) affected either a) change in number of correct predictions from before to after the discussion moments, b) agreement of predictions after discussion moments, and c) adopting the other child's pre-discussion prediction for post-discussion predicting. Similar to coherence, the differences between the relatively small values of $R^2_{marginalized}$ vs. the relatively large values of $R^2_{conditional}$ suggest that the effect of average transformed relative phase angle on the three indices of task performance mostly differs between dyads.

Discussion

The first goal of this study was to investigate how children (age: 6-10 years) coordinate their speech, hand movements, and head movements at multiple timescales when they solve balance scale problems together. We applied cross wavelet analysis coherence, which enabled us to study the *similarity of two systems' oscillations* at multiple and *distinct* timescales across a particular episode. Relative phase angle characterizes the direction of the relation between these oscillations. We found that both modality and timescale significantly affect transformed coherence (see Method section for details of the transformation). More specifically, the interacting children coordinated their speech mostly at the fast timescale (8 - 2 Hz), whereby only the difference between the slow timescale (0.5 - 0.25 Hz) and fast timescale was statistically reliable. Furthermore, our results showed that children coordinated both their head movements and hand movements reliably more at the middle (2 - 0.5 Hz) and slow timescales than at the fast timescale.

In addition, the results of our exploratory analyses showed that modality and timescale also significantly affect transformed relative phase angle. Specifically, for speech, the transformed relative phase angle at the fast and middle timescale was reliably different than the transformed relative phase angle at the slow timescale. Notably, speech at the fast and middle timescale often showed in-phase coordination (i.e. between 0° and 90°), while the transformed relative phase angle at the slow timescale was evenly distributed between 0° and 180° , so spanning across both in-phase and anti-phase coordination. For head movements and hand movements we did not find reliable differences in transformed relative phase angle. Similar to the coordination of speech at the slow timescale, the transformed relative phase angle of head movements and hand movements was not centered in either the in-phase coordination or anti-phase coordination regimes.

This study's second goal was to research how these measures of interpersonal coordination would predict children's task performance. With regard to transformed coherence, we found no statistically reliable effects of either speech (fast timescale), head movements (slow timescale), and hand movements (middle timescale) on the change in number of correct predictions from before to after the discussion moments, the agreement of predictions after

discussion moments, nor the tendency to adopt the other child's pre-discussion prediction for post-discussion predicting.

Our exploratory analyses yielded a statistically reliable effect of transformed relative phase angle of speech (fast timescale) on adopting the other child's pre-discussion prediction for post-discussion predicting, whereby not adopting the other child's prediction went together with an in-phase relation between children's speech. It is important to note that such results should be treated carefully given issues of multiple comparisons. This in-phase relation between children's speech might indicate that both children talked at the same time, which is different from a pattern of turn taking. However, turn taking on a timescale of 8 -2 Hz is very fast, which makes this explanation for the in-phase relation between children's speech unlikely. There was no reliable effect of transformed relative phase angle of speech on either of the other indices of task performance, nor did we find significant effects of transformed relative phase angle of head movements (slow timescale) and hand movements (middle timescale) on either of the indices of task performance.

Multimodality at different timescales for interacting children

Interpersonal coordination in terms of (transformed) coherence

Our findings with regard to multimodality at different timescales for interacting children are not identical to the findings for adults, albeit there is some notable overlap. With regard to transformed coherence, we found mostly, but not (always) reliably more, interpersonal coordination of speech at the fast timescale. Previous studies have overwhelmingly shown that adults mostly interpersonally coordinate their speech at the fast timescale (for a recent review, see Poeppel & Assaneo, 2020). Furthermore, while we found significantly more interpersonal coordination of hand movements at both the middle and slow timescale, previous studies in adults suggest that hand movements are usually interpersonally coordinated most at the middle timescale only (Pouw & Dixon, 2019b; Xiong & Quek, 2006). In addition, our findings of significantly higher transformed coherence at the middle and slow timescale (2 – 0.125 Hz) for head movements are in line with Hale et al. (2019), who found significant coherence between head movements at a timescale from 1.1 to 0.2 Hz, that is, the timescale of mimicry. Lastly, Hale et al. (2019) found lower than chance coherence (i.e. decoupling) between high frequency (2.6 – 6.5 Hz) head movements, which they attributed to fast nodding by the listener. Although differences in calculation prohibit a direct comparison of Hale et al.'s (2019) coherence measure with the current study's transformed coherence values, it should be noted that transformed coherence of head movements at the fast timescale was the lowest of all transformed coherence values in our study. This could mean that a similar decoupling of high frequency head movements is evident in children as well. In summary, while our study shows

Multiscale multimodal coordination during dyadic problem solving

similarities with previous findings in adults, our findings about multimodality at *distinct* timescales in children seem to be less pronounced than in adults.

One possible explanation for the differences between children and adults in multimodality at different timescales pertains to developmental differences. The literature about the development of interpersonal coordination of speech, hand movements, and head movements in conversations across middle childhood is scarce, however. One observational case study about a 10-year-old child showed no significant linguistic alignment between the child and various interaction partners (Martin et al., 2010), in contrast to findings about linguistic alignment in adults (e.g. Coupland, 1980; Fusaroli et al., 2012). That said, there are some studies about the development of interpersonal coordination across middle childhood in general (see also Rauchbauer, 2020, for a recent review about motor mimicry and synchrony across childhood). One recent study showed that children's alignment of hand movements during a dyadic joystick task increases between 6 and 9 years, but is still far below adult's alignment (Satta, 2017). Furthermore, Xavier et al. (2018) found that interpersonal synchronization of children and adolescents, between 6 and 19 years, and a virtual character increases with age. These studies suggest that general interpersonal coordination in children at the age of our sample (6-10 years) is still developing and is therefore still quite different from interpersonal coordination in adults. This may manifest itself during dyadic conversations as less pronounced coordination of speech, hand movements, and head movements at distinct timescales.

Another possible explanation for the differences in multimodality at different timescales between our study and previous studies stems from our study set-up. First, we collected data in a science center. As a consequence, the data is inherently more noisy, and the interaction between children less stable, than in a controlled lab setting. It is possible that the coordination of speech, hand movements, and head movements as is found in controlled lab settings is more pronounced than in naturalistic settings. In addition, one might speculate that the interpersonal coordination between children is less mature than in adults, which might amplify this effect of naturalistic vs. controlled setting even more. Second, the duration of discussion moments in our study are shorter than the episodes of conversation which are typically investigated in other studies (e.g. Pouw & Dixon, 2019b; Romero et al., 2018; Wiltshire et al., 2019). During short discussion moments, coordination of speech, hand movements, and head movements might manifest itself differently than in long discussion moments. For example, some coordination patterns might take longer to establish than the average duration of 22 s of bouts of discussion in this study allowed for. Previous studies have reported on coordination patterns on such slow timescales, such as a 20 to 30 s lag between the vocalizations of 4-month-olds and their mothers (Jaffe et al., 2001), or an average lag of 18 s between Kindergartner's gestures and speech (De Jonge-Hoekstra et al., 2016).

Phase relations

Besides interpersonal coordination in terms of transformed coherence, we also exploratively investigated the transformed phase relations of the children's interpersonal coordination of the three modalities. To our knowledge, phase relations between one of the three modalities during conversations have only been investigated by Hale et al. (2019). They found 30° to 90° relative phase angles, that is, in-phase relation, between leaders' and followers' head movements at a timescale range from 0.2 to 0.5 Hz. Aside from the three modalities, relative phase angles have been investigated between people's whole body movements during conversations. Schmidt et al. (2014) found above chance relative phase angles from 0° to 40°, again an in-phase relation, between joke tellers' and joke listeners' body movements, while Romero et al. (2018) found no clear phase relation between the body movements of autistic children and clinicians.

Similar to Romero et al. (2018), we did not find clear phase relations between children's hand movements and head movements at either of the timescales. Maybe there were no clear phase relations between movements in both Romero et al.'s (2018) and our study, because the participants did not engage in a task with a clear and assigned role structure, in contrast to the studies by Hale et al. (2019; leader and follower) and Schmidt et al. (2014; joke teller and joke listener). However, Fujiwara et al. (2019a) researched unstructured conversations, and using a more strict procedure to define in- or anti-phase (i.e. in-phase = 0° - 20°; anti-phase = 160° - 180°), they found above chance in-phase coordination and below chance anti-phase coordination for women. Furthermore, the explanation of assigned role structure still leaves open the question of why we did find an in-phase relation between children's speech at the fast and middle timescale. More studies are needed to investigate the similarities and differences between interpersonal coordination of speech and body movements, and of children and adults.

Effect of multimodal interpersonal coordination measures on task performance

Except for the effect of transformed relative phase angle of speech on adopting the other child's prediction, we did not find an effect of either transformed coherence nor transformed relative phase angle between the three modalities on our three indices of task performance. Our null-findings are unlike findings in previous studies, which used a variety of measures of interpersonal coordination. For example, a recent cross wavelet analysis study showed that both coherence and in-phase coordination between dyads' whole body movements at the 1 Hz timescale predicted their performance on a complex collaborative problem solving task (Wiltshire et al., 2019). With regard to other measures of interpersonal coordination, both

Multiscale multimodal coordination during dyadic problem solving

complexity matching and Cross Recurrence Quantification Analysis (CRQA) have been related to task performance, which will be addressed in the next paragraphs.

Complexity matching characterizes the *similarity in (multi)fractal structure* of two timeseries' (e.g. participant's speech) variability at multiple timescales across a particular episode (West et al., 2008; also see Abney, Paxton et al., 2014; de Jonge-Hoekstra et al., 2021). Fusaroli et al. (2013) found that degree of complexity matching between participant's speech timing correlated positively with the degree to which the dyad benefited from working together during a joint decision task. Furthermore, a recent study showed that degree of complexity matching between participant's speech timing during a construction task with marshmallows and uncooked spaghetti was positively correlated to the height of the tower (Abney et al., 2021). They found no correlation between degree of complexity matching between participant's whole body movement timing and task performance, however.

CRQA measures characterize the *coupling between two systems* at multiple timescales across a particular episode (e.g. Marwan et al., 2007; Shockley et al., 2002; Webber Jr. & Zbilut, 2005). Subjecting the data of the whole body movements during the construction task (Abney et al., 2021) to CRQA, Abney et al. (2015) found that coupling strength negatively predicted tower height. Furthermore, in a study in which dyads of children (8 – 13 years) solved tangram puzzles together, Vink et al. (2017) used CRQA measures to characterize the coupling between dyads' postural sway, in terms of degree of stability, synchronization, and complexity. They found a negative correlation between the CRQA measures and dyadic task performance.

When comparing the findings of the current study and the above mentioned studies, it is important to note the differences between the studies. Besides differences in analysis methods (i.e. cross wavelet analysis, complexity matching, CRQA), there were differences in tasks (e.g. discussion only, Fusaroli et al., 2013, vs. playing a video game together, Wiltshire et al., 2019), and movement and/or speech measurements (e.g. spike trains of speech and movements, Abney et al., 2021, vs. postural sway, Vink et al., 2017). Moreover, only children were included in the current study and in the study by Vink et al. (2017). These differences make it difficult to compare these studies' findings. Future, more systematic, research is needed to further understand the relations between task characteristics, which patterns are captured at which level of detail, specific measures of interpersonal coordination, age, and task performance.

Learning about balance scale problems

With regard to task characteristics, in the current study children solved balance scale problems together by means of discussion. Balance scale problems have been widely used in developmental psychology (e.g. Boom et al., 2001; Jansen & van der Maas, 2002; Pine et al., 2004, 2007; Siegler, 1976; van Rijn et al., 2003) because they enable researchers to

systematically study learning within a short period of time. A limited number of studies have used balance scale problems to investigate dyadic learning, using either paper-and-pencil task to predict the outcome of the problems (Denessen et al., 2008; Krol et al., 2004), or a real balance scale and weights (Tudge, 1991). Instead of focusing on dyadic learning, in the current study we used balance scale problems primarily as a vehicle to elicit discussion and collaborative problem solving in children. This does not make the question about dyadic learning within our study any less interesting or intriguing, however.

In our study, dyads of children often improved their number of correct predictions by either 1 (43.4%) or 2 (13.2%), and they agreed about the second prediction 62.5% of the time (see Descriptives in Results-section). This might suggest that some kind of learning and mutual influence was going on between the dyads. To better understand whether and how children learn from each other during the experiment, we need to investigate the bouts of discussion in a qualitative and *detailed* manner in a follow-up study (cf. Van der Steen et al., 2012; Van Der Steen et al., 2014), thereby capturing what the dyads actually discussed. In addition, we need to investigate the development of children's predictions and discussion over time, across the task. Lastly, children not only interacted during discussion moments, but also in between. Researching what happens in between discussion moments might also shed light on dyadic learning within our study.

From our anecdotal observations during data collection, we know there were large differences between dyads in terms of quality and content of discussion moments. These large between-dyad differences might be captured by the relatively large $R^2_{conditional}$ (compared to the small $R^2_{marginalized}$ of the fixed effects) of the models which predicted our three indices of task performance. Among the many possible explanations of these large differences between dyads, previous studies have shown that the composition of the dyad in terms of individual ability influences dyadic learning about balance scale problems (Denessen et al., 2008; Krol et al., 2004; Tudge, 1991). In general, previous studies have shown that when a higher performing peer teams up with a lower performing peer, the performance of the lower performing peer often increases (e.g. Bahrami et al., 2010; Young et al., 2019). These findings fit within the concept of *scaffolding*. Scaffolding refers to providing support to someone, so that they are able to do a task that they would not have been able to do without support (van de Pol et al., 2010; van Geert & Steenbeek, 2005; Wood et al., 1976). Follow-up studies need to investigate whether dyadic learning in the current study is related to composition of the dyad, and the related possible occurrence of scaffolding, by creating dyads in a specific way focusing on initial understanding level of the children.

Multiscale multimodal coordination during dyadic problem solving

Limitations

Despite thorough preparation, our data collection procedure did not run flawless, with some technical and procedural issues leading to the removal of a number of the discussion moments from the analyses. A large part of these metaphorical bumps are due to collecting data at a science center, which is inherently noisy. Also, an additional number of discussion moments needed to be removed because children did not follow the task instructions. Finally, in the current study we have treated the coherence coordination and relative phase angle as relatively abstract measures of interpersonal coordination. We did not go into qualitative detail of how these different measures of coherence coordination and relative phase angle manifest itself as actual behavior in the discussion moments.

Conclusion

We found that children coordinate their speech, hand movements, and head movements together, and at different timescales, whereby children's speech often shows an in-phase relation. Furthermore, if neither of the children adopts the other child's prediction, this often goes together with an in-phase relation between children's speech. We found no further relations between transformed coherence or phase relation and other indices of task performance, however. Our study is the first to provide important understanding about school-aged children's multimodal interpersonal coordination, and its relation with task performance.

Acknowledgements

We would like to thank the Connecticut Science Center for the opportunity to participate in the Living Lab. Furthermore, we would like to thank the Center for Information Technology of the University of Groningen for their support and for providing access to the Peregrine high performance computing cluster. This work was supported by the Prins Bernhard Cultuurfonds (ref. no: 40021263).



6

General Discussion

General Discussion

In this dissertation, my aim was to understand how cognitive development is related to how children move their hands and how they speak during cognitive tasks, over time and at multiple (time) scales. I therefore examined how children's hand movements, gestures, and speech relate to each other and to the physical and social environment, when children engage in science and technology tasks. Previous studies have shown that children's hand movements lead cognitive development, over speech. For example, children use their hands to explore (e.g. Adolph & Franchak, 2017), and they show understanding in gestures before they can put this into words (e.g. Cook & Goldin-Meadow, 2006; Goldin-Meadow, 2001). My research questions and hypotheses, as well as the methodologies and data-analyses, were based on the theoretical perspectives of *complex dynamical systems*, *coordination dynamics*, and *affordances*, because I expected these perspectives to bring unique insights about hand movements' leading role in cognitive development. A detailed explanation of the theoretical perspectives can be found in the General Introduction. In this dissertation's studies I investigated *hand movements*, *gestures*, or a *combination* of the two¹. Depending on whether a study emphasized hand movements or gestures, I will choose one of the terms in the study descriptions below.

Summary of studies

Study 1 - Asymmetric dynamic attunement of speech and gestures in the construction of children's understanding

In the first study (Chapter 2), I investigated whether the leading role of hand movements and gestures (hereafter: gestures) in children's cognitive development is also evident *within* tasks, as opposed to *between* tasks, which previous studies have shown. Using a multiple case study design ($N = 12$), I researched how children performed a science and technology task together with an adult. I coded children's gestures and speech, and assigned levels of understanding to both their gestures and speech. Furthermore, Chromatic and Anisotropic Cross Recurrence Quantification Analysis (Cross RQA) was used to investigate the coupling between levels of understanding of gestures and speech, in terms of which modality is leading in time, and which modality attracts the other more strongly (see Figure 1). In addition, I examined whether these within-task-measures of coupling were related to more static measures of cognitive development.

¹ Gestures are hand movements which are typically tightly coordinated with speech, while hand movements are a broader category and are not necessarily coordinated with speech. The boundary for when a hand movements becomes a gesture is fuzzy, however. I will also touch upon this later in the General Discussion.

General Discussion

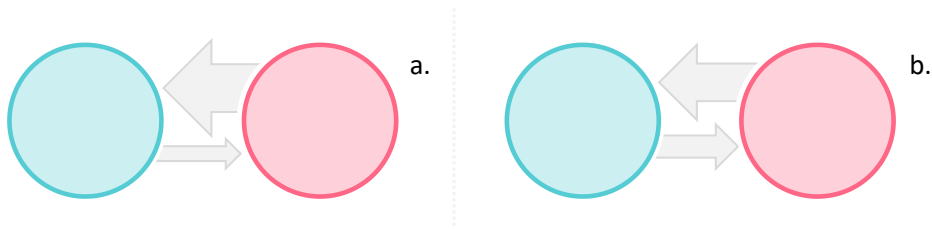


Figure 1. Schematic visualization of coupling between levels of understanding of gestures (blue) and speech (pink). Panel a shows that understanding in speech attracts understanding in gestures much more than vice versa, as displayed by the large difference in arrow size. This pattern was typical of children with a high score on a standardized language test. Panel b also shows that understanding in speech attracts understanding in gestures more than vice versa, but it is more balanced than the coupling relation between gestures and speech in panel a. This last pattern was typical for children with a high score on math or past science and technology tasks.

The results showed that Kindergartners' ($n = 5$) understanding in gestures was leading understanding in speech by 18 s, on average. For first-graders ($n = 7$) understanding in gestures and speech was more synchronized, whereby speech was slightly leading ($M = 0.71$ s). I thus only found a leading role of gestures in time for younger children. Furthermore, for all children speech attracted gestures more than vice versa. Interestingly, this asymmetry in speech attracting gestures was more pronounced for children who scored higher on a standardized language test². For children who scored higher on a standardized math test², or on past science and technology tasks, the asymmetry was less pronounced, and speech and gestures were thus more balanced. This last finding could be taken to suggest that, for these children, speech constrains gestures relatively less, and that children benefit from this during math or science and technology tasks.

Besides the empirical findings, in this first study I proposed a theoretical perspective on the coupling between gestures and speech. I proposed that gestures and speech were two coupled, yet separate synergies. In other words, both gesturing and speaking involves the functional organization of many components (e.g. muscles, bones, neurons) at many scales (e.g. cells, muscles, brain) throughout the neuromuscular system. Critically, there is overlap in the components which are involved in the simultaneously occurring synergies of gestures and speech (Figure 2a). For instance, our lungs, muscles in our back, and the neural structure of Broca's area are involved in both gesturing and speaking. Based on this proposal, I suggested that gesture-speech mismatches are due to a competition and subsequent decoupling between the synergies of gestures and speech during difficult tasks (Figure 2b). This idea was empirically challenged in the second study.

² We used the Kindergarten CITO as our standardized language and math test. This test is not used anymore in primary schools in the Netherlands, because its predictive value of later language and math scores is limited.

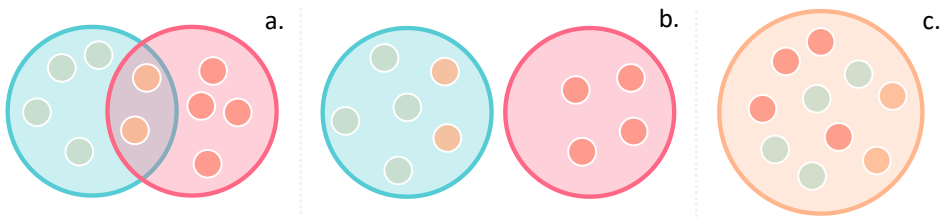


Figure 2. Schematic visualization of synergies of gestures and speech. The small dots represent components at many scales. The large circles represent synergies. Panel a displays the coupled synergies of gestures (blue circle) and speech (pink circle), whereby there is an overlap between the two synergies with respect to the involved components. Panel b displays the decoupled synergies of gestures and speech during gesture-speech mismatches, due to competition between the two synergies during difficult tasks (as proposed in Study 1/Chapter 2). Panel c displays how the components involved in the synergies of gestures and speech have organized into one gesture-speech-synergy (orange circle), due to difficult tasks (as proposed in Study 2/Chapter 3).

Study 2 - Easier said than done? Task difficulty's influence on temporal alignment, semantic similarity, and complexity matching between gestures and speech

In the second study (Chapter 3), I investigated whether task difficulty influences gesture-speech synchronization. Different from the other three studies in this dissertation, I investigated bachelor students instead of children. We chose to investigate bachelor students instead of children, because bachelor students would be able to persevere for 1100 trials, while this would be nearly impossible for children.

In the experiment, task difficulty was manipulated by means of the task layout (also see Figure 4, Chapter 3, p. 74). In the easy condition participants had to match targets of the same color by pointing to locations and saying the word of the locations in a *regular* pattern. In the difficult condition participants also had to match targets of the same color, but they had to point to locations and say the word of the locations, which were presented in a *random* pattern. I analyzed three forms of gesture-speech synchronization: Temporal alignment, semantic similarity, and complexity matching. Temporal alignment pertained to the time between pointing to a location and saying the word of a location. Semantic similarity referred to whether the location to which a participant pointed matched with the location that the participant said in words. Important to note is that gesture-speech mismatches can be seen as semantic *dissimilarity* between gestures and speech. Complexity matching can be thought of as a similarity in the multiscale (i.e. including a whole range of shorter and longer timescales) organization of gestures and speech. I used Multi Fractal Detrended Fluctuation Analysis (MFDFA), which informs about systems' multiscale organization, to investigate complexity matching between gestures and speech.

General Discussion

I found that task difficulty indeed influences students' gesture-speech synchronization. I found less temporal alignment, less semantic similarity (thus more gesture-speech mismatches), and more complexity matching between gestures and speech in the difficult than in the easy condition. This last finding suggests that the synergies of gestures and speech *do not* compete and subsequently decouple during difficult tasks, as we suggested in the first study. Instead, it seems like the components involved in the synergies of gestures and speech organized into a combined gesture-speech-synergy (Figure 2c). Interestingly, this gesture-speech-synergy during difficult tasks seems to be characterized by less temporal alignment and less semantic similarity than the coupled synergies of gestures and speech during easy tasks (Figure 2a).

A second interesting finding from the second study was that participants never *pointed* to the incorrect location, whereas they frequently *said* the incorrect location. This means that all the instances of semantic dissimilarity were due to incorrectly saying the location. I drew a parallel between this finding of gestures always being correct while speech was not, and gestures' leading role in cognitive development during gesture-speech mismatches. To explain both phenomena, I proposed that gestures and hand movements, regardless of age, are strongly coupled to spatiotemporal properties of the environment, while speech is not strongly coupled to these properties. I further investigated this proposal in the third study.

Study 3 - Movers and shakers of cognition: Hand movements, speech, task properties, and variability

Previous studies have shown that a transition from "old" to "new" understanding is a reorganization of the system which goes together with an increase in variability³ (Stephen, Boncoddio, et al., 2009; Stephen, Dixon, et al., 2009). As a brief reminder, in the second study I proposed that hand movements' leading role in cognitive development is related to hand movements' typically stronger coupling with spatiotemporal task properties. If this is true, one would expect a difference in these participants' hand movements' variability upon a change in relevant spatiotemporal task properties (i.e. length of the balance scale, or relative difference in mass between the two weights), but not in speech. In the third study (Chapter 4), I researched whether a change in the spatiotemporal task properties of a balance scale task differentially affected the variability of speech and hand movements (including gestures; hereafter: hand movements) during children's explanations. Children between 4 and 7 years were investigated, which is the age range in which children typically transition from only regarding the mass of the

³ I explained this increase in variability during transitions and reorganization using a LEGO-metaphor (Chapter 4, p. 99): One can only build a new structure from an old structure if one breaks the old structure (increase the variability) and uses the bricks to build a new structure.

weights to solve balance scale problems to including the distance of the weights from the fulcrum in their explanation (see Figure 3).

Children were asked to predict and explain about balance scale problems to an experiment leader. They hereby participated in one of two experiments (also see Figure 1, Chapter 4, p. 100). In the first experiment, children worked with a long balance scale in the first half of the task, and with a short balance scale in the second half (Long-Short condition), or vice versa (Short-Long condition). The difference in length of the balance scale was related to the relevant dimension of distance from the fulcrum for solving balance scale problems. In the second experiment, children worked with a large difference in mass of the weights in the first half of the task, and with a small difference in mass in the second half (Large-Small condition), or vice versa (Small-Large condition). The difference in mass of the weights was related to the relevant dimension of weight. Children's hand movements and speech were separately coded, in broad behavioral categories, and with a frequency of 2 Hz (2 data points per second). On these time series of hand movements and speech, I calculated two measures of variability. The first measure was *Diversity*, which was calculated by means of the Shannon entropy of the frequency distribution of the behavioral categories and their duration. Diversity indicates a system's adaptability (e.g. Adolph et al., 2015), and is related to the range of behaviors across the task. The second measure was *Complexity*, which was derived by applying categorical RQA, and calculating the block entropy (Leonardi, 2018). Complexity indicates a system's flexibility, and is related to the temporal structure of behavior across the task (Adolph et al., 2015).

Across the two experiments, and thereby four conditions, I found a difference in Complexity after a change in relevant spatiotemporal task properties, but not in Diversity, for *both* hand movements and speech in all but the Long-Short condition. Furthermore, only in the Large-Small condition the change in Complexity (and Diversity) of hand movements and speech was related. In other words, changes in relevant spatiotemporal task properties thus often affected the flexibility (but not adaptability) of both hand movements and speech. The differences in flexibility of hand movements and speech *within* children were mostly unrelated, however.

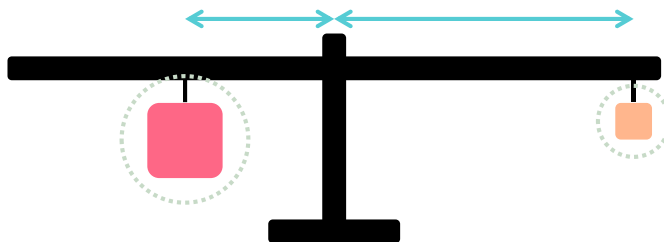


Figure 3. Example of a balance scale task. The dotted lines are drawn around the weights. The arrows indicate the distance of the weights from the fulcrum.

General Discussion

These findings suggest that changing the relevant spatiotemporal properties of a task thus influences the flexibility of children's hand movements and speech, but it is yet unclear whether and how such changes benefit cognitive change in children. Furthermore, I suggested that changes in spatiotemporal task properties might not even have to be related to the *relevant* dimensions of the task in order to elicit an increase in variability. In line with this, Stephen, Dixon, et al. (2009) found that *random* changes in spatiotemporal task properties increased the variability of participants' hand movements, and thereby raised the likelihood of discovering a new cognitive strategy. In addition, children not only needed to attune their hand movements (and speech) to spatiotemporal task properties, but also to the experiment leader to which they explained their answers. These social interactions shape children's hand movements and speech as well, which might explain why I found changes in both hand movements' and speech's flexibility. I further addressed how children attune their multimodal behavior to each other during social interactions in the fourth and last study.

Study 4 - Get it together: How do children interact multimodally and at multiple time scales when they solve dyadic balance scale problems?

In the final study, I investigated interpersonal coordination between children's speech, hand movements, and head movements, when children solve balance scale problems in pairs. I collected data of 25 dyads ($N_{children} = 50$), between 6 and 10 years of age, in the Connecticut Science Center. For each balance scale problem, children were asked to first individually discuss about the outcome of the balance scale problem, by pressing a button on a game controller. If they disagreed or if they both gave the wrong prediction, the children needed to discuss what they individually predicted and why. After these discussion moments, children were asked to give a second, individual prediction about the particular balance scale problem. I measured children's hand movements and head movements during discussion moments, and recorded their speech. To investigate interpersonal coordination, I applied cross wavelet analysis on the time series of intensity of speech, hand movements, and head movements during discussion moments. Cross wavelet analysis allowed me to look for similarities in the periodicities (i.e. frequency of the waves [~] which make up communication) of each modality between children in a dyad over time (see Figure 1, Chapter 5, p. 130, and Figure 4, Chapter 5, p. 138, for a more elaborate explanation), and thereby investigate which multimodal coordination patterns occurred at which timescales when children solve dyadic balance scale problems. Furthermore, I examined whether these measures of interpersonal coordination predict dyadic task performance after discussing.

I found that children together coordinated their speech mostly at a (fast) timescale from 2-8 Hz (i.e. one ~ per 0.5 to 0.125 s), while they coordinated their hand movements and head movements mostly at a (slow) timescale from 0.25-2 Hz (i.e. one ~ per 4 to 0.5 s). These findings

are comparable to results about interpersonal coordination between adults, albeit less pronounced. Furthermore, I found that children's speech at a timescale from 0.5-8 Hz (i.e. one \sim per 2 to 0.125 s) is mostly in an in-phase (i.e. 0-90°) relation, which means that one of the children typically followed the other child's talking, with a delay between 0 and 1 s (depending on the timescale as well as the exact phase relation). There were no clear phase relations between the other modalities at either of the timescales, however. Lastly, except for an anti-phase relation between children's speech being predictive of *not* adopting the other child's prediction, there were no relations between the measures of interpersonal coordination and indices of dyadic task performance. In line with our findings, previous studies found interpersonal coordination between people who do something together, but findings are mixed about whether there is a relation between measures of interpersonal coordination and task performance. I therefore suggested that more *systematic*⁴ and *qualitative* research is needed to learn how interpersonal coordination contributes to how well people work together in general, and how children learn from each other during cognitive tasks in particular.

In the following sections I will discuss the findings summarized above from several angles. Firstly, I will deal with how to understand the role of children's hand movements and speech in cognitive development from the perspective of complex dynamical systems, coordination dynamics, and affordances, and what the studies, methods, and analyses, in this dissertation contribute to this understanding. Secondly, based on this as well as on the studies' limitations, I will lay out important directions for future and follow-up research. Finally, I will discuss which practical implications, for instance for education, follow from this dissertation.

Coupling between hand movements and speech during science and technology tasks

The findings of Study 1 and 3 (Chapter 2 and 4) point into one direction: When children are asked by someone else to explain about science and technology tasks, their hand movements and/or gestures and speech are clearly coordinated and coupled. This finding resonates with many previous studies (e.g. Alibali et al., 2014; Church & Goldin-Meadow, 1986; Goldin-Meadow, 2001; Goldin-Meadow, 2017; Kita, 2000). With regard to the specifics of this coupling, in Study 1 (Chapter 2) I found that understanding in speech attracts understanding in gestures more than vice versa (i.e. *asymmetric* coupling). This result possibly reflects that speech often

⁴ At the moment, different studies have applied diverse measures of interpersonal coordination to a diverse range of interpersonal tasks. This makes it difficult to compare the outcomes of the studies. If we would systematically apply the same measure of interpersonal coordination to different tasks, and/or apply diverse measures of interpersonal coordination to one task, it would be possible to disentangle the influence of task specifics and measurement specifics on the research findings.

General Discussion

recruits hand movements when someone is speaking (Iverson & Thelen, 1999), while many of our hand movements, such as manipulating objects, happen without speech. Similar asymmetric coupling has been found between movements of someone's dominant and non-dominant hand (de Poel et al., 2007). Furthermore, in Study 3 (Chapter 4) I found that changes in spatiotemporal task properties influence both hand movements' and speech's variability. This might also indicate a bidirectional coupling between hand movements and speech, whereby changes in hand movements recruit changes in speech as well. In line with this, Pouw, Harrison et al. (2019) found entrainment between forceful, vertical hand movements, and the fundamental frequency and amplitude envelope of speech.

The above findings are supplemented and extended by the results of Study 2 (Chapter 3), in which students (instead of children) performed a simple cognitive task. In an easy, repetitive task, I found phase synchronization between gestures and speech, in line with the entrainment found by Pouw, Harrison et al. (2019). Furthermore, a difficult task with random variation even induced synergetic coupling between gestures and speech, which entails synchronization on multiple timescales. One way to interpret these findings is that they reflect two possible coordination patterns between gestures and speech. In the first coordination pattern, gestures and speech are like two phase-locked oscillators. This coordination pattern is characterized by close temporal alignment and high semantic similarity (Wagner et al., 2014). I hypothesize that this coordination pattern is the "default" mode of gesture-speech coordination, in line with how weakly coupled systems ubiquitously have been shown to synchronize (Rosenblum et al., 2001). This coordination pattern could be seen as a strong and stable attractor. In the second coordination pattern gestures and speech no longer behave as two oscillators, but instead form one large functional organization with different behavioral characteristics, such as low temporal alignment, low semantic similarity, but high accuracy of gestures. This second coordination pattern is a much weaker and less stable attractor than the first. Only situations and tasks with specific characteristics are able to elicit the second coordination pattern. I will further address the relation between the coupling between hand movements and speech and tasks with specific characteristics in the next section.

Hand movement-speech coupling is embedded and nested

Children's hand movements and speech, and the coupling between them, are embedded within the characteristics of the environment. Embedded means that the environment opens up and constrains possibilities for action (i.e. affordances; Adolph, 2019; Adolph & Hoch, 2019; Gibson, 1966), such as hand movements and speech. For example, in Study 2 (Chapter 3), students engaged in a task with a specific physical structure (targets in a regular or random pattern) and also a very specific instruction (match targets of the same color by means of pointing and saying

their location). The differences in physical structure between the two conditions resulted in clear differences in the coupling between students' hand movements and speech.

Also with regard to the science and technology tasks that I used in Study 1, 3, and 4 (Chapter 2, 4 and 5), children's hand movements and speech are embedded within tasks with very specific characteristics. These tasks require children to verbally, and preferably correctly, answer someone else's questions, and are thus characterized by a particular social structure. Anything a child does during these tasks is directed towards someone else. In addition, the verbal answering causes most of the child's hand movements to happen together with speech. In line with the social structure of such tasks, in Study 4 (Chapter 5), I found that children coordinate their speech, hand movements, and also head movements with the person with whom they are doing the task. Furthermore, the task material itself has a specific physical structure. For example, in a balance scale task the physical task characteristics are laid out so that only the distance from the fulcrum and the weights can change. Moreover, the social and physical structure of the task are also intertwined, as behaving according to the social structure (correctly answering the questions) requires children to attune to the relevant physical structure of the task material (also see E. J. Gibson & Pick, 2000; Adolph & Kretch, 2015).

Besides being embedded in science and technology tasks with particular social and physical characteristics, children's hand movements and speech, and the coupling between the two, are nested within cognitive understanding (see Figure 4 for a schematic visualization; also see General Introduction, p. 19-22; Adolph & Hoch, 2019; Kloos & Van Orden, 2009; Newen et al., 2018). Cognitive understanding can be seen as a coordination pattern which is functional within a specific task (for a more in-depth explanation, please see the General Introduction, p. 19-22). In the case of a science and technology task, this coordination pattern involves things like perceiving and attuning to particular physical properties of the environment, and communicating with someone else. Being nested means that cognitive understanding constrains and enables children's hand movements, speech, and hand movement-speech coupling, while children's hand movements, speech, and hand movement-speech coupling also constrain and enable cognitive understanding. Furthermore, the above definition of cognitive understanding as a functional coordination pattern within a specific task implies that cognitive understanding is also embedded within the characteristics of the environment, such as the science and technology tasks in this dissertation (Study 1, 3, and 4/Chapter 2, 4, and 5). The embeddedness and nestedness of children's (coupling of) hand movements and speech and cognitive understanding has consequences for hand movements' role in cognitive development.

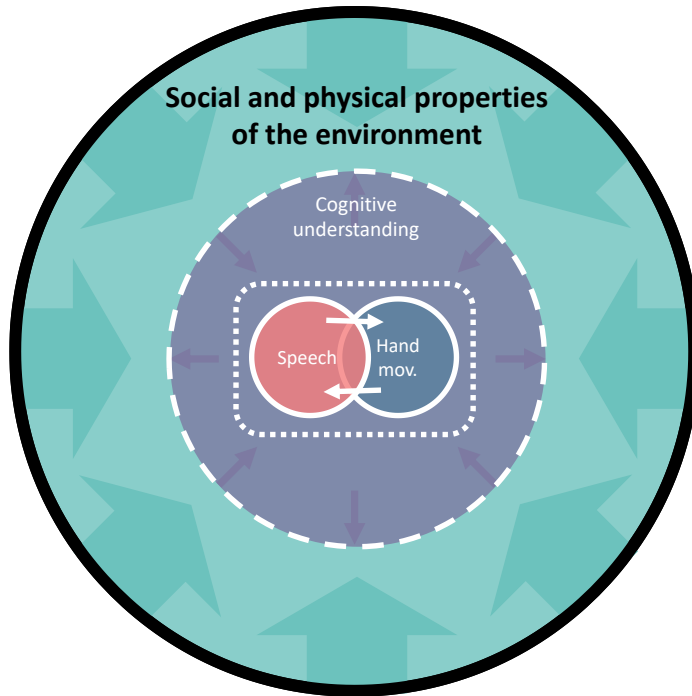


Figure 4. Speech, hand movements, and hand movement-speech-coupling are nested within cognitive understanding. This means that cognitive understanding constrains and enables children's hand movements, speech, and hand movement-speech coupling, while children's hand movements, speech, and hand movement-speech coupling also constrain and enable cognitive understanding (the bidirectional purple arrows). Furthermore, both cognitive understanding and speech, hand movements, and hand movement-speech-coupling are embedded in the social and physical properties of the environment (the uni-directional blue-green arrows).

Hand movements' (leading?) role in cognitive development

If children's hand movements and speech, and the coupling between them, as well as cognitive understanding are embedded within the characteristics of the environment, then hand movements' (leading) role in cognitive development is also dependent on, and embedded in, the task. For example, manual exploration enables someone to learn about the physical structure (e.g. shape, size, surface) of objects, among other things. People use their hands for this exploration, because hands are usually the parts of the body which are simply most suited for functionally exploring objects. Using speech (i.e. sound waves) to explore things like the shape, size or surface of an object would provide you with little to no useful information about the physical structure of the object. I therefore propose that it is strange to say that hand

movements are leading exploration of objects over speech, while speech was never able to provide any useful information about the physical structure of objects in the first place⁵.

With regard to gestures in specific, researchers often have used tasks similar to the science and technology tasks that I used in this dissertation (Study 1, 3, and 4/Chapter 2, 4, and 5) to investigate gestures' role in cognitive development. As described in the previous section, these tasks have very specific social and physical characteristics, whereby children need to perceive and attune to the relevant physical, spatial properties of the task in order to correctly answer someone's questions. Similar to manual exploration, children's hands are the prime candidates to explore the physical, spatial properties of the task material, while this is simply impossible for speech (also see Study 3 and the Discussion of Study 2). In other words, these tasks are devised to elicit hand movements. Different from manual exploration, such tasks also require children to talk, in order to answer the questions. As described before, speaking often recruits hand movements. In line with this, I could speculate that (deictic and representational) gestures reflect hand movements' simultaneous coordination with the rhythm of speech (coordination dynamics) and the typical possibilities of action of hands (affordances) (also see Pouw et al., 2018; Wagner et al., 2014). The above has implications for gestures' leading role in cognitive development, in the form of gesture-speech mismatches.

Why do gesture-speech mismatches occur?

During gesture-speech mismatches, children's hands correspond to the relevant physical properties of the task material for answering the questions, while their speech does not. Analogous to manual exploration, it is strange to say that gestures are leading cognitive development about tasks, which have specific physical and spatial properties, while speech is just not suited to fully correspond to these properties. Interestingly, also structural changes in eye movements, which are capable of moving according to the spatial structure of the task as well, have been shown to precede changes in verbal explanations (e.g. Dixon et al., 2012; Grant & Spivey, 2003). In that sense, instead of asking whether gestures are leading cognitive development, the following question about gesture-speech mismatches might be more useful: Why do children not speak the right words to indicate the relevant physical structure of the task, while their hands (or eyes) have started to attune and correspond to this physical structure?

⁵ This does **not** mean that speech is never important for manual exploration. Instead, many of children's exploratory actions will emerge in the context of social interactions, in which children speak with others. Speech and exploration will thus often elicit and constrain each other, which again emphasizes the tight coupling between speech and hand movement on many levels, their embeddedness in the physical and social environment, and their nestedness within cognitive development.

General Discussion

One important ingredient for answering this question is that gesture-speech mismatches are known to happen around cognitive insights (e.g. Church & Goldin-Meadow, 1986; Perry et al., 1992). As cognitive understanding is a functional coordination pattern, a cognitive insight is the emergence of a new coordination pattern. In line with this, a cognitive insight shows characteristics of a self-organized transition of the system (e.g. Stephen, Boncoddio, et al., 2009; Stephen, Dixon, et al., 2009). Such a transition is typically depicted as moving from a stable coordination pattern, through a variable transition phase, to another (more or less) stable coordination pattern (also see Study 3/Chapter 4). The variable transition phase means that both children's hand movements and speech will be variable too around cognitive insights.

Although hand movements are variable during such a transition phase, children can still (relatively easily) move and shape their hands according to the physical properties of the environment, as these physical properties are directly specified by the environment. However, there is relatively little in the environment that directly corresponds to children's speech, or the structure of sounds which corresponds to particular words (also see Study 2/Chapter 3). Therefore the variability in speech, in the form of searching for the right words, is more apparent than the variability of children's hand movements, even though hand movements also become more variable around a cognitive insight (Stephen, Boncoddio, et al., 2009; Stephen, Dixon, et al., 2009). The findings in Study 2 (Chapter 3), in which the difficult task led to a merged gesture-speech synergy, with low temporal alignment, low semantic similarity, but high gesture-accuracy, might also reflect such a variable phase.

Within science and technology tasks, in extreme cases the variability in speech might even cause children to reiterate a structure of sounds that they are familiar with and are able to put into words, and which one might interpret as "old" understanding⁶. When children's hands movements simultaneously attune to and correspond to the relevant physical properties of the task, one might interpret this combination as a gesture-speech mismatch. In addition, the coding systems which have led to the discovery of gesture-speech mismatches might exaggerate subtle differences between the content of gestures and speech (also see Koschmann, 2017), leading to even more coded gesture-speech mismatches.

In conclusion, hand movements' role in cognitive development is embedded in and dependent on the specifics of the task. When tasks have a physical, spatial structure, children's hand movements are often the most direct and suitable body parts to explore and learn about this physical structure. When a task also requires children to speak and answer questions, hand

⁶ In a sense, this might be similar to the perseveration in the A-not-B error (see General Introduction, p. 21-22 for an elaborate exploration), whereby children continue to search at location A while the toy has been hidden at location B.

movements are recruited by speech and become gestures, which are attuned to the rhythm of speech while also corresponding to the physical properties of the environment. During cognitive insights, children's hand movements and speech become variable, whereby hand movements can be structured according to the directly specified physical properties of the task, while there is nothing in the environment that directly specifies speech. While this might seem as if understanding in hand movements is ahead of speech, and is therefore coded as such, both hand movements and speech reflect the variability of the system during cognitive insights. In other words, neither hand movements nor speech are driving or leading cognitive development over the other. Instead, cognitive development and cognitive insights are fuzzy processes, which arise from the complex and nuanced interactions between the system as a whole, the coupling between a child's hand movements and speech, and the social and physical properties of the environment (also see the General Introduction, p. 11-14; 19-22).

Capturing the nuance of cognitive development

Cognitive development and cognitive insights are thus fuzzy processes, which emerge from what children do, whereby their actions are nested in an environment with specific characteristics. Capturing cognitive development in all its typical nuance, fuzziness, and nestedness requires specific methods, which are grounded in the theoretical perspectives of complex dynamical systems, coordination dynamics, and affordances - which I did. Most of these methods had never been applied to study children's (and students') hand movements, gestures, and speech during (science and technology) tasks.

To be specific, Study 1 (Chapter 2) was the first study to apply Cross RQA to investigate the coupling between gestures and speech. Moreover, it was the first study ever to apply anisotropic Cross RQA to examine the directionality of coupling between two systems. Applying anisotropic Cross RQA showed that speech typically attracts gestures more than vice versa, which would not have been found using other methods. Furthermore, Study 2 (Chapter 3) was the first study to simultaneously investigate gesture-speech synchronization in terms of temporal alignment, semantic similarity, and complexity matching, whereby complexity matching between gestures and speech had never been examined before. This led to the counterintuitive finding of more complexity matching between gestures and speech in the difficult than in the easy condition. To my knowledge, Study 3 (Chapter 4) was the first study which *empirically* applied a new categorical RQA measure for system's variability, namely block entropy. Using this RQA measure, I found changes in variability which were not picked up by another measure for variability that did not include temporal order. Lastly, Study 4 (Chapter 5) was the first study to simultaneously research interpersonal coordination between speech, hand movements, and head movements during collaborative problem solving of dyads of

General Discussion

children. I found that the timescales of interpersonal coordination between children's speech, hand movements, and head movements are similar but less pronounced than in adults, while the degree of interpersonal coordination between children did not seem to influence task performance.

In conclusion, applying methods which are grounded in the theoretical perspectives of complex dynamical systems, coordination dynamics, and affordances has yielded new as well as counterintuitive and nuanced findings. Furthermore, the studies in this dissertation show that these methods are applicable to research and capture a whole range of change processes in (cognitive) development and particularly its relation to hand movements and speech, including the nuanced and fuzzy processes that are typical for development, which future research could benefit from.

Different perspectives on children's hand movements and speech within cognitive development

Although the studies in this dissertation yielded new understanding of children's hand movements and speech within cognitive development, a number of important areas of study are also missing. In the next sections, I will focus on the most important ones, and discuss possible venues for future research.

First, investigating children's hand movements, gestures, and speech from the theoretical perspectives of complex dynamical systems, coordination dynamics, and affordances, meant that I focused on temporal patterns in this dissertation, instead of qualitative descriptions of children's behavior. However, connecting findings on the level of temporal patterns with findings on a qualitative level will yield additional and more pronounced understanding of the processes, coordination patterns, attractors, and relations between children and their environment, that underlie cognitive development. For example, such an approach would make it possible to empirically address my proposal that gesture-speech mismatches are (merely) a reflection of hand movements and speech being variable around cognitive insights. This could entail 1) capturing the intensity of children's and experimenters' hand movements and speech over time, and investigate hand movements' and speech's variability and coupling, 2) detailed coding of the timing *and* semantic content of children's and experimenter's hand movements and speech on multiple timescales (i.e. word by word, movement by movement, utterance by utterance, different strategies), and 3) a detailed description of the tasks' physical and social structure. Next to providing important understanding about cognitive development, I think qualitative descriptions are essential for bridging between quantitative findings and educational practice.

Second, while I focused on children's hand movements, gestures and speech within experimental and controlled tasks with a duration of 10 to 20 minutes, cognitive development extends far beyond these limited settings. If we are taking the theoretical perspectives of complex dynamical systems, coordination dynamics, and affordances, and the nested timescales which are inherent to these perspectives, seriously, we need to complement experimental studies with studies in which we investigate children longitudinally, with dense and diverse sampling of datapoints, and in a range of naturalistic settings (Adolph, 2019). In light of this, technological advances, such as OpenPose (Cao et al., 2021), DeepLabCut (Mathis et al., 2018), and wearables are promising. Such automated capturing and analysis techniques and devices allow accessible, cheap, and non-intrusive measuring of many different behaviors, such as movements, speech, heart rate, location, etc. (Developmental) Psychologists should profit from this, as these techniques and devices open up new possibilities to bring the methods we use to study behavior over time out of the lab and into the real world. Furthermore, I think studying children longitudinally and in naturalistic settings, using the technological advances that I just described, is essential for research findings to be meaningful in educational practice.

Lastly, in none of the studies did I address why children would want to learn something new in the first place. Moreover, this crucial issue is often not taken into account in studies about cognitive development, as well as in most studies which are grounded in the perspectives of complex dynamical systems, coordination dynamics, and affordances. However, as a mother observing and interacting with my children, it is impossible to overlook the importance of things like joy, curiosity, laughing, and mutual reinforcement for things my children learn. From the perspective of affordances, affordances are framed as possibilities for action, which enable and constrain, but *not* determine, what children (or other animals) do (e.g. Withagen et al., 2017). This leaves room for children to be motivated, at least to some degree, to engage with either one affordance or another. In the words by Withagen et al. (2017), the child is an agent, that is, an individual who is able to affect the degree to which an affordance has an effect on them. Furthermore, Wagman et al. (2016) propose that at any given time, many, hierarchically nested, affordances are present for any given individual. This could imply that researchers, parents, or teachers, can modulate this hierarchy of nested affordances, by means of shaping the physical and social characteristics of the environment, and thereby affording individual children to learn. We need more research to establish how we could do this for individual children at any point in time.

General Discussion

Practical implications for education

In line with the process of cognitive development, the practical implications of this dissertation are not straightforward (as they hardly ever are). However, I have some modest things to share with people involved in educational practice.

I found that children and students (hereafter: children) use both their hands, speech, and even move their heads according to the social and physical structure of the environment when they engage in (science and technology) tasks. Furthermore, I found that children's hand movements, gestures and speech are tightly and continuously coupled. I think it is important to be aware of this continuous coupling between children's modalities, and to allow children to communicate in such multimodal ways, and not only attend to, or favor, their speech. In addition, I found differences between kindergartners and first graders with regard to coupling between gestures and speech, which suggest a potential age difference in multimodal communication. This age difference could be important to take into account while teaching. Besides being coupled to each other, children's hand movements, gestures and speech are also attuned to the physical and social environment. Emphasizing particular properties of a task, either by means of the physical or social structure that is provided, influences children's hand movements, gestures, and speech, and thereby their cognitive development. Lastly, the findings in the studies show that cognitive development is not linear or straightforward, but instead is a nuanced, complex, fuzzy process, comprised of and nested within many other processes. This implies that children both need space to explore and try things out, using the parts of their body most suited for the job, as well as functional constraints as provided by the social and physical structure of the task.

I genuinely hope that, one day, I will have the chance to discuss these rather general practical implications with teachers. I would love to devise a plan *together* for how to translate these implications into an applied research program, deemed useful by teachers themselves (and yes, this is an invitation!)



Appendices

Appendix A

Coding procedure [Study 1/Chapter 2]

Coding of verbal expressions

All children's verbal expressions were coded in four steps using the computer program MediaCoder (Bos & Steenbeek, 2006). First, we started with the determination of the exact points in time when utterances of children started and ended. Then we classified the verbal utterances into six categories: Descriptive, predictive, and explanatory utterances; requests; content-related questions, and miscellaneous.

After these two steps, meaningful units of the children's coherent descriptive, predictive, and explanatory utterances were formed, the so-called 'units'. Each utterance corresponded to one unit. However, when two or more utterances only had a short pause in between (< 2 s), and focused on the very same topic, we also considered this as one unit, which meant that we could group them together in the next step of the coding process (see below). Each unit ended when the next expression of the child fell into another category, when there was a longer pause between the child's utterances, or when the researcher interrupted the child (e.g., by asking another question, or by making a procedural remark). An exception was made for short and simple expressions of encouragement of the researcher (e.g., "I see").

In the fourth step, the complexity of the utterances within a unit was determined. This meant that each unit was rated on a scale based on the model of dynamic skill theory developed by Fischer (1980). At Level 1 (sensorimotor actions), children stated single characteristics of the task, such as "This tube is long". At Level 2 (sensorimotor mappings), two elements of the task were coupled, such as "I can push this [piston] into here [pump]". At Level 3 (sensorimotor systems), simple causal mechanisms were stated, such as "If I push this [pump] in, the balloon grows bigger". At Level 4 (single representations), two causal mechanisms were coupled, or an "invisible" causal mechanism was mentioned, such as "When I push this [pump], *air* travels to the balloon". Explanations involving two causal relationships and an additional step were classified at Level 5 (representational mappings), e.g. "The piston pushes the air down, which goes through the tube to the other syringe, which piston then gets pushed out by the air". Level 6 (representational systems) comprised utterances in which all relevant representations that play a role within the task are mentioned. Level 7 comprised abstract utterances, for example about air pressure, or compression. Level 1-3 are part of the sensorimotor tier, level 4-6 are part of the representational tier, and level 7 is part of the abstract tier. In the original theory, 3

Coding procedure [Study 1/Chapter 2]

more levels are specified, but these develop at later ages and were therefore not specified for this study. Incorrect, irrelevant, and “don’t know”-answers were rated as incorrect.¹

The questions and units of answers received a code on an ordinal scale from 1 to 7 (ranging from sensorimotor actions to single abstractions). The coding 0 was used to mark the end of each utterance. Only utterances that displayed correct characteristics or possible task operations or mechanisms were coded as a skill level.

Coding of gestures

Gestures and task manipulations were coded in three steps. First, we coded the exact points in time when gestures and task manipulations of children started and ended. In this step we also noted whether the gesture could be characterized as 1) a short answer (short, task-related gestures, usually serving as an answer to a question), 2) a representation of the task or a task manipulation, or 3) an emblem. The latter category comprised task-*un*related short gestures with a rather universal character (e.g., ‘thumbs up’), which were not subject to further analysis.

In the second step, we further classified the categories short answers and representations/manipulations. Short answers were classified into: Nodding yes, shaking the head (“no”), lifting both shoulders (“don’t know”), and pointing toward (part of) the task. The representations/manipulations were further classified into: Representing a *characteristic* of the task, representing a *movement* of (elements of) the task, representing a *relationship* between two or more task elements, representing an *abstraction*, *single manipulations* (simple procedural manipulations of the task, e.g., pushing the syringe, turning a tap), and miscellaneous.

In the last step, we classified the short answers into ‘right’, ‘wrong’, or ‘other’, and we assigned skill levels to the representations/manipulations, again based on Skill Theory (Fischer, 1980; Fischer & Bidell, 2006). At Level 1 (sensorimotor actions), the child described (in gestures) a single characteristic of the task or an object that was directly observable (e.g., stating that something is heavy, soft, hard, small, etc.). At level 2 (sensorimotor mappings), the gestures of the child represented simple, observable relationships between elements of the task, for example, a gesture that depicts a simple direction of movement. At level 3 (sensorimotor systems), gestures depicted observable causal relationships between elements, such as describing two subsequent movements, or gestures depicting a cause and effect. Level 4 (single representations) comprised gestures not involving direct observable elements, such as when

¹ For earlier use and more examples of this scale, see Van Der Steen, Steenbeek, Wielinski, & Van Geert, 2012; Van Der Steen, Steenbeek, & Van Geert, 2012; see also Rappolt-Schlichtmann, Tenenbaum, Koepke, & Fischer, 2007 for another application of this theory.

the child made a prediction, or gestured about invisible mechanisms (air), or when he or she connected two causal relationships. Level 5 (representational mappings) is assigned when the child's gestures connect two or more single representations, such as correctly predicting (single representation 1) the flow of air (representation 2) within the task. Level 6 (representational systems) covers gestures in which all relevant representations that play a role within the task are mentioned. Finally, we scored level 7 when the gesture contained an abstraction, such as a representation of air compression. Incorrect, irrelevant, and "don't know"-answers were rated as incorrect.

The gestures received a code on an ordinal scale from 1 to 7 (ranging from sensorimotor actions to single abstractions). The coding 0 was used to mark the end of each utterance, and for utterances. Only utterances that displayed correct characteristics or possible task operations or mechanisms were coded as a skill level.

Appendix B

Detailed and accessible description of Multifractal Detrended Fluctuation Analysis (MFDFA) [Study 2/Chapter 3]

To provide an accessible introduction to Multifractal Detrended Fluctuation Analysis (MFDFA) to readers from diverse academic backgrounds, we will introduce the method in three main steps. First, we will illustrate how the box counting method is used to approximate the fractal dimension of objects. Second, we will illustrate how Detrended Fluctuation Analysis (DFA), which shares similarities with the box counting method, is used to approximate the temporal fractality of time series. Third, we will illustrate how MFDFA extends from DFA, and how it is used to approximate time series' temporal multifractality. For the first step, we largely follow David Feldman's (2019) highly accessible explanation of the box counting dimension, which is part of the [Fractals and Scaling course](#) from the Sante Fe Institute. For the second and third step, we largely follow the clear and recommended [explanation by Ihlen \(2012\)](#), which includes a script to perform MFDFA in Matlab.

Box counting method

As described in the Introduction of the main paper, objects that show self-similarity, i.e. that look similar at different levels of magnification, are *fractal*. However, the relation between level of magnification (s) and number of perfect “copies” (n) of the object differs for different fractal objects. For example, if we would dissect a line into two equal line segments, we would need to magnify the two line segments by a factor of two ($s = 1/2$), to see two perfect copies ($n = 2$) of our initial line (see Figure A1, left panel). Similarly, if we would dissect a line into three equal line segments, we would need to magnify the three line segments by a factor of three ($s = 1/3$), to see three perfect copies ($n = 3$) of our initial line. However, if we would dissect the lines of a square into two equal line segments each, we would create four smaller squares (see Figure A1, right panel). We would need to magnify these four smaller squares by a factor of two ($s = 1/2$), to see four perfect copies ($n = 4$) of our initial square. Similarly, if we would dissect the lines of a square into three equal line segments each, we would create nine smaller squares. We would need to magnify these nine smaller squares by a factor of three ($s = 1/3$), to see nine perfect copies ($n = 9$) of our initial square.

This relation between level of magnification (scaling factor) and number of copies (segments) is captured by the Hausdorff dimension, which is a form of fractal dimension. For mathematical objects, such as a line, a square, or the Koch snowflake (see Figure 2 in main paper, panel c), we can calculate the fractal dimension D by means of the following formula: $D = \frac{\log n}{\log 1/s}$, whereby

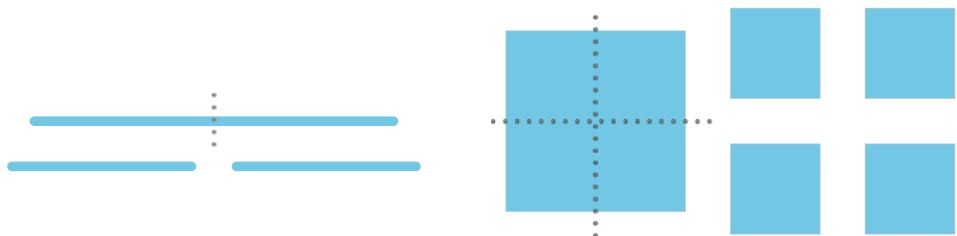


Figure A1. Dissecting the lines of a line and a square into 2 equal line segments ($s = \frac{1}{2}$). For the line, this creates 2 self-similar copies ($n = 2$). For the square this creates 4 self-similar copies ($n = 4$).

n = number of segments, and s = scaling factor. When we apply this formula to the previous examples of dissecting objects' lines into two equal line segments, D of a line is 1, and D of a square is 2. Using this same formula, D of the Koch snowflake is calculated to be around 1.26. Roughly speaking, the fractal dimension D is a measure for an object's complexity.

Next to mathematical objects, which show *perfect self-similarity*, many real world objects, such as Romanesco broccoli (see Figure 2 in main paper, panel d) or the coast of Britain (see Figure A2), are self-similar and thus fractal too, which is called *statistical self-similarity*. Different from mathematical objects, we cannot calculate the fractal dimension of real world objects exactly. Instead, we need to estimate their fractal dimension. The box counting method is a widely used method to estimate an object's fractal dimension. If we apply the box counting method to estimate an object's fractal dimension, we basically draw a grid of boxes of a certain size over that object and count the number of boxes of that particular size that are necessary to completely overlay the object. We repeat this procedure for grids with different box size (i.e. different side length). We subsequently plot the number of boxes that are needed to cover the object on the y-axis, and $1/\text{box side length}$ on the x-axis, on log-log scales. We can find the fractal dimension D of the object by drawing a regression line through the dots on the log-log plot, and calculating the slope of that line, that is the scaling relation. Figure A2 illustrates how we can use the box counting method to estimate the fractal dimension of a line, a square, and Britain's coast. Due to specific characteristics of biological time series, which we will explain next, we cannot directly apply the box counting method to estimate the fractal dimension of time series, however.

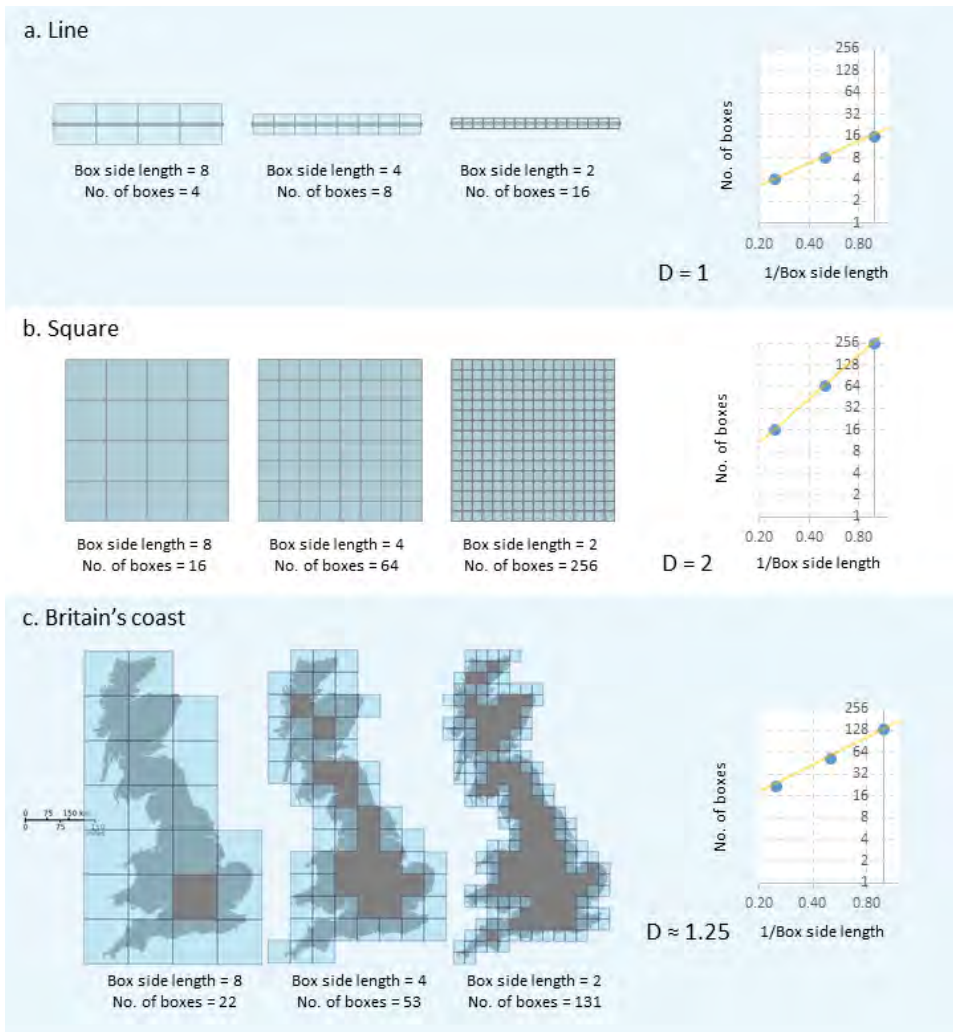


Figure A2. The box counting method to estimate the fractal dimension of a line (panel a), a square (panel b), and Britain's coast (panel c). We can estimate the fractal dimension by plotting the box side length against $[1/(\text{the number of boxes that are needed})]$ on a log-log plot, and calculating the slope of the resulting regression line. The fractal dimension D of a line is 1, of a square is 2, and Mandelbrot (1967) estimated the fractal dimension of Britain's coast to be 1.25.

Detrended Fluctuation Analysis

DFA is a method to determine the statistical self-similarity of a time series' variability (Ihlen, 2012; Peng et al., 1995). DFA's first step (Ihlen, 2012) is to transform the raw, noise-like, time series (see Figure A3, panel a) to an integrated, random-walk like time series (see Figure A3, panel b). The second step is to divide the time series into non-overlapping bins and calculate the variability within these bins, and repeat this for different bin sizes (see Figure A3, panel c). This second step has similarities with the box counting method, where now the bin size refers to the size of temporal window ('box') etc.

However, for biological time series, calculating the variability is not as straightforward as it may seem. Biological time series are typically non-stationary, which means that they stem from systems of which behavior changes over time (Peng et al., 1995). One part of these changes comes from random influences in the environment that we do not intend to measure. The other part of the changes comes from the system's internal dynamics, that we do want to measure. Peng et al. (1995) showed that accidental changes present themselves as changing trends in the biological time series. To calculate the scale-invariant variability for bins of non-stationary time series, we need to measure the variability *around these trends*, instead of the raw variability. For each bin, we therefore fitted a linear trend to the data (see the orange lines in Figure A3, panel c) and calculated the variability as the Root Mean Square of the residual variability (see orange, transparent, area around the orange lines in Figure A3, panel c), i.e. the *detrended fluctuation* or $RMS_{bins-scale}$.

After calculating the variability of all the bins with different sizes, DFA's next step is to calculate the overall Root Mean Square of each bin size scale, by means of the following formula: $RMS_{overall-scale} = \sqrt{RMS_{bins-scale}^2}$. We subsequently need to plot the $RMS_{overall}$ for the different scales on the y-axis, and the bin size on the x-axis, on log-log scale (see Figure A5, panel a). When we draw a regression line through this dots in the plot, the slope of this line corresponds to the Hurst exponent H . The Hurst exponent is a measure for the interdependence of datapoints in a time series. For example, for more random timeseries (i.e. Gaussian white noise) the datapoints are more independent, which corresponds to a $H \approx 0.5$. For time series with datapoints that are in between dependent and independent (i.e. pink noise; see Figure 2, panel a, in the main paper), $H \approx 1.0$. Following Wijnants et al. (2012), the Hurst exponent H is related to the fractal Dimension D according to the following formula: $D = 0.4H^2 - 1.2H + 2$.

Description of MFDFA [Study 2/Chapter 3]

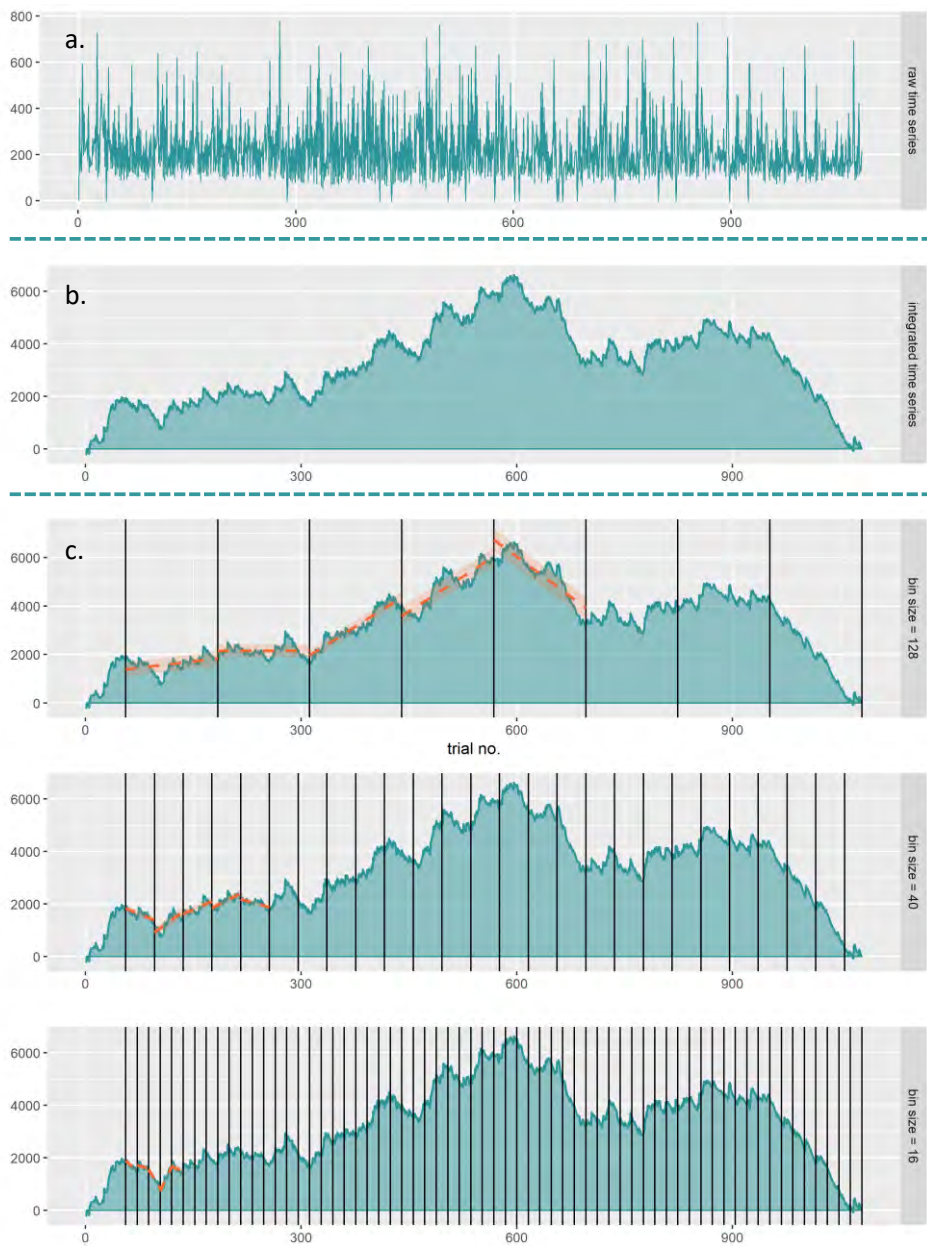


Figure A3. Steps of Detrended Fluctuation Analysis, illustrated with data from one participant in our experiment. Panel a shows the raw time series. This raw time series is then transformed to an integrated time series, which is shown in panel b. Panel c shows how the integrated time series is divided in increasingly smaller bin sizes, and the detrended fluctuation in each bin is calculated (orange lines).

Multifractal Detrended Fluctuation Analysis

Strictly speaking, only mathematical objects can be monofractal, that is, can be captured perfectly by *one* fractal dimension only. Real-world objects, such as Romanesco broccoli or Britain's coast, are more irregular and therefore better described by a range of fractal dimensions, although the size of this range varies from object to object. Cumulonimbus clouds, which usually develop into a thunderstorm, are a clear example of a multifractal natural object (see Figure A4). Different parts of the cloud are self-similar and fractal, with a different fractal dimension, and yet the fractality of these different parts is also related and intertwined. Also time series can have a multifractal structure. When time series are multifractal, periods of pink noise-like variability are intermitted by periods of much larger and much smaller fluctuations. These intermittent periods of larger and smaller fluctuations stem from processes at intertwined time series, and are thus not random but occur systematically. MFDFA is a method to approximate the range of fractal dimensions that characterize the variability of a time series.

To approximate the range of fractal dimensions of a time series, we need to measure and quantify its periods of larger and smaller fluctuations – something that DFA is unable to. MFDFA deals with this 'problem' by means of extending DFA with the *q-order*. As a brief reminder, for



Figure A4. Cumulonimbus cloud. The self-similarity of this cloud cannot be captured by one fractal dimension only, but varies for different parts of the cloud. This cloud is thus a multifractal object.

Description of MFDFA [Study 2/Chapter 3]

DFA, we calculate the overall Root Mean Square of each bin size scale by means of the following formula: $RMS_{overall-scale} = \sqrt{RMS_{bins-scale}^2}$. We thus calculate the variation at a bin size scale using the *second order* statistical moment (2). For MFDFA, we calculate the variation at a bin size scale using a range of *q-order* statistical moments. As a first step, we transform $RMS_{bins-scale}$ to $RMS_{bins-scale[q]}$, by means of the following formula: $RMS_{bins-scale[q]} = RMS_{bins-scale}^q$. As a second step, we calculate the overall q-order RMS: $RMS_{overall-scale[q]} = \overline{RMS_{bins-scale[q]}}^{1/q}$. The *q-order* essentially weights the influence of segments with large and small fluctuations on the overall q-order RMS. For negative q's, $RMS_{overall-scale[q]}$ is influenced by small fluctuations, while for positive q's, $RMS_{overall-scale[q]}$ is influenced by large fluctuations, whereby increasingly negative q-values emphasize increasingly smaller fluctuations, and increasingly positive q-values emphasize increasingly larger fluctuations. We subsequently can plot the $RMS_{overall[q]}$ for the different scales and different q-orders on the y-axis, and the bin size on the x-axis, on a log-log plot (see Figure A5, panel b). When a timeseries is multifractal, the slope of the regression line is different for different values of q.

While DFA uses the slope of the regression line as the outcome measure, i.e. the Hurst exponent H , MFDFA converts the q-order Hurst exponents $H(q)$ to the so-called multifractal spectrum. Researchers typically use the multifractal spectrum width as the outcome measure of MFDFA. We can create the multifractal spectrum in four steps. First, we convert $H(q)$ to the q-order mass exponent $t(q)$: $t(q) = H(q) * (q - 1)$. Second, we convert $t(q)$ to the q-order singularity exponent $h(q)$: $h(q) = \frac{dt(q)}{dq}$. Third, we convert $t(q)$ and $h(q)$ to the singularity dimension $D(q)$: $D(q) = 1 + qh(q) - t(q)$. Fourth, by plotting $h(q)$ on the x-axis and $D(q)$ on the y-axis, we create the multifractal spectrum (see Figure A6 for the multifractal spectrums of

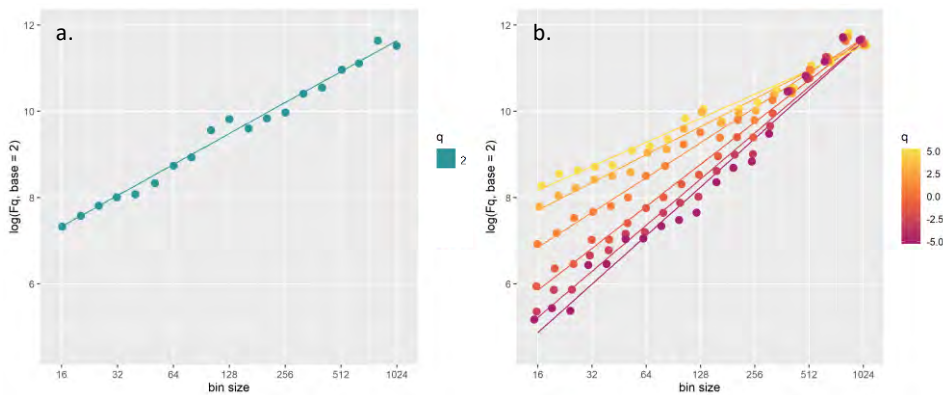


Figure A5. Log-log plots with $RMS_{overall}$ (Fq) on the y-axis and bin size on the x-axis, for one participant in our experiment. Panel a displays the dots and regression line of $q = 2$, as is the procedure for DFA. Panel b displays the dots and regression line of $-5 \leq q \leq 5$, as is the procedure for MFDFA.

gestures and speech of one participant in the difficult condition and one participant in the easy condition).

The multifractal spectrum is an arc (see Figure A6), and its shape informs us about the fractality of the timeseries (for a complete overview, see Ihlen, 2012). The central tendency of the multifractal spectrum (i.e. top of the arc) is closely related to the average fractal structure of the timeseries, or the Hurst exponent that is the outcome measure of DFA. The width of the arc informs us about the degree to which the timeseries' large and small fluctuations diverge from this average fractal structure. This means that timeseries that can be mostly characterized by one scaling relation will have a small multifractal spectrum width, while timeseries that can be characterized by a whole range of scaling relations will have a large multifractal spectrum width.

Complexity matching as difference in multifractal spectrum width

In the current study, we define complexity matching between gestures and speech as the difference in multifractal spectrum width. Similarly, Davis, Brooks and Dixon (2016) performed MFDFA and compared multifractal spectrum widths to investigate how two participants coordinate their hand movements during a joint task. Furthermore, using a different technique to create the multifractal spectrums, Stephen and Dixon (2011) compared multifractal spectrum widths to investigate how participants coordinate their finger tapping with a metronome that beats in a particular, and sometimes multifractal, pattern. It should be noted that Delignières, Almurad, Roume and Marmelat (2016) proposed a different method than comparing multifractal spectrum widths to investigate multifractal complexity matching.

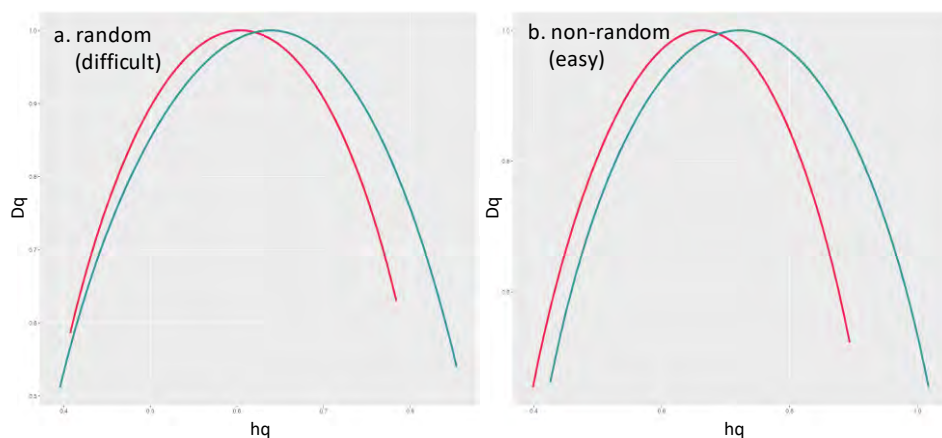


Figure A6. Multifractal spectrums of gestures (blue arc) and speech (red arc) for a participant in the difficult condition (panel a) and a participant in the easy condition (panel b). The difference in multifractal spectrum width is 0.081 for the participant in the difficult condition and 0.096 for the participant in the easy condition. We would interpret this as more complexity matching between gestures and speech for the participant in the difficult condition, compared to the participant in the easy condition.

Appendix C

Tables with results of stepwise mixed regression of transformed coherence and transformed relative phase angle on the three indices of task performance [Study 4/Chapter 5]:

- a) change in number of correct predictions
- b) agreement of (post-discussion) predictions
- c) adopting the other child's pre-discussion prediction for post-discussion predicting

Table A1

Results of stepwise regression of transformed coherence on indices of task performance

Dep. variable	Random effects		Fixed effects	R^2_{marg}	R^2_{cond}	Comparison with prev. model		
	Slope	Intercept				df	χ^2	p
change in number of correct predictions	coherence speech + coherence hand mov. + coherence head mov.	dyad	-	-	.156	-	-	-
			coherence speech	.018	.203	1	1.942	.163
			coherence speech + coherence hand mov.	.040	.203	1	3.090	.079
			coherence speech + coherence hand mov. + coherence head mov.	.051	.219	1	1.335	.248
			coherence speech * coherence hand mov. * coherence head mov.	.096	.276	4	5.816	.213
agreement of predictions	coherence speech + coherence hand mov. + coherence head mov.	dyad	-	-	.311	-	-	-
			coherence speech	.002	.311	1	0.059	.809
			coherence speech + coherence hand mov.	.002	.311	1	0.079	.778
			coherence speech + coherence hand mov. + coherence head mov.	.008	.327	1	.534	.465
			coherence speech * coherence hand mov. * coherence head mov.	.070	.419	4	4.003	.406
adopting the other child's pre- discussion prediction	coherence speech + coherence hand mov. + coherence head mov.	dyad	-	-	.132	-	-	-
			coherence speech	.010	.122	1	0.423	.515
			coherence speech + coherence hand mov.	.033	.199	1	2.349	.125
			coherence speech + coherence hand mov. + coherence head mov.	.036	.197	1	0.376	.540
			coherence speech * coherence hand mov. * coherence head mov.	.098	.327	4	4.161	.385

Note. The model with the interaction effect also contains the individual fixed effects of transformed coherence of speech, hand movements, and head movements.

Tables results stepwise mixed regression [Study 4/Chapter 5]

Table A2

Results of stepwise regression of transformed relative phase angle on indices of task performance

Dep. variable	Random effects		Fixed effects	R^2_{marg}	R^2_{cond}	Comparison with prev. model		
	Slope	Intercept				df	χ^2	p
change in number of correct predictions	r. p. angle speech + r. p. angle hand mov. + r. p. angle head mov.	dyad	-	-	.186	-	-	-
			r. p. angle speech	.004	.197	1	0.474	.491
			r. p. angle speech + r. p. angle hand mov.	.023	.225	1	2.521	.112
			r. p. angle speech + r. p. angle hand mov. + r. p. angle head mov.	.027	.231	1	0.319	.572
			r. p. angle speech * r. p. angle hand mov. * r. p. angle head mov.	.029	.221	4	0.699	.951
agreement of predictions	r. p. angle speech + r. p. angle hand mov. + r. p. angle head mov.	dyad	-	-	.244	-	-	-
			r. p. angle speech	.008	.271	1	0.430	.512
			r. p. angle speech + r. p. angle hand mov.	.009	.273	1	0.076	.782
			r. p. angle speech + r. p. angle hand mov. + r. p. angle head mov.	.011	.271	1	0.195	.659
			r. p. angle speech * r. p. angle hand mov. * r. p. angle head mov.	.110	.380	4	5.900	.207
adopting the other child's pre-discussion prediction	r. p. angle speech + r. p. angle hand mov. + r. p. angle head mov.	dyad	-	-	.421	-	-	-
			r. p. angle speech	.138	.389	1	0.423	.007
			r. p. angle speech + r. p. angle hand mov.	.155	.427	1	2.349	.197
			r. p. angle speech + r. p. angle hand mov. + r. p. angle head mov.	.171	.420	1	0.376	.260
			r. p. angle speech * r. p. angle hand mov. * r. p. angle head mov.	.329	.559	4	4.161	.080

Note. The model with the interaction effect also contains the individual fixed effects of transformed relative phase angle of speech, hand movements, and head movements.



References

- Abbarbanell, L., & Li, P. (2020). Unraveling the contribution of left-right language on spatial perspective taking. *Spatial Cognition & Computation*, 21(1), 1–38.
<https://doi.org/10.1080/13875868.2020.1825442>
- Abney, D. H. (2016). *The Complexity Matching hypothesis for human communication*. University of California.
- Abney, D. H., Dale, R., Louwerse, M. M., & Kello, C. T. (2018). The Bursts and Lulls of Multimodal Interaction: Temporal Distributions of Behavior Reveal Differences Between Verbal and Non-Verbal Communication. *Cognitive Science*, 42(4), 1297–1316. <https://doi.org/10.1111/cogs.12612>
- Abney, D. H., Paxton, A., Dale, R., & Kello, C. T. (2014). Complexity matching in dyadic conversation. *Journal of Experimental Psychology: General*, 143(6), 2304–2315. <https://doi.org/10.1037/xge0000021>
- Abney, D. H., Paxton, A., Dale, R., & Kello, C. T. (2015). Movement dynamics reflect a functional role for weak coupling and role structure in dyadic problem solving. *Cognitive Processing*, 16(4), 325–332.
<https://doi.org/10.1007/s10339-015-0648-2>
- Abney, D. H., Paxton, A., Dale, R., & Kello, C. T. (2021). Cooperation in sound and motion: Complexity matching in collaborative interaction. *Journal of Experimental Psychology: General*.
<https://doi.org/10.1037/xge0001018>
- Abney, D. H., Warlaumont, A. S., Haussman, A., Ross, J. M., & Wallot, S. (2014). Using nonlinear methods to quantify changes in infant limb movements and vocalizations. *Frontiers in Psychology*, 5.
<https://doi.org/10.3389/fpsyg.2014.00771>
- Abney, D. H., Warlaumont, A. S., Oller, D. K., Wallot, S., & Kello, C. T. (2016). Multiple Coordination Patterns in Infant and Adult Vocalizations. *Infancy*, 22(4), 514–539. <https://doi.org/10.1111/infa.12165>
- Adolph, K. E. (2020). An ecological approach to learning in (not and) development. *Human Development*, 63(3–4), 180–201. <https://doi.org/10.1159/000503823>
- Adolph, K. E., Cole, W. G., & Vereijken, B. (2015). Intraindividual variability in the development of motor skills in childhood. In *Handbook of intraindividual variability across the life span* (pp. 59–83). Routledge/Taylor & Francis Group. <https://doi.org/10.4324/9780203113066-14>
- Adolph, K. E., & Franchak, J. M. (2017). The development of motor behavior. *Wiley Interdisciplinary Reviews: Cognitive Science*, 8(1–2). <https://doi.org/10.1002/wcs.1430>
- Adolph, K. E., & Hoch, J. E. (2019). Motor development: Embodied, embedded, enculturated, and enabling. *Annual Review of Psychology*, 70, 141–164.
<https://doi.org/10.1146/annurev-psych-010418-102836>
- Adolph, K. E., Hoch, J. E., & Cole, W. G. (2018). Development (of walking): 15 suggestions. *Trends in Cognitive Sciences*, 22(8), 699–711. <https://doi.org/10.1016/j.tics.2018.05.010>
- Adolph, K. E., & Kretch, K. S. (2015). Gibson's theory of perceptual learning. In H. Keller (Ed.), *International Encyclopedia of the Social & Behavioral Sciences: Second Edition* (Vol. 2, pp. 127–134). Elsevier Inc. <https://doi.org/10.1016/b978-0-08-097086-8.23096-1>

References

- Alibali, M. W., Boncoddio, R., & Hostetter, A. B. (2014). Gesture in Reasoning: An Embodied Perspective. In *The Routledge Handbook of Embodied Cognition*. Routledge.
<https://doi.org/10.4324/9781315775845.ch15>
- Alibali, M. W., Kita, S., & Young, A. J. (2000). Gesture and the process of speech production: We think, therefore we gesture. *Language and Cognitive Processes*, 15(6), 593–613.
<https://doi.org/10.1080/016909600750040571>
- Alibali, M. W., & Nathan, M. J. (2012). Embodiment in mathematics teaching and learning: Evidence from learners' and teachers' gestures. *Journal of the Learning Sciences*, 21(2), 247–286.
<https://doi.org/10.1080/10508406.2011.611446>
- Alibali, M. W., & Goldin-Meadow, S. (1993a). Gesture-Speech Mismatch and Mechanisms of Learning: What the Hands Reveal about a Child's State of Mind. *Cognitive Psychology*, 25(4), 468–523.
<https://doi.org/10.1006/cogp.1993.1012>
- Alibali, M. W., & Goldin-Meadow, S. (1993b). Modeling learning using evidence from speech and gesture. *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society*, 203–208.
- Allen, S., Özyürek, A., Kita, S., Brown, A., Furman, R., Ishizuka, T., & Fujii, M. (2007). Language-specific and universal influences in children's syntactic packaging of Manner and Path: A comparison of English, Japanese, and Turkish. *Cognition*, 102(1), 16–48.
<https://doi.org/10.1016/j.cognition.2005.12.006>
- Almurad, Z. M. H., & Delignières, D. (2016). Evenly spacing in Detrended Fluctuation Analysis. *Physica A: Statistical Mechanics and Its Applications*, 451, 63–69.
<https://doi.org/10.1016/j.physa.2015.12.155>
- Almurad, Z. M. H., Roume, C., & Delignières, D. (2017). Complexity matching in side-by-side walking. *Human Movement Science*, 54, 125–136. <https://doi.org/10.1016/j.humov.2017.04.008>
- Alviar, C., Dale, R., Dewitt, A., & Kello, C. (2020). Multimodal Coordination of Sound and Movement in Music and Speech. *Discourse Processes*, 57(8), 682–702.
<https://doi.org/10.1080/0163853x.2020.1768500>
- Anastas, J. R., Stephen, D. G., & Dixon, J. A. (2011). The scaling behavior of hand motions reveals self-organization during an executive function task. *Physica A: Statistical Mechanics and Its Applications*, 390(9), 1539–1545. <https://doi.org/10.1016/j.physa.2010.11.038>
- Anderson, M. L., Richardson, M. J., & Chemero, A. (2012). Eroding the Boundaries of Cognition: Implications of Embodiment. *Topics in Cognitive Science*, 4(4), 717–730.
<https://doi.org/10.1111/j.1756-8765.2012.01211.x>
- Aßmann, B., Romano, M. C., Thiel, M., & Niemitz, C. (2007). Hierarchical organization of a reference system in newborn spontaneous movements. *Infant Behavior and Development*, 30(4), 568–586.
<https://doi.org/10.1016/j.infbeh.2007.04.004>
- Bahrami, B., Olsen, K., Latham, P. E., Roepstorff, A., Rees, G., & Frith, C. D. (2010). Optimally Interacting Minds. *Science*, 329(5995), 1081–1085. <https://doi.org/10.1126/science.1185718>

- Barabási, A.-L. (2005). The origin of bursts and heavy tails in human dynamics. *Nature*, 435(7039), 207–211. <https://doi.org/10.1038/nature03459>
- Barton, K. (2020). *Multi-Model Inference* (1.43.17) [Computer software].
- Bassano, D., & Van Geert, P. (2007). Modeling continuity and discontinuity in utterance length: A quantitative approach to changes, transitions and intra-individual variability in early grammatical development. *Developmental Science*, 10(5), 588–612. <https://doi.org/10.1111/j.1467-7687.2006.00629.x>
- Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Scheipl, F., Grothendieck, G., Green, P., Fox, J., Bauer, A., & Krivitsky, P. N. (2020). *Linear Mixed-Effects Models using "Eigen" and S4* (1.1-26) [Computer software].
- Beilock, S. L., & Goldin-Meadow, S. (2010). Gesture Changes Thought by Grounding It in Action. *Psychological Science*, 21(11), 1605–1610. <https://doi.org/10.1177/0956797610385353>
- Bergmann, K., Aksu, V., & Kopp, S. (2011). The relation of speech and gestures: Temporal synchrony follows semantic synchrony. *Proceedings of the 2nd Workshop on Gesture and Speech in Interaction (GeSpIn 2011)*.
- Bergmann, K., & Kopp, S. (2010). Modeling the production of coverbal iconic gestures by learning bayesian decision networks. *Applied Artificial Intelligence*, 24(6), 530–551. <https://doi.org/10.1080/08839514.2010.492162>
- Bergmann, K., & Kopp, S. (2012). Gestural alignment in natural dialogue. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 34(34).
- Bernstein, N. A. (1967). *The co-ordination and regulation of movements*. Pergamon Press.
- Bishop, L., & Goebel, W. (2017). Communication for coordination: Gesture kinematics and conventionality affect synchronization success in piano duos. *Psychological Research*, 82(6), 1177–1194. <https://doi.org/10.1007/s00426-017-0893-3>
- Boersma, P., & Weenink, D. (2018). *Praat: Doing Phonetics by Computer* (6.0.42) [Computer software]. <http://www.praat.org/>
- Boncoddo, R., Dixon, J. A., & Kelley, E. (2010). The emergence of a novel representation from action: Evidence from preschoolers. *Developmental Science*, 13(2), 370–377. <https://doi.org/10.1111/j.1467-7687.2009.00905.x>
- Boom, J., Hoijtink, H., & Kunnen, S. (2001). Rules in the balance: Classes, strategies, or rules for the balance scale task? *Cognitive Development*, 16(2), 717–735. [https://doi.org/10.1016/S0885-2014\(01\)00056-9](https://doi.org/10.1016/S0885-2014(01)00056-9)
- Bos, J., & Steenbeek, H. (2006). *MediaCoder: A Simple Application for Coding Behavior Within Media Files*. University of Groningen.
- Broaders, S. C., Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2007). Making children gesture brings out implicit knowledge and leads to learning. *Journal of Experimental Psychology: General*, 136(4), 539–550. <https://doi.org/10.1037/0096-3445.136.4.539>

References

- Brooks, N., & Goldin-Meadow, S. (2016). Moving to Learn: How Guiding the Hands Can Set the Stage for Learning. *Cognitive Science*, 40(7), 1831–1849. <https://doi.org/10.1111/cogs.12292>
- Brookshire, G., Lu, J., Nusbaum, H. C., Goldin-Meadow, S., & Casasanto, D. (2017). Visual cortex entrains to sign language. *Proceedings of the National Academy of Sciences*, 114(24), 6352–6357. <https://doi.org/10.1073/pnas.1620350114>
- Butcher, C., & Goldin-Meadow, S. (2000). Gesture and the transition from one- to two-word speech: When hand and mouth come together. In *Language and Gesture* (pp. 235–258). Cambridge University Press. <https://doi.org/10.1017/cbo9780511620850.015>
- Butterworth, G., & Hopkins, B. (1988). Hand-mouth coordination in the new-born baby. *British Journal of Developmental Psychology*, 8(4), 303–314. <https://doi.org/10.1111/j.1464-835X.1988.tb01103.x>
- Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., & Sheikh, Y. (2021). OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(1), 172–186. <https://doi.org/10.1109/tpami.2019.2929257>
- Carey, S., & Spelke, E. (1994). Domain-specific knowledge and conceptual change. In *Mapping the Mind* (pp. 169–200). Cambridge University Press. <https://doi.org/10.1017/cbo9780511752902.008>
- Castillo, R. D., Kloos, H., Richardson, M. J., & Waltzer, T. (2015). Beliefs as Self-Sustaining Networks: Drawing Parallels Between Networks of Ecosystems and Adults' Predictions. *Frontiers in Psychology*, 6. <https://doi.org/10.3389/fpsyg.2015.01723>
- Castillo, R. D., Waltzer, T., & Kloos, H. (2017). Hands-on experience can lead to systematic mistakes: A study on adults' understanding of sinking objects. *Cognitive Research: Principles and Implications*, 2(1). <https://doi.org/10.1186/s41235-017-0061-8>
- Chandrasekaran, C., Trubanova, A., Stillitano, S., Caplier, A., & Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. *PLoS Comput Biol*, 5(7), e1000436. <https://doi.org/10.1371/journal.pcbi.1000436>
- Chase, W. (2020, January 30). *The Glamour of Graphics*. rstudio::conf 2020, San Francisco. <https://rstudio.com/resources/rstudioconf-2020/the-glamour-of-graphics/>
- Chemero, A. (2011). *Radical embodied cognitive science*. MIT Press.
- Chemero, A. (2003). An Outline of a Theory of Affordances. *Ecological Psychology*, 15(2), 181–195. https://doi.org/10.1207/s15326969eco1502_5
- Chui, K. (2014). Mimicked gestures and the joint construction of meaning in conversation. *Journal of Pragmatics*, 70, 68–85. <https://doi.org/10.1016/j.pragma.2014.06.005>
- Church, R. B., & Goldin-Meadow, S. (1986). The mismatch between gesture and speech as an index of transitional knowledge. *Cognition*, 23(1), 43–71. [https://doi.org/10.1016/0010-0277\(86\)90053-3](https://doi.org/10.1016/0010-0277(86)90053-3)
- Cirelli, L. K. (2018). How interpersonal synchrony facilitates early prosocial behavior. *Current Opinion in Psychology*, 20, 35–39. <https://doi.org/10.1016/j.copsyc.2017.08.009>

- Clark, J. E., Whittall, J., & Phillips, S. J. (1988). Human interlimb coordination: The first 6 months of independent walking. *Developmental Psychobiology*, 21(5), 445–456.
<https://doi.org/10.1002/dev.420210504>
- Clearfield, M. W. (2004). The role of crawling and walking experience in infant spatial memory. *Journal of Experimental Child Psychology*, 89(3), 214–241. <https://doi.org/10.1016/j.jecp.2004.07.003>
- Coco, M. I., & Dale, R. (2014). Cross-recurrence quantification analysis of categorical and continuous time series: An R package. *Frontiers in Psychology*, 5. <https://doi.org/10.3389/fpsyg.2014.00510>
- Coey, C. A., Kallen, R. W., Chemero, A., & Richardson, M. J. (2018). Exploring complexity matching and asynchrony dynamics in synchronized and syncopated task performances. *Human Movement Science*, 62, 81–104. <https://doi.org/10.1016/j.humov.2018.09.006>
- Coey, C. A., Washburn, A., Hassebrock, J., & Richardson, M. J. (2016). Complexity matching effects in bimanual and interpersonal syncopated finger tapping. *Neuroscience Letters*, 616, 204–210. <https://doi.org/10.1016/j.neulet.2016.01.066>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum Associates.
- Cook, S. W., Duffy, R. G., & Fenn, K. M. (2013). Consolidation and Transfer of Learning After Observing Hand Gesture. *Child Development*, 84(6), 1863–1871. <https://doi.org/10.1111/cdev.12097>
- Cook, S. W., & Goldin-Meadow, S. (2006). The Role of Gesture in Learning: Do Children Use Their Hands to Change Their Minds? *Journal of Cognition and Development*, 7(2), 211–232.
https://doi.org/10.1207/s15327647jcd0702_4
- Cook, S. W., Yip, T. K., & Goldin-Meadow, S. (2012). Gestures, but not meaningless movements, lighten working memory load when explaining math. *Language and Cognitive Processes*, 27(4), 594–610. <https://doi.org/10.1080/01690965.2011.567074>
- Coupland, N. (1980). Style-shifting in a Cardiff work-setting. *Language in Society*, 9(1), 1–12. <https://doi.org/10.1017/s0047404500007752>
- Cox, R. F. A., Van der Steen, S., Guevara, M., De Jonge-Hoekstra, L., & Van Dijk, M. (2016). Chromatic and Anisotropic Cross-Recurrence Quantification Analysis of Interpersonal Behavior. In *Springer Proceedings in Physics* (pp. 209–225). Springer International Publishing.
https://doi.org/10.1007/978-3-319-29922-8_11
- Cox, R. F. A., & Van Dijk, M. (2013). Microdevelopment in Parent-Child Conversations: From Global Changes to Flexibility. *Ecological Psychology*, 23(3), 304–315. <https://doi.org/10.1080/10407413.2013.810095>
- Cox, R. F. A. (2016, July). *Complexity matching in a cooperative Wiimote game*. 14th European Workshop on Ecological Psychology, Groningen, the Netherlands.
- Crowder, E. M., & Newman, D. (1993). Telling what they know: The role of gesture and language in children's science explanations. *Pragmatics and Cognition*, 1(2), 341–376. <https://doi.org/10.1075/pc.1.2.06cro>

References

- Dale, R., & Spivey, M. J. (2006). Unraveling the Dyad: Using Recurrence Analysis to Explore Patterns of Syntactic Coordination Between Children and Caregivers in Conversation. *Language Learning*, 56(3), 391–430. <https://doi.org/10.1111/j.1467-9922.2006.00372.x>
- Davis, T. J., Brooks, T. R., & Dixon, J. A. (2016). Multi-scale interactions in interpersonal coordination. *Journal of Sport and Health Science*, 5(1), 25–34. <https://doi.org/10.1016/j.jshs.2016.01.015>
- De Graag, J. A., Cox, R. F. A., Hasselman, F., Jansen, J., & De Weerth, C. (2012). Functioning within a relationship: Mother–infant synchrony and infant sleep. *Infant Behavior and Development*, 35(2), 252–263. <https://doi.org/10.1016/j.infbeh.2011.12.006>
- De Jonge-Hoekstra, L., Cox, R. F. A., Van der Steen, S., & Dixon, J. A. (2021). Easier said than done? Task difficulty's influence on temporal alignment, semantic similarity, and complexity matching between gestures and speech. *Cognitive Science*. <https://doi.org/10.1111/cogs.12989>
- De Jonge-Hoekstra, L., Van der Steen, S., & Cox, R. F. A. (2020). Movers and shakers of cognition: Hand movements, speech, task properties, and variability. *Acta Psychologica*, 211. <https://doi.org/10.1016/j.actpsy.2020.103187>
- De Jonge-Hoekstra, L., Van der Steen, S., Van Geert, P., & Cox, R. F. A. (2016). Asymmetric Dynamic Attunement of Speech and Gestures in the Construction of Children's Understanding. *Frontiers in Psychology*, 7. <https://doi.org/10.3389/fpsyg.2016.00473>
- De Poel, H. J., Peper, C. E., & Beek, P. J. (2007). Handedness-related asymmetry in coupling strength in bimanual coordination: Furthering theory and evidence. *Acta Psychologica*, 124(2), 209–237. <https://doi.org/10.1016/j.actpsy.2006.03.003>
- De Vries, J. I., Visser, G. H., & Prechtl, H. F. (1982). The emergence of fetal behavior: I Qualitative aspects. *Early Human Development*, 7(4), 301–322. [https://doi.org/10.1016/0378-3782\(82\)90033-0](https://doi.org/10.1016/0378-3782(82)90033-0)
- Delgado, B., Gómez, J. C., & Sarriá, E. (2011). Pointing gestures as a cognitive tool in young children: Experimental evidence. *Journal of Experimental Child Psychology*, 110(3), 299–312. <https://doi.org/10.1016/j.jecp.2011.04.010>
- Delignières, D., Almurad, Z. M. H., Roume, C., & Marmelat, V. (2016). Multifractal signatures of complexity matching. *Experimental Brain Research*, 234(10), 2773–2785. <https://doi.org/10.1007/s00221-016-4679-4>
- Den Hartigh, R. J. R., Cox, R. F. A., Gernigon, C., Van Yperen, N. W., & Van Geert, P. L. C. (2015). Pink noise in rowing ergometer performance and the role of skill level. *Motor Control*, 19(4), 355–369. <https://doi.org/10.1123/mc.2014-0071>
- Den Hartigh, R. J. R., Marmelat, V., & Cox, R. F. A. (2018). Multiscale coordination between athletes: Complexity matching in ergometer rowing. *Human Movement Science*, 57, 434–441. <https://doi.org/10.1016/j.humov.2017.10.006>
- Denessen, E., Veenman, S., Dobbela, J., & Van Schilt, J. (2008). Dyad Composition Effects on Cognitive Elaboration and Student Achievement. *The Journal of Experimental Education*, 76(4), 363–386. <https://doi.org/10.3200/jexe.76.4.363-386>

- Dixon, J. A., Holden, J. G., Mirman, D., & Stephen, D. G. (2011). Multifractal Dynamics in the Emergence of Cognitive Structure. *Topics in Cognitive Science*, 4(1), 51–62.
<https://doi.org/10.1111/j.1756-8765.2011.01162.x>
- Dumas, G., & Fairhurst, M. T. (2019). *Reciprocity and alignment: Quantifying coupling in dynamic interactions*. <https://doi.org/10.31234/osf.io/nmg4x>
- Edelman, G. M., & Gally, J. A. (2001). Degeneracy and complexity in biological systems. *Proceedings of the National Academy of Sciences of the United States of America*, 98(24), 13763–13768.
[https://doi.org/10.1016/0378-3782\(82\)90033-0](https://doi.org/10.1016/0378-3782(82)90033-0)
- Ehrlich, S. B., Levine, S. C., & Goldin-Meadow, S. (2006). The Importance of Gesture in Children's Spatial Reasoning. *Developmental Psychology*, 42(6), 1259–1268.
<https://doi.org/10.1037/0012-1649.42.6.1259>
- Endedijk, H. M., Ramenzoni, V. C. O., Cox, R. F. A., Cillessen, A. H. N., Bekkering, H., & Hunnius, S. (2015). Development of interpersonal coordination between peers during a drumming task. *Developmental Psychology*, 51(5), 714–721. <https://doi.org/10.1037/a0038980>
- Esteve-Gibert, N., Borràs-Comes, J., Asor, E., Swerts, M., & Prieto, P. (2017). The timing of head movements: The role of prosodic heads and edges. *The Journal of the Acoustical Society of America*, 141(6), 4727–4739. <https://doi.org/10.1121/1.4986649>
- Esteve-Gibert, N., & Guellaï, B. (2018). Prosody in the auditory and visual domains: A developmental perspective. *Frontiers in Psychology*, 9. <https://doi.org/10.3389/fpsyg.2018.00338>
- Esteve-Gibert, N., & Prieto, P. (2014). Infants temporally coordinate gesture-speech combinations before they produce their first words. *Speech Communication*, 57, 301–316.
<https://doi.org/10.1016/j.specom.2013.06.006>
- Fajen, B. R., Riley, M. A., & Turvey, M. T. (2009). Information, affordances, and the control of action in sport. *International Journal of Sport Psychology*, 40(1), 79–107.
- Fawcett, L. M., & Garton, A. F. (2005). The effect of peer collaboration on children's problem-solving ability. *British Journal of Educational Psychology*, 75(2), 157–169.
<https://doi.org/10.1348/000709904x23411>
- Feldman, D. (2019). *Fractals and Scaling* [Massive Open Online Course].
<https://www.complexityexplorer.org/courses/93-fractals-and-scaling-winter-2019/segments/7786?summary>
- Fischer, K. W., & Bidell, T. R. (2006). Dynamic development of action, thought and emotion. In W. Damon & R. M. Lerner (Eds.), *Theoretical models of human development. Handbook of child psychology* (Vol. 6, pp. 313–399). Wiley.
- Fischer, K. W. (1980). A theory of cognitive development: The control and construction of hierarchies of skills. *Psychological Review*, 87(6), 477–531. <https://doi.org/10.1037/0033-295X.87.6.477>

References

- Fisher, C. B., & Camenzuli, C. A. (1987). Influence of body rotation on children's left-right confusion: A challenge to bilateral symmetry theory. *Developmental Psychology*, 23(2), 187. <https://doi.org/10.1037/0012-1649.23.2.187>
- Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47(6), 381–391. <https://doi.org/10.1037/h0055392>
- Fitzpatrick, P., Bui, P., & Garry, A. (2018). The Role of Perception–Action Systems in the Development of Tool-Using Skill. *Ecological Psychology*, 30(1), 74–98. <https://doi.org/10.1080/10407413.2017.1410044>
- Flynn, E., Pine, K., & Lewis, C. (2007). Using the microgenetic method to investigate cognitive development: An introduction. *Infant & Child Development*, 18(1), 1–6. <https://doi.org/10.1002/icd.503>
- Fowler, C. A. (2010). Embodied, Embedded Language Use. *Ecological Psychology*, 22(4), 286–303. <https://doi.org/10.1080/10407413.2010.517115>
- Fowler, C. A., Richardson, M. J., Marsh, K. L., & Shockley, K. D. (2008). Language use, coordination, and the emergence of cooperative action. In *Coordination: Neural, behavioral and social dynamics* (pp. 261–279). Springer. https://doi.org/10.1007/978-3-540-74479-5_13
- Fujiwara, K., Kimura, M., & Daibo, I. (2019a). Gender differences in synchrony: Females in sync during unstructured dyadic conversation. *European Journal of Social Psychology*. <https://doi.org/10.1002/ejsp.2587>
- Fujiwara, K., Kimura, M., & Daibo, I. (2019b). Rhythmic Features of Movement Synchrony for Bonding Individuals in Dyadic Interaction. *Journal of Nonverbal Behavior*, 44(1), 173–193. <https://doi.org/10.1007/s10919-019-00315-0>
- Furuyama, N. (2002). Prolegomena of a theory of between-person coordination of speech and gesture. *International Journal of Human-Computer Studies*, 57(4), 347–374. <https://doi.org/10.1006/ijhc.2002.1021>
- Fusaroli, R., Abney, D. H., Ashwini, R., Bahrami, B., Kello, C. T., & Tylén, K. (2013). *Conversation, coupling and complexity: Matching scaling laws predict performance in a joint decision task*. 35th Annual Conference of the Cognitive Science Society.
- Fusaroli, R., Bahrami, B., Olsen, K., Roepstorff, A., Rees, G., Frith, C., & Tylén, K. (2012). Coming to Terms: Quantifying the Benefits of Linguistic Coordination. *Psychological Science*, 23(8), 931–939. <https://doi.org/10.1177/0956797612436816>
- Fusaroli, R., Rączaszek-Leonardi, J., & Tylén, K. (2014). Dialog as interpersonal synergy. *New Ideas in Psychology*, 32, 147–157. <https://doi.org/10.1016/j.newideapsych.2013.03.005>
- Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, 13(3), 361–377. <https://doi.org/10.3758/BF03193857>
- Garber, P., & Goldin-Meadow, S. (2002). Gesture offers insight into problem-solving in adults and children. *Cognitive Science*, 26(6), 817. https://doi.org/10.1207/s15516709cog2606_5

- Gelman, R. (2002). Cognitive Development. In *Stevens' Handbook of Experimental Psychology*. John Wiley & Sons, Inc. <https://doi.org/10.1002/0471214426.pas0212>
- Gershkoff-Stowe, L., & Smith, L. B. (1997). A Curvilinear Trend in Naming Errors as a Function of Early Vocabulary Growth. *Cognitive Psychology*, 34(1), 37–71. <https://doi.org/10.1006/cogp.1997.0664>
- Gibson, E. J., & Pick, A. D. (2000). *An ecological approach to perceptual learning and development*. Oxford University Press.
- Gibson, J. J. (1979). *The ecological approach to visual perception*. Houghton, Mifflin and Company.
- Gibson, J.J. (1966). *The senses considered as perceptual systems*. Houghton Mifflin.
- Goldin-Meadow, S. (1998). The development of gesture and speech as an integrated system. *New Directions for Child and Adolescent Development*, 1998(79), 29–42. <https://doi.org/10.1002/cd.23219987903>
- Goldin-Meadow, S. (2003). *Hearing gesture: How our hands help us think*. Harvard University Press.
- Goldin-Meadow, S. (2007). Pointing Sets the Stage for Learning Language—And Creating Language. *Child Development*, 78(3), 741–745. <https://doi.org/10.1111/j.1467-8624.2007.01029.x>
- Goldin-Meadow, S. (2011). Learning through gesture. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(6), 595–607. <https://doi.org/10.1002/wcs.132>
- Goldin-Meadow, S. (2016). What the hands can tell us about language emergence. *Psychonomic Bulletin & Review*, 24(1), 213–218. <https://doi.org/10.3758/s13423-016-1074-x>
- Goldin-Meadow, S., & Alibali, M. W. (2013). Gesture's role in speaking, learning, and creating language. *Annual Review of Psychology*, 64, 257–283. <https://doi.org/10.1146/annurev-psych-113011-143802>
- Goldin-Meadow, S., Alibali, M. W., & Church, R. B. (1993). Transitions in Concept Acquisition: Using the Hand to Read the Mind. *Psychological Review*, 100(2), 279–297. <https://doi.org/10.1037/0033-295x.100.2.279>
- Goldin-Meadow, S., Cook, S. W., & Mitchell, Z. A. (2009). Gesturing Gives Children New Ideas About Math. *Psychological Science (0956-7976)*, 20(3), 267–272. <https://doi.org/10.1111/j.1467-9280.2009.02297.x>
- Goldin-Meadow, S., Levine, S. C., Zinchenko, E., Yip, T. K., Hemani, N., & Factor, L. (2012). Doing gesture promotes learning a mental transformation task better than seeing gesture. *Developmental Science*, 15(6), 876–884. <https://doi.org/10.1111/j.1467-7687.2012.01185.x>
- Goldin-Meadow, S., Nusbaum, H. C., Garber, P., & Church, R. B. (1993). Transitions in learning: Evidence for simultaneously activated strategies. *Journal of Experimental Psychology: Human Perception and Performance*, 19(1), 92–107. <https://doi.org/10.1037/0096-1523.19.1.92>
- Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining Math: Gesturing Lightens the Load. *Psychological Science (0956-7976)*, 12(6), 516. <https://doi.org/10.1111/1467-9280.00395>

References

- Goldin-Meadow, S., Wein, D., & Chang, C. (1992). Assessing knowledge through gesture: Using children's hands to read their minds. *Cognition and Instruction*, 9(3), 201–219. https://doi.org/10.1207/s1532690xci0903_2
- Granott, N., Fischer, K. W., & Parziale, J. (2002). Bridging to the unknown: A transition mechanism in learning and development. In N. Granott & J. Parziale (Eds.), *Microdevelopment: Transition processes in development and learning* (pp. 131–156). Cambridge University Press. <https://doi.org/10.1017/cbo9780511489709.006>
- Granott, N., & Parziale, J. (2002). *Microdevelopment: Transition processes in development and learning*. Cambridge University Press. <https://doi.org/10.1017/cbo9780511489709>
- Grant, E. R., & Spivey, M. J. (2003). Eye Movements and Problem Solving: Guiding Attention Guides Thought. *Psychological Science*, 14(5), 462–466. <https://doi.org/10.1111/1467-9280.02454>
- Grinsted, A., Moore, J. C., & Jevrejeva, S. (2004). Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear Processes in Geophysics*, 11(5/6), 561–566. <https://doi.org/10.5194/npg-11-561-2004>
- Guevara Guerrero, M. (2015). *Peer Interaction and Scientific Reasoning Processes in Preschoolers* [PhD Thesis]. University of Groningen.
- Haken, H., Kelso, J. A., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological Cybernetics*, 51(5), 347–356. <https://doi.org/10.1007/BF00336922>
- Haken, H. (1977). *Synergetics*. <https://doi.org/10.1007/978-3-642-96363-6>
- Hale, J., Ward, J. A., Buccheri, F., Oliver, D., & Hamilton, A. F. de C. (2019). Are You on My Wavelength? Interpersonal Coordination in Dyadic Conversations. *Journal of Nonverbal Behavior*, 44(1), 63–83. <https://doi.org/10.1007/s10919-019-00320-3>
- Harbourne, R. T., & Stergiou, N. (2009). Movement Variability and the Use of Nonlinear Tools: Principles to Guide Physical Therapist Practice. *Physical Therapy*, 89(3), 267–282. <https://doi.org/10.2522/ptj.20080130>
- He, L., & Dellwo, V. (2016, September). A Praat-Based Algorithm to Extract the Amplitude Envelope and Temporal Fine Structure Using the Hilbert Transform. *Proceedings of the Annual Conference of the International Speech Communication Association*. INTERSPEECH. <https://doi.org/10.21437/interspeech.2016-1447>
- He, L., & Dellwo, V. (2017). Amplitude envelope kinematics of speech: Parameter extraction and applications. *The Journal of the Acoustical Society of America*, 141(5), 3582–3582. <https://doi.org/10.1121/1.4987638>
- Hilliard, C., & Cook, S. W. (2015). A technique for continuous measurement of body movement from video. *Behavior Research Methods*, 49(1), 1–12. <https://doi.org/10.3758/s13428-015-0685-x>
- Hoekstra, L. (2012). *Curious Minds: Children's Hands and their Scientific Understanding: The construction of a coding system for gestures and acting on material, and an exploratory case study* [MSc. thesis].

- Holler, J., & Levinson, S. C. (2019). Multimodal Language Processing in Human Communication. *Trends in Cognitive Sciences*, 23(8), 639–652. <https://doi.org/10.1016/j.tics.2019.05.006>
- Holler, J., & Wilkin, K. (2011). Co-Speech Gesture Mimicry in the Process of Collaborative Referring During Face-to-Face Dialogue. *Journal of Nonverbal Behavior*, 35(2), 133–153. <https://doi.org/10.1007/s10919-011-0105-6>
- Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review*, 15(3), 495–514. <https://doi.org/10.3758/pbr.15.3.495>
- Hostetter, A. B., & Alibali, M. W. (2010). Language, Gesture, Action! A Test of the Gesture as Simulated Action Framework. *Journal of Memory and Language*, 63(2), 245–257. <https://doi.org/10.1016/j.jml.2010.04.003>
- Ihlen, E. A. F. (2012). Introduction to Multifractal Detrended Fluctuation Analysis in Matlab. *Frontiers in Physiology*, 3. <https://doi.org/10.3389/fphys.2012.00141>
- Ihlen, E. A. F., & Vereijken, B. (2010). Interaction-dominant dynamics in human cognition: Beyond 1/f fluctuation. *Journal of Experimental Psychology: General*, 139(3), 436–463. <https://doi.org/10.1037/a0019098>
- Iverson, J. M. (2010). Multimodality in infancy: Vocal-motor and speech-gesture coordinations in typical and atypical development. *Enfance*, 62(3), 257–274. <https://doi.org/10.4074/S0013754510003046.Multimodality>
- Iverson, J. M., & Fagan, M. K. (2004). Infant Vocal-Motor Coordination: Precursor to the Gesture-Speech System? *Child Development*, 75(4), 1053–1066. <https://doi.org/10.1111/j.1467-8624.2004.00725.x>
- Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture Paves the Way for Language Development. *Psychological Science* (0956-7976), 16(5), 367–371. <https://doi.org/10.1111/j.0956-7976.2005.01542.x>
- Iverson, J. M., & Thelen, E. (1999). Hand, mouth and brain: The dynamic emergence of speech and gesture. *Journal of Consciousness Studies*, 8(11–12), 19–40.
- Jacoby, W. G. (2000). Loess: A nonparametric, graphical tool for depicting relationships between variables. *Electoral Studies*, 19(4), 577–613. [https://doi.org/10.1016/s0261-3794\(99\)00028-1](https://doi.org/10.1016/s0261-3794(99)00028-1)
- Jaffe, J., Beatrice, B., Stanley, F., Crown, C. L., & Jasnow, M. D. (2001). Rhythms of dialogue in infancy: Coordinated timing in development. *Monographs of the Society for Research in Child Development*, 66(2), vi–131. <https://doi.org/10.2307/3181589>
- Jansen, B. R. J., & Van der Maas, H. L. J. (2002). The development of children's rule use on the balance scale task. *Journal of Experimental Child Psychology*, 81(4), 383–416. <https://doi.org/10.1006/jecp.2002.2664>
- Jarvis, S. (2013). Capturing the Diversity in Lexical Diversity. *Language Learning*, 63, 87–106. <https://doi.org/10.1111/j.1467-9922.2012.00739.x>
- Jost, L. (2006). Entropy and diversity. *Oikos*, 113(2), 363–375. <https://doi.org/10.1111/j.2006.0030-1299.14714.x>

References

- Kantelhardt, J. W., Zschiegner, S. A., Koscielny-Bunde, E., Havlin, S., Bunde, A., & Stanley, H. E. (2002). Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A: Statistical Mechanics and Its Applications*, 313(1–4), 87–114.
[https://doi.org/10.1016/S0378-4371\(02\)01383-3](https://doi.org/10.1016/S0378-4371(02)01383-3)
- Karsai, M., Kaski, K., Barabási, A.-L., & Kertész, J. (2012). Universal features of correlated bursty behaviour. *Scientific Reports*, 2(1). <https://doi.org/10.1038/srep00397>
- Kello, C. T., Beltz, B. C., Holden, J. G., & Van Orden, G. C. (2007). The emergent coordination of cognitive function. *Journal of Experimental Psychology: General*, 136(4), 551–568.
<https://doi.org/10.1037/0096-3445.136.4.551>
- Kelso, J. A. S. (1994). The informational character of self-organized coordination dynamics. *Human Movement Science*, 13(3–4), 393–413. [https://doi.org/10.1016/0167-9457\(94\)90047-7](https://doi.org/10.1016/0167-9457(94)90047-7)
- Kelso, J. A. S. (1995). *Dynamic patterns: The self-organization of brain and behavior*. The MIT Press.
- Kelso, J. A. S. (2013). Coordination Dynamics. In *Encyclopedia of Complexity and Systems Science* (pp. 1–41). Springer New York. https://doi.org/10.1007/978-3-642-27737-5_101-3
- Kelty-Stephen, D. G., Palatinus, K., Saltzman, E., & Dixon, J. A. (2013). A Tutorial on Multifractality, Cascades, and Interactivity for Empirical Time Series in Ecological Science. *Ecological Psychology*, 25(1), 1–62.
<https://doi.org/10.1080/10407413.2013.753804>
- Kendon, A. (1972). Some Relationships Between Body Motion and Speech. In *Studies in Dyadic Communication* (pp. 177–210). Elsevier. <https://doi.org/10.1016/b978-0-08-015867-9.50013-7>
- Kita, S. (2000). How representational gestures help speaking. In D. McNeill (Ed.), *Language and Gesture* (pp. 162–185). Cambridge University Press. <https://doi.org/10.1017/cbo9780511620850.011>
- Kita, S., Alibali, M. W., & Chu, M. (2017). How do gestures influence thinking and speaking? The gesture-for-conceptualization hypothesis. *Psychological Review*, 124(3), 245–266.
<https://doi.org/10.1037/rev0000059>
- 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. [https://doi.org/10.1016/s0749-596x\(02\)00505-3](https://doi.org/10.1016/s0749-596x(02)00505-3)
- Kita, S., Van Gijn, I., & Van der Hulst, H. (1998). Movement phases in signs and co-speech gestures, and their transcription by human coders. In I. Wachsmuth & M. Fröhlich (Eds.), *Gesture and Sign Language in Human-Computer Interaction, International Gesture Workshop Bielefeld, Germany, September 17-19, 1997, Proceedings. Lecture Notes in Artificial Intelligence* (pp. 23–35). Springer Berlin Heidelberg. <https://doi.org/10.1007/bfb0052986>
- Kloos, H., Fisher, A., & Van Orden, G. C. (2010). Situated naïve physics: Task constraints decide what children know about density. *Journal of Experimental Psychology: General*, 139(4), 625–637.
<https://doi.org/10.1037/a0020977>
- Kloos, H., & Van Orden, G. (2010). Voluntary behavior in cognitive and motor tasks. *Mind and Matter*, 8(1), 19–43.

- Kloos, H., & Van Orden, G. C. (2009). Soft-Assembled Mechanisms for the Unified Theory. In *Toward a Unified Theory of Development Connectionism and Dynamic System Theory Re-Consider* (pp. 253–267). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195300598.003.0012>
- Konvalinka, I., Xygalatas, D., Bulbulia, J., Schjodt, U., Jegindo, E.-M., Wallot, S., Van Orden, G., & Roepstorff, A. (2011). Synchronized arousal between performers and related spectators in a fire-walking ritual. *Proceedings of the National Academy of Sciences*, 108(20), 8514–8519. <https://doi.org/10.1073/pnas.1016955108>
- Koschmann, T. (2017). Why would the discovery of gestures produced by signers jeopardize the experimental finding of gesture-speech mismatch? *The Behavioral and Brain Sciences*, 40, e60. <https://doi.org/10.1017/s0140525x15002952>
- Krol, K., Janssen, J., Veenman, S., & Van der Linden, J. (2004). Effects of a Cooperative Learning Program on the Elaborations of Students Working in Dyads. *Educational Research and Evaluation*, 10(3), 205–237. <https://doi.org/10.1076/edre.10.3.205.30269>
- Kugler, P. N., & Turvey, M. T. (1987). *Information, natural law, and the self-assembly of rhythmic movement*. Lawrence Erlbaum Associates, Inc. <https://doi.org/10.4324/9781315648460>
- Kuhn, D., Pease, M., & Wirkala, C. (2009). Coordinating the effects of multiple variables: A skill fundamental to scientific thinking. *Journal of Experimental Child Psychology*, 103(3), 268–284. <https://doi.org/10.1016/j.jecp.2009.01.009>
- Latash, M. L. (2008). *Synergy*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195333169.001.0001>
- Lenth, R. V., Buurkner, P., Herve, M., Love, J., Riebl, H., & Singmann, H. (2020). *Estimated Marginal Means, aka Least-Squares Means* (1.5.3) [Computer software].
- Leonard, T., & Cummins, F. (2011). The temporal relation between beat gestures and speech. *Language and Cognitive Processes*, 26(10), 1457–1471. <https://doi.org/10.1080/01690965.2010.500218>
- Leonardi, G. (2018). A Method for the computation of entropy in the Recurrence Quantification Analysis of categorical time series. *Physica A: Statistical Mechanics and Its Applications*, 512, 824–836. <https://doi.org/10.1016/j.physa.2018.08.058>
- Lichtwarck-Aschoff, A., Hasselman, F., Cox, R., Pepler, D., & Granic, I. (2012). A Characteristic Destabilization Profile in Parent-Child Interactions Associated with Treatment Efficacy for Aggressive Children. *Nonlinear Dynamics, Psychology, and Life Sciences*, 16(3), 353–379.
- Lockman, J. J. (2000). A perception-action perspective on tool use development. *Child Development*, 71(1), 137–144. <https://doi.org/10.1111/1467-8624.00127>
- Loehr, D. (2007). Aspects of rhythm in gesture and speech. *Gesture*, 7(2), 179–214. <https://doi.org/10.1111/j.1551-6709.2012.01269.x>
- Louwerse, M. M., Dale, R., Bard, E. G., & Jeuniaux, P. (2012). Behavior Matching in Multimodal Communication Is Synchronized. *Cognitive Science*, 36(8), 1404–1426. <https://doi.org/10.1111/j.1551-6709.2012.01269.x>

References

- Lund, U., Agostinelli, C., Hiroyoshi, A., Gagliardi, A., Garcia Portugues, E., Giunchi, D., Irisson, J.-O., Pocernich, M., & Rotolo, F. (2017). *Circular Statistics* (0.4-93) [Computer software].
- Macuch Silva, V., Holler, J., Ozyurek, A., & Roberts, S. G. (2020). Multimodality and the origin of a novel communication system in face-to-face interaction. *Royal Society Open Science*, 7(1), 182056. <https://doi.org/10.1098/rsos.182056>
- Marghetis, T., Núñez, R., & Bergen, B. K. (2014). Doing arithmetic by hand: Hand movements during exact arithmetic reveal systematic, dynamic spatial processing. *The Quarterly Journal of Experimental Psychology*, 67(8), 1579–1596. <https://doi.org/10.1080/17470218.2014.897359>
- Marmelat, V., & Delignières, D. (2012). Strong anticipation: Complexity matching in interpersonal coordination. *Experimental Brain Research*, 222(1–2), 137–148. <https://doi.org/10.1007/s00221-012-3202-9>
- Marsh, K. L., Richardson, M. J., & Schmidt, R. C. (2009). Social connection through joint action and interpersonal coordination. *Topics in Cognitive Science*, 1(2), 320–339. <https://doi.org/10.1111/j.1756-8765.2009.01022.x>
- Martin, N., Chevrot, J.-P., & Barbu, S. (2010). Stylistic variations in the social network of a 10-year-old child: Pragmatic adjustments or automatic alignment? *Journal of Sociolinguistics*, 14(5), 678–692. <https://doi.org/10.1111/j.1467-9841.2010.00459.x>
- Marwan, N., Romano, M. C., Thiel, M., & Kurths, J. (2007). Recurrence plots for the analysis of complex systems. *Physics Reports. A Review Section of Physics Letters*, 438(5–6), 237. <https://doi.org/10.1016/j.physrep.2006.11.001>
- Marwan, N., & Webber Jr., C. L. (2014). Mathematical and computational foundations of recurrence quantifications. In N. Marwan & C. L. Webber (Eds.), *Recurrence Quantification Analysis: Theory and best practices* (pp. 3–43). Springer. https://doi.org/10.1007/978-3-319-07155-8_1
- Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., & Bethge, M. (2018). DeepLabCut: Markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience*, 21(9), 1281–1289. <https://doi.org/10.1038/s41593-018-0209-y>
- Mathôt, S., Schreij, D., & Theeuwes, J. (2011). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, 44(2), 314–324. <https://doi.org/10.3758/s13428-011-0168-7>
- Max Planck Institute for Psycholinguistics, The Language Archive. (2020). *ELAN* (6.0) [Computer software]. <https://archive.mpi.nl/tla/elan>
- McKinley, J., Dempster, M., & Gormley, G. J. (2015). ‘Sorry, I meant the patient’s left side’: Impact of distraction on left–right discrimination. *Medical Education*, 49(4), 427–435. <https://doi.org/10.1111/medu.12658>
- McNeill, D. (1985). So you think gestures are nonverbal? *Psychological Review*, 92(3), 350–371. <https://doi.org/10.1037/0033-295x.92.3.350>

- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago Press.
<http://search.ebscohost.com.proxy-ub.rug.nl/login.aspx?direct=true&db=psyh&AN=1992-98214-000&site=ehost-live&scope=site>
- Meindertsma, H. B. (2014). *Predictions and explanations* [PhD Thesis]. University of Groningen.
- Melinger, A., & Kita, S. (2007). Conceptualisation Load Triggers Gesture Production. *Language and Cognitive Processes*, 22(4), 473–500. <https://doi.org/10.1080/01690960600696916>
- Messer, D. J., Pine, K. J., & Butler, C. (2008). Children's behaviour and cognitions across different balance tasks. *Learning and Instruction*, 18(1), 42–53. <https://doi.org/10.1016/j.learninstruc.2006.09.008>
- Meyer, M., & Hunnius, S. (2020). Becoming better together: The early development of interpersonal coordination. In *Progress in Brain Research* (pp. 187–204). Elsevier.
<https://doi.org/10.1016/bs.pbr.2020.06.012>
- Molenaar, I., Slegers, P., & Van Boxtel, C. (2014). Metacognitive scaffolding during collaborative learning: A promising combination. *Metacognition and Learning*, 3(3), 309–332.
<https://doi.org/10.1007/s11409-014-9118-y>
- Molenaar, P. C. M., & Campbell, C. G. (2009). The New Person-Specific Paradigm in Psychology. *Current Directions in Psychological Science*, 18(2), 112–117.
<https://doi.org/10.1111/j.1467-8721.2009.01619.x>
- Myowa-Yamakoshi, M., & Takeshita, H. (2006). Do Human Fetuses Anticipate Self-Oriented Actions? A Study by Four-Dimensional (4D) Ultrasonography. *Infancy*, 15(3), 289–301.
https://doi.org/10.1207/s15327078in1003_5
- Nakagawa, S., & Schielzeth, H. (2012). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 3(2), 133–142.
<https://doi.org/10.1111/j.2041-210x.2012.00261.x>
- Newen, A., De Bruin, L., & Gallagher, S. (Eds.). (2018). *The Oxford Handbook of 4E Cognition*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198735410.001.0001>
- Ninness, C., Newton, R., Saxon, J., Rumph, R., Bradfield, A., Harrison, C., Vasquez, E. I., & Ninness, S. K. (2002). Small group statistics: A Monte Carlo comparison of parametric and randomization tests. *Behavior and Social Issues*, 12(1), 53–63. <https://doi.org/10.5210/bsi.v12i1.79>
- Novack, M. A., Congdon, E. L., Hemani-Lopez, N., & Goldin-Meadow, S. (2014). From action to abstraction: Using the hands to learn math. *Psychological Science*, 25(4), 903–910.
<https://doi.org/10.1177/0956797613518351>
- Novack, M., & Goldin-Meadow, S. (2015). Learning from Gesture: How Our Hands Change Our Minds. *Educational Psychology Review*, 27(3), 405–412. <https://doi.org/10.1007/s10648-015-9325-3>
- Olthof, M., Hasselman, F., Strunk, G., Aas, B., Schiepek, G., & Lichtwarck-Aschoff, A. (2020). Destabilization in self-ratings of the psychotherapeutic process is associated with better treatment outcome in patients with mood disorders. *Psychotherapy Research*, 30(4), 520–531.
<https://doi.org/10.1080/10503307.2019.1633484>

References

- Olthof, M., Hasselman, F., Strunk, G., Van Rooij, M., Aas, B., Helmich, M. A., Schiepek, G., & Lichtwarck-Aschoff, A. (2020). Critical fluctuations as an early-warning signal for sudden gains and losses in patients receiving psychotherapy for mood disorders. *Clinical Psychological Science*, 8(1), 25–35. <https://doi.org/10.1177/2167702619865969>
- Paxton, A., & Dale, R. (2017). Interpersonal Movement Synchrony Responds to High- and Low-Level Conversational Constraints. *Frontiers in Psychology*, 8. <https://doi.org/10.3389/fpsyg.2017.01135>
- Pedersen, T. L. (2020). *Accelerating "ggplot2"* (0.3.2) [Computer software].
- Pellecchia, G. L., & Shockley, K. (2005). Application of recurrence quantification analysis: Influence of cognitive activity on postural fluctuations. *Tutorials in Contemporary Nonlinear Methods for the Behavioral Sciences*, 95–141.
- Peng, C.-K., Havlin, S., Stanley, H. E., & Goldberger, A. L. (1995). Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 5(1), 82–87. <https://doi.org/10.1063/1.166141>
- Pennings, H. J. M., Brekelmans, M., Sadler, P., Claessens, L. C. A., Van der Want, A. C., & Van Tartwijk, J. (2018). Interpersonal adaptation in teacher-student interaction. *Learning & Instruction*, 55, 41–57. <https://doi.org/10.1016/j.learninstruc.2017.09.005>
- Perry, M., Church, R. B., & Goldin-Meadow, S. (1988). Transitional knowledge in the acquisition of concepts. *Cognitive Development*, 3(4), 359–400. [https://doi.org/10.1016/0885-2014\(88\)90021-4](https://doi.org/10.1016/0885-2014(88)90021-4)
- Perry, M., Church, R. B., & Goldin-Meadow, S. (1992). Is gesture-speech mismatch a general index of transitional knowledge? *Cognitive Development*, 7(1), 109–122. [https://doi.org/10.1016/0885-2014\(92\)90007-E](https://doi.org/10.1016/0885-2014(92)90007-E)
- Piaget, J., & Cook, M. (Trans.). (1952). *The origins of intelligence in children* (2nd ed.). W W Norton & Co.
- Piaget, J., & Cook, M. (Trans.). (1954). The development of object concept. In *The construction of reality in the child* (pp. 3–96). Basic Books. <https://doi.org/10.1037/11168001>
- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(2). <https://doi.org/10.1017/s0140525x04000056>
- Pikovsky, A., Kurths, J., Rosenblum, M., & Kurths, J. (2001). *Synchronization: A universal concept in nonlinear sciences*. Cambridge University Press. <https://doi.org/10.1017/cbo9780511755743>
- Pine, K. J., Lufkin, N., Kirk, E., & Messer, D. (2007). A Microgenetic Analysis of the Relationship between Speech and Gesture in Children: Evidence for Semantic and Temporal Asynchrony. *Language and Cognitive Processes*, 22(2), 234–246. <https://doi.org/10.1080/01690960600630881>
- Pine, K. J., Lufkin, N., & Messer, D. (2004). More Gestures Than Answers: Children Learning about Balance. *Developmental Psychology*, 40(6), 1059–1067. <https://doi.org/10.1037/0012-1649.40.6.1059>
- Poeppel, D., & Assaneo, M. F. (2020). Speech rhythms and their neural foundations. *Nature Reviews Neuroscience*, 21(6), 322–334. <https://doi.org/10.1038/s41583-020-0304-4>
- Pouw, W., & Dixon, J. A. (2019a). Quantifying gesture-speech synchrony. *Proceedings of the 6th Gesture and Speech in Interaction Conference*. <https://doi.org/10.17619/UNIPB/1-815>

- Pouw, W., & Dixon, J. A. (2019b). Entrainment and Modulation of Gesture–Speech Synchrony Under Delayed Auditory Feedback. *Cognitive Science*, 43(3), e12721. <https://doi.org/10.1111/cogs.12721>
- Pouw, W., Harrison, S. A., & Dixon, J. A. (2018). *Gesture-Speech Physics: The Biomechanical Basis for the Emergence of Gesture-Speech Synchrony*. <https://doi.org/10.31234/osf.io/tgua4>
- Pouw, W. T. J. L., De Jonge-Hoekstra, L., & Dixon, J. A. (2018). *Stabilizing Speech Production through Gesture-Speech Coordination*. <https://doi.org/10.31234/osf.io/arzne>
- Pouw, W. T. J. L., De Nooijer, J. A., Van Gog, T., Zwaan, R. A., & Paas, F. (2014). Toward a more embedded/extended perspective on the cognitive function of gestures. *Frontiers in Psychology*, 5. <https://doi.org/10.3389/fpsyg.2014.00359>
- Pouw, W. T. J. L., Van Gog, T., Zwaan, R. A., & Paas, F. (2017). Are gesture and speech mismatches produced by an integrated gesture-speech system? A more dynamically embodied perspective is needed for understanding gesture-related learning. *The Behavioral and Brain Sciences*, 40, e68. <https://doi.org/10.1017/S0140525X15003039>
- Pouw, W., & Trujillo, J. P. (2019). *Tutorial Gesp2019—Using video-based motion tracking to quantify speech-gesture synchrony*. <https://doi.org/10.17605/OSF.IO/RXB8J>
- Pouw, W., Trujillo, J. P., & Dixon, J. A. (2019). The quantification of gesture–speech synchrony: A tutorial and validation of multimodal data acquisition using device-based and video-based motion tracking. *Behavior Research Methods*, 52(2), 723–740. <https://doi.org/10.3758/s13428-019-01271-9>
- Prieto, P., & Roseano, P. (2018). Prosody: Stress, Rhythm, and Intonation. In K. L. Geeslin (Ed.), *The Cambridge Handbook of Spanish Linguistics* (Vol. 1, pp. 211–236). Cambridge University Press. <https://doi.org/10.1017/9781316779194.011>
- Ramirez-Aristizabal, A. G., Médé, B., & Kello, C. T. (2018). Complexity matching in speech: Effects of speaking rate and naturalness. *Chaos, Solitons & Fractals*, 111, 175–179. <https://doi.org/10.1016/j.chaos.2018.04.021>
- Rasenberg, M., Özyürek, A., & Dingemanse, M. (2020). Alignment in Multimodal Interaction: An Integrative Framework. *Cognitive Science*, 44(11). <https://doi.org/10.1111/cogs.12911>
- Rauchbauer, B., & Grosbras, M.-H. (2020). Developmental trajectory of interpersonal motor alignment: Positive social effects and link to social cognition. *Neuroscience & Biobehavioral Reviews*, 118, 411–425. <https://doi.org/10.1016/j.neubiorev.2020.07.032>
- Reed, B. S. (2010). Prosody and alignment: A sequential perspective. *Cultural Studies of Science Education*, 5(4), 859–867. <https://doi.org/10.1007/s11422-010-9289-z>
- Reed, E. S. (1982). An outline of a theory of action systems. *Journal of Motor Behavior*, 14(2), 98–134. <https://doi.org/10.1080/00222895.1982.10735267>
- Reed, E. S. (1995). The Ecological Approach to Language Development: A Radical Solution to Chomsky's and Quine's Problems. *Language & Communication*, 15(1), 1–29. [https://doi.org/10.1016/0271-5309\(94\)E0010-9](https://doi.org/10.1016/0271-5309(94)E0010-9)

References

- Reuzel, E., Embregts, P., Bosman, A., Cox, R., Van Nieuwenhuijzen, M., & Jahoda, A. (2013). Conversational Synchronization in Naturally Occurring Settings: A Recurrence-Based Analysis of Gaze Directions and Speech Rhythms of Staff and Clients with Intellectual Disability. *Journal of Nonverbal Behavior*, 37(4), 281–305. <https://doi.org/10.1007/s10919-013-0158-9>
- Reuzel, E., Embregts, P. J. C. M., Bosman, A. M. T., Cox, R. F. A., Van Nieuwenhuijzen, M., & Jahoda, A. (2014). Verbal Interactional Dominance and Coordinative Structure of Speech Rhythms of Staff and Clients with an Intellectual Disability. *Nonlinear Dynamics, Psychology, and Life Sciences*, 18(4), 371–396.
- Richardson, D. C., & Dale, R. (2005). Looking to Understand: The Coupling between Speakers' and Listeners' Eye Movements and Its Relationship to Discourse Comprehension. *Cognitive Science*, 29(6), 1045–1060. https://doi.org/10.1207/s15516709cog0000_29
- Richardson, D. C., Dale, R., & Kirkham, N. Z. (2007). The Art of Conversation Is Coordination: Common Ground and the Coupling of Eye Movements During Dialogue. *Psychological Science (0956-7976)*, 18(5), 407–413. <https://doi.org/10.1111/j.1467-9280.2007.01914.x>
- Richardson, M. J. (2009). *Polhemus applications and example code*.
- Richardson, Michael J., Marsh, K. L., Isenhower, R. W., Goodman, J. R. L., & Schmidt, R. C. (2007). Rocking together: Dynamics of intentional and unintentional interpersonal coordination. *Human Movement Science*, 26(6), 867–891. <https://doi.org/10.1016/j.humov.2007.07.002>
- Richardson, Michael J., Marsh, K. L., & Schmidt, R. C. (2005). Effects of Visual and Verbal Interaction on Unintentional Interpersonal Coordination. *Journal of Experimental Psychology: Human Perception and Performance*, 31(1), 62–79. <https://doi.org/10.1037/0096-1523.31.1.62>
- Richardson, Michael J., Schmidt, R. C., & Kay, B. A. (2007). Distinguishing the noise and attractor strength of coordinated limb movements using recurrence analysis. *Biological Cybernetics*, 90(1), 59–78. <https://doi.org/10.1007/s00422-006-0104-6>
- Rieber, R. W., & Carton, A. S. (Eds.). (1988). *The Collected Works of L. S. Vygotsky*. Springer US. <https://doi.org/10.1007/978-1-4613-1655-8>
- Rigoli, L. M., Holman, D., Spivey, M. J., & Kello, C. T. (2014). Spectral convergence in tapping and physiological fluctuations: Coupling and independence of 1/f noise in the central and autonomic nervous systems. *Frontiers in Human Neuroscience*, 8. <https://doi.org/10.3389/fnhum.2014.00713>
- Riley, M. A., & Turvey, M. T. (2002). Variability and Determinism in Motor Behavior. *Journal of Motor Behavior*, 34(2), 99–125. <https://doi.org/10.1080/00222890209601934>
- Romero, V., Fitzpatrick, P., Roulier, S., Duncan, A., Richardson, M. J., & Schmidt, R. C. (2018). Evidence of embodied social competence during conversation in high functioning children with autism spectrum disorder. *PLOS ONE*, 13(3), e0193906. <https://doi.org/10.1371/journal.pone.0193906>
- Rösch, A., & Schmidbauer, H. (2016). *WaveletComp: A guided tour through the R-package*. 38.
- Rösch, A., & Schmidbauer, H. (2018). *Computational Wavelet Analysis* (1.1) [Computer software].

- Rosenblum, M., Pikovsky, A., Kurths, J., Schäfer, C., & Tass, P. A. (2001). Phase synchronization: From theory to data analysis. In *Neuro-Informatics and Neural Modelling* (pp. 279–321). Elsevier.
[https://doi.org/10.1016/s1383-8121\(01\)80012-9](https://doi.org/10.1016/s1383-8121(01)80012-9)
- Roth, W.-M. (2002). From action to discourse: The bridging function of gestures. *Cognitive Systems Research*, 3(1–4), 535–554. [https://doi.org/10.1016/s1389-0417\(02\)00056-6](https://doi.org/10.1016/s1389-0417(02)00056-6)
- Rowe, S. M., & Wertsch, J. V. (2002). Vygotsky's model of cognitive development. *Blackwell Handbook of Childhood Cognitive Development*, 538–554. <https://doi.org/10.1002/9780470996652.ch24>
- Roy, B. C., Frank, M. C., DeCamp, P., Miller, M., & Roy, D. (2015). Predicting the birth of a spoken word. *Proceedings of the National Academy of Sciences*, 112(41), 12663–12668.
<https://doi.org/10.1073/pnas.1419773112>
- Roy, D. (2011). *Deb Roy: The birth of a word*. TED Talk.
- Rusiewicz, H., & Esteve-Gibert, N. (2018). Set in time: Temporal coordination of prosody and gesture in the development of spoken language production. In P. Prieto & N. Esteve-Gibert (Eds.), *The Development of Prosody in First Language Acquisition* (pp. 103–124). Benjamins Publishing Company. <https://doi.org/10.1075/tilar.23.06rus>
- Rusiewicz, H. L., Shaiman, S., Iverson, J. M., & Szuminsky, N. (2013). Effects of Prosody and Position on the Timing of Deictic Gestures. *Journal of Speech, Language, and Hearing Research*, 56(2), 458–470.
[https://doi.org/10.1044/1092-4388\(2012\)11-0283](https://doi.org/10.1044/1092-4388(2012)11-0283)
- Rusiewicz, H. L., Shaiman, S., Iverson, J. M., & Szuminsky, N. (2014). Effects of perturbation and prosody on the coordination of speech and gesture. *Speech Communication*, 57, 283–300.
<https://doi.org/10.1016/j.specom.2013.06.004>
- Satta, E., Ferrari-Toniolo, S., Visco-Comandini, F., Caminiti, R., & Battaglia-Mayer, A. (2017). Development of motor coordination during joint action in mid-childhood. *Neuropsychologia*, 105, 111–122.
<https://doi.org/10.1016/j.neuropsychologia.2017.04.027>
- Schloesser, D. S., Kello, C. T., & Marmelat, V. (2019). Complexity matching and coordination in individual and dyadic performance. *Human Movement Science*, 66, 258–272.
<https://doi.org/10.1016/j.humov.2019.04.010>
- Schmidt, R. C., Fitzpatrick, P., Caron, R., & Mergeche, J. (2011). Understanding social motor coordination. *Human Movement Science*, 30(5), 834–845. <https://doi.org/10.1016/j.humov.2010.05.014>
- Schmidt, R. C., Nie, L., Franco, A., & Richardson, M. J. (2014). Bodily synchronization underlying joke telling. *Frontiers in Human Neuroscience*, 8. <https://doi.org/10.3389/fnhum.2014.00633>
- Schmidt, R. C., & O'Brien, B. (1997). Evaluating the Dynamics of Unintended Interpersonal Coordination. *Ecological Psychology*, 9(3), 189–206. https://doi.org/10.1207/s15326969eco0903_2
- Schmidt, R. C., & Turvey, M. T. (1994). Phase-entrainment dynamics of visually coupled rhythmic movements. *Biological Cybernetics*, 70(4), 369–376. <https://doi.org/10.1007/s004220050040>

References

- Schmidt, Richard C., & Richardson, M. J. (2008). Dynamics of Interpersonal Coordination. In A. Fuchs & V. K. Jirsa (Eds.), *Coordination: Neural, Behavioral and Social Dynamics* (pp. 281–308). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-74479-5_14
- Schneider, S., Ramirez-Aristizabal, A. G., Gavilan, C., & Kello, C. T. (2019). Complexity matching and lexical matching in monolingual and bilingual conversations. *Bilingualism: Language and Cognition*, 23(4), 845–857. <https://doi.org/10.1017/s1366728919000774>
- Schöner, G., & Dineva, E. (2007). Dynamic instabilities as mechanisms for emergence. *Developmental Science*, 10(1), 69–74. <https://doi.org/10.1111/j.1467-7687.2007.00566.x>
- Schöner, G., & Thelen, E. (2006). Using Dynamic Field Theory to Rethink Infant Habituation. *Psychological Review*, 113(2), 273–299. <https://doi.org/10.1037/0033-295X.113.2.273>
- Schrauf, C., Call, J., & Pauen, S. (2011). The Effect of Plausible Versus Implausible Balance Scale Feedback on the Expectancies of 3- to 4-Year-Old Children. *Journal of Cognition and Development*, 12(4), 518–536. <https://doi.org/10.1080/15248372.2011.571647>
- Schwartz, M., & Fischer, K. W. (2004). Building general knowledge and skill: Cognition and microdevelopment in science learning. In *Cognitive developmental change: Theories, models and measurement* (pp. 157–185). Cambridge University Press. <https://doi.org/10.1017/cbo9780511489938.006>
- Seifert, L., Komar, J., Araújo, D., & Davids, K. (2016). Neurobiological degeneracy: A key property for functional adaptations of perception and action to constraints. *Neuroscience and Biobehavioral Reviews*, 69, 159–165. <https://doi.org/10.1016/j.neubiorev.2016.08.006>
- Shanahan, L. E., & Roof, L. M. (2013). Developing Strategic Readers: A Multimodal Analysis of a Primary School Teacher's Use of Speech, Gesture and Artefacts. *Literacy*, 47(3), 157–164. <https://doi.org/10.1111/lit.12002>
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3), 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- Sherwin, W. B., Chao, A., Jost, L., & Smouse, P. E. (2017). Information Theory Broadens the Spectrum of Molecular Ecology and Evolution. *Trends in Ecology & Evolution*, 32(12), 948–963. <https://doi.org/10.1016/j.tree.2017.09.012>
- Shockley, K., Butwill, M., Zbilut, J. P., & Webber Jr., C. L. (2002). Cross recurrence quantification of coupled oscillators. *Physics Letters A*, 305(1/2), 59. [https://doi.org/10.1016/S0375-9601\(02\)01411-1](https://doi.org/10.1016/S0375-9601(02)01411-1)
- Shockley, K., Richardson, D. C., & Dale, R. (2009). Conversation and Coordinative Structures. *Topics in Cognitive Science*, 1(2), 305–319. <https://doi.org/10.1111/j.1756-8765.2009.01021.x>
- Shockley, K., Santana, M.-V., & Fowler, C. A. (2003). Mutual Interpersonal Postural Constraints Are Involved in Cooperative Conversation. *Journal of Experimental Psychology: Human Perception & Performance*, 29(2), 326. <https://doi.org/10.1037/0096-1523.29.2.326>
- Siegler, R. S. (1976). Three aspects of cognitive development. *Cognitive Psychology*, 8(4), 481–520. [https://doi.org/10.1016/0010-0285\(76\)90016-5](https://doi.org/10.1016/0010-0285(76)90016-5)

- Siegler, R. S. (1989). Mechanisms of cognitive development. *Annual Review of Psychology*, 40(1), 353.
<https://doi.org/10.1146/annurev.ps.40.020189.002033>
- Siegler, R. S. (2006). Microgenetic Analyses of Learning. In *Handbook of child psychology: Cognition, perception, and language*, Vol. 2, 6th ed (6th ed., pp. 464–510). John Wiley & Sons Inc.
<https://doi.org/10.1002/9780470147658.chpsy0211>
- Siegler, R. S. (2007). Cognitive variability. *Developmental Science*, 10(1), 104–109.
<https://doi.org/10.1111/j.1467-7687.2007.00571.x>
- Smith, L. B. (2005). Cognition as a dynamic system: Principles from embodiment. *Developmental Review*, 25(3–4), 278–298. <https://doi.org/10.1016/j.dr.2005.11.001>
- Smith, L. B. (2010). Action as Developmental Process: A Commentary on Iverson's "Developing Language in a Developing Body: The Relationship between Motor Development and Language Development." *Journal of Child Language*, 37(2), 263–267. <https://doi.org/10.1017/S0305000909990535>
- Smith, L. B., & Thelen, E. (2003). Development as a dynamic system. *Trends in Cognitive Sciences*, 7(8), 343.
[https://doi.org/10.1016/S1364-6613\(03\)00156-6](https://doi.org/10.1016/S1364-6613(03)00156-6)
- Smith, L. B., Thelen, E., Titzer, R., & McLin, D. (1999). Knowing in the context of acting: The task dynamics of the A-not-B error. *Psychological Review*, 106(2), 235–260. <https://doi.org/10.1037/0033-295x.106.2.235>
- Smith, L., & Gasser, M. (2005). The Development of Embodied Cognition: Six Lessons from Babies. *Artificial Life*, 11(1/2), 13–30. <https://doi.org/10.1162/1064546053278973>
- Smitsman, A. W., & Bongers, R. M. (2003). Tool Use and Tool Making: A Developmental Action Perspective. In *Handbook of Developmental Psychology* (pp. 172–193). SAGE Publications Ltd.
<https://doi.org/10.4135/9781848608306.n8>
- Sparling, J. W., Arena, J., & Chescheir, N. C. (1999). Fetal and neonatal hand movement. *Physical Therapy*, 79(1), 24–39. <https://doi.org/10.1093/ptj/79.1.24>
- Spencer, J. P., Smith, L. B., & Thelen, E. (2001). Tests of a dynamic systems account of the A-not-B error: The influence of prior experience on the spatial memory abilities of two-year-olds. *Child Development*, 72(5), 1327–1346. <https://doi.org/10.1093/ptj/79.1.24>
- Stephen, D. G., Anastas, J. R., & Dixon, J. A. (2012). Scaling in cognitive performance reflects multiplicative multifractal cascade dynamics. *Frontiers in Physiology*, 3.
<https://doi.org/10.3389/fphys.2012.00102>
- Stephen, D. G., Boncoddò, R. A., Magnuson, J. S., & Dixon, J. A. (2009). The dynamics of insight: Mathematical discovery as a phase transition. *Memory & Cognition*, 37(8), 1132–1149.
<https://doi.org/10.3758/MC.37.8.1132>
- Stephen, D. G., Dixon, J. A., & Isenhowe, R. W. (2009). Dynamics of representational change: Entropy, action, and cognition. *Journal of Experimental Psychology: Human Perception and Performance*, 35(6), 1811–1832. <https://doi.org/10.1037/a0014510>

References

- Stephen, D. G., Stepp, N., Dixon, J. A., & Turvey, M. T. (2008). Strong anticipation: Sensitivity to long-range correlations in synchronization behavior. *Physica A: Statistical Mechanics and Its Applications*, 387(21), 5271–5278. <https://doi.org/10.1016/j.physa.2008.05.015>
- Stern, D. N., Spieker, S., Barnett, R. K., & MacKain, K. (1983). The Prosody of Maternal Speech: Infant Age and Context-Related Changes. *Journal of Child Language*, 10(1), 1–15. <https://doi.org/10.1017/S0305000900005092>
- Sternberg, R. J. (1984). *Mechanisms of cognitive development*. Freeman.
- Stoffregen, T. A. (2003). Affordances as Properties of the Animal-Environment System. *Ecological Psychology*, 15(2), 115–134. https://doi.org/10.1207/s15326969eco1502_2
- Storch, N. (2001). How collaborative is pair work? ESL tertiary students composing in pairs. *Language Teaching Research*, 5(1), 29–53. <https://doi.org/10.1191/136216801666650977>
- Strogatz, S. H. (2012). *Sync: How order emerges from chaos in the universe, nature, and daily life*. Hyperion.
- Tamis-LeMonda, C. S., Bornstein, M. H., & Baumwell, L. (2001). Maternal responsiveness and children's achievement of language milestones. *Child Development*, 72(3), 748–767. <https://doi.org/10.1111/1467-8624.00313>
- Tamis-LeMonda, C. S., Bornstein, M. H., Kahana-Kalman, R., Baumwell, L., & Cyphers, L. (1998). Predicting variation in the timing of language milestones in the second year: An events history approach. *Journal of Child Language*, 25(3), 675–700. <https://doi.org/10.1017/s0305000998003572>
- Teasley, S. D. (1995). The role of talk in children's peer collaborations. *Developmental Psychology*, 31(2), 207. <https://doi.org/10.1037/0012-1649.31.2.207>
- Thelen, E. (2000). Grounded in the World: Developmental Origins of the Embodied Mind. *Infancy*, 1(1), 3–28. https://doi.org/10.1207/S15327078IN0101_02
- Thelen, E., & Corbetta, D. (2002). Microdevelopment and dynamic systems: Applications to infant motor development. In *Microdevelopment: Transition processes in development and learning* (pp. 59–79). Cambridge University Press. <https://doi.org/10.1017/cbo9780511489709.003>
- Thelen, E., Kelso, J. S., & Fogel, A. (1987). Self-organizing systems and infant motor development. *Developmental Review*, 7(1), 39–65. [https://doi.org/10.1016/0273-2297\(87\)90004-9](https://doi.org/10.1016/0273-2297(87)90004-9)
- Thelen, E., Schöner, G., Scheier, C., & Smith, L. B. (2001). The dynamics of embodiment: A field theory of infant perseverative reaching. *Behavioral and Brain Sciences*, 24(1), 1–86. <https://doi.org/10.1017/S0140525X01003910>
- Thelen, E., & Smith, L. B. (1994). *A Dynamic Systems Approach to the Development of Cognition and Action*. MIT Press. <http://search.ebscohost.com.proxy-ub.rug.nl/login.aspx?direct=true&db=eric&AN=ED407072&site=ehost-live&scope=site>
- Thelen, E., & Smith, L. B. (2007). Dynamic Systems Theories. In *Handbook of Child Psychology*. John Wiley & Sons, Inc. <https://doi.org/10.1002/9780470147658.chpsy0106>

- Tilsen, S., & Arvaniti, A. (2013). Speech rhythm analysis with decomposition of the amplitude envelope: Characterizing rhythmic patterns within and across languages. *The Journal of the Acoustical Society of America*, 134(1), 628–639. <https://doi.org/10.1121/1.4807565>
- Todman, J. B., & Dugard, P. (2001). *Single-case and small-n experimental designs: A practical guide to randomization tests*. Lawrence Erlbaum Associates Publishers.
<https://doi.org/10.4324/9781410600943>
- Tomasello, M. (2008). *Origins of human communication*. MIT Press.
<https://doi.org/10.7551/mitpress/7551.001.0001>
- Treffner, P., & Peter, M. (2002). Intentional and attentional dynamics of speech-hand coordination. *Human Movement Science*, 21(5–6), 641–697. [https://doi.org/10.1016/S0167-9457\(02\)00178-1](https://doi.org/10.1016/S0167-9457(02)00178-1)
- Treffner, P., Peter, M., & Kleidon, M. (2008). Gestures and Phases: The Dynamics of Speech-Hand Communication. *Ecological Psychology*, 20(1), 32–64.
<https://doi.org/10.1080/10407410701766643>
- Trudeau, J. J., & Dixon, J. A. (2007). Embodiment and abstraction: Actions create relational representations. *Psychonomic Bulletin & Review*, 14(5), 994–1000. <https://doi.org/10.3758/BF03194134>
- Tschacher, W., & Meier, D. (2019). Physiological synchrony in psychotherapy sessions. *Psychotherapy Research*, 30(5), 558–573. <https://doi.org/10.1080/10503307.2019.1612114>
- Tudge, J. (1991, April 19). *Peer collaboration: The case for treating the dyad as the unit of analysis*. Biennial Meeting of the Society for Research in Child Development, Seattle, WA.
- Turvey, M. T. (2007). Action and perception at the level of synergies. *Human Movement Science*, 26(4), 657–697. <https://doi.org/10.1016/j.humov.2007.04.002>
- Van de Langenberg, R., Kingma, I., & Beek, P. J. (2006). Mechanical invariants are implicated in dynamic touch as a function of their salience in the stimulus flow. *Journal of Experimental Psychology: Human Perception and Performance*, 32(5), 1093–1106.
<https://doi.org/10.1037/0096-1523.32.5.1093>
- Van de Pol, J., Volman, M., & Beishuizen, J. (2010). Scaffolding in Teacher–Student Interaction: A Decade of Research. *Educational Psychology Review*, 22(3), 271–296.
<https://doi.org/10.1007/s10648-010-9127-6>
- Van der Maas, H. L., & Molenaar, P. C. (1992). Stages of cognitive development: An application of catastrophe theory. *Psychological Review*, 99(3), 395–417.
<https://doi.org/10.1037/0033-295X.99.3.395>
- Van der Steen, S. (2014). *“How Does It Work?” A Longitudinal Microgenetic Study On The Development Of Young Children's Understanding Of Scientific Concepts* [PhD Thesis]. University of Groningen.
- Van der Steen, S., Steenbeek, H., Van Dijk, M., & Van Geert, P. (2014). A process approach to children's understanding of scientific concepts: A longitudinal case study. *Learning and Individual Differences*, 30, 84–91. <https://doi.org/10.1016/j.lindif.2013.12.004>

References

- Van der Steen, S., Steenbeek, H., & Van Geert, P. (2012). Using the Dynamics of a Person-Context System to Describe Children's Understanding of Air Pressure. In *Current Topics in Children's Learning and Cognition*. InTech. <https://doi.org/10.5772/53935>
- Van der Steen, S., Steenbeek, H. W., Den Hartigh, R. J. R., & Van Geert, P. L. C. (2019). The Link between Microdevelopment and Long-Term Learning Trajectories in Science Learning. *Human Development*, 63(1), 4–32. <https://doi.org/10.1159/000501431>
- Van Dijk, M., & Van Geert, P. (2014). The Nature and Meaning of Intraindividual Variability in Development in the Early Life Span. In M. Diehl, K. Hooker, & M. J. Sliwinski (Eds.), *Handbook of Intraindividual Variability Across the Life Span* (1st ed., pp. 57–78). Routledge. <https://doi.org/10.4324/9780203113066-13>
- Van Dijk, Marijn, & Van Geert, P. (2011). Heuristic techniques for the analysis of variability as a dynamic aspect of change. *Infancia y Aprendizaje: Journal for the Study of Education and Development*, 34(2), 151–167. <https://doi.org/10.1174/021037011795377557>
- Van Dijk, Marijn, Van Geert, P., Korecky-Kröll, K., Maillochon, I., Laaha, S., Dressler, W. U., & Bassano, D. (2013). Dynamic Adaptation in Child-Adult Language Interaction. *Language Learning*, 63(2), 243–270. <https://doi.org/10.1111/lang.12002>
- Van Geert, P. L. C. (2020). Dynamic systems, process and development. *Human Development*, 63(3–4), 153–179. <https://doi.org/10.1159/000503825>
- Van Geert, P., & Steenbeek, H. (2005). The dynamics of scaffolding. *New Ideas in Psychology*, 23(3), 115–128. <https://doi.org/10.1016/j.newideapsych.2006.05.003>
- Van Geert, P., & Van Dijk, M. (2002). Focus on Variability: New Tools to Study Intra-Individual Variability in Developmental Data. *Infant Behavior and Development: An International and Interdisciplinary Journal*, 25(4), 340–374. [https://doi.org/10.1016/S0163-6383\(02\)00140-6](https://doi.org/10.1016/S0163-6383(02)00140-6)
- Van Geert, P. (2003). Dynamic Systems Approaches and Modeling of Developmental Processes. In J. Valsiner & K. J. Conolly (Eds.), *Handbook of Developmental Psychology* (pp. 640–672). SAGE Publications Ltd. <https://doi.org/10.4135/9781848608306.n28>
- Van Geert, P. (2009). Nonlinear Complex Dynamical Systems in Developmental Psychology. In S. J. Guastello, M. Koopmans, & D. Pincus (Eds.), *Chaos and Complexity in Psychology* (pp. 242–281). Cambridge University Press. <https://doi.org/10.1017/cbo9781139058544.009>
- Van Geert, P. (1998). A dynamic systems model of basic developmental mechanisms: Piaget, Vygotsky, and beyond. *Psychological Review*, 105(4), 634–677. <https://doi.org/10.1037/0033-295X.105.4.634-677>
- Van Geert, P. (2011). The contribution of complex dynamic systems to development. *Child Development Perspectives*, 5(4), 273–278. <https://doi.org/10.1111/j.1750-8606.2011.00197.x>
- Van Geert, P., & Fischer, K. W. (2009). Dynamic Systems and the Quest for Individual-Based Models of Change and Development. In J. P. Spencer, M. S. C. Thomas, & J. McClelland (Eds.), *Toward a Unified Theory of Development Connectionism and Dynamic System Theory Re-Consider* (pp. 313–336). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195300598.003.0016>

- Van Geert, P., & Steenbeek, H. (2005). Explaining After by Before: Basic Aspects of a Dynamic Systems Approach to the Study of Development. *Developmental Review*, 23(3–4), 408–442. <https://doi.org/10.1016/j.dr.2005.10.003>
- Van Orden, G. C., Holden, J. G., & Turvey, M. T. (2003). Self-organization of cognitive performance. *Journal of Experimental Psychology: General*, 132(3), 331–350. <https://doi.org/10.1037/0096-3445.132.3.331>
- Van Orden, G. C., Holden, J. G., & Turvey, M. T. (2005). Human Cognition and 1/f Scaling. *Journal of Experimental Psychology: General*, 134(1), 117–123. <https://doi.org/10.1037/0096-3445.134.1.117>
- Van Orden, G. C., Kloos, H., & Wallot, S. (2011). Living in the pink: Intentionality, wellbeing, and complexity. In *Philosophy of complex systems* (pp. 629–672). Elsevier.
- Van Rijn, H., Van Someren, M., & Van der Maas, H. (2003). Modeling developmental transitions on the balance scale task. *Cognitive Science*, 27(2), 227–257. https://doi.org/10.1207/s15516709cog2702_4
- Vingerhoets, G., & Sarrechia, I. (2009). Individual differences in degree of handedness and somesthetic asymmetry predict individual differences in left-right confusion. *Behavioural Brain Research*, 204(1), 212–216. <https://doi.org/10.1016/j.bbr.2009.06.004>
- Vink, R., Wijnants, M. L., Cillessen, A. H. N., & Bosman, A. M. T. (2017). Cooperative learning and interpersonal synchrony. *Nonlinear Dynamics, Psychology, and Life Sciences*, 21(2), 189–215.
- Von Holst, E. (1938). The behavioral physiology of man and animals. In R. Martin (Ed.), *The collected papers of Erich von Holst*.
- Vygotsky, L. S. (1994). The development of academic concepts in school aged children. In R. Van der Veer & J. Valsiner (Eds.), *The Vygotsky Reader*. Wiley-Blackwell.
- Wagman, J. B., Caputo, S. E., & Stoffregen, T. A. (2016). Hierarchical nesting of affordances in a tool use task. *Journal of Experimental Psychology: Human Perception and Performance*, 42(10), 1627–1642. <https://doi.org/10.1037/xhp0000251>
- Wagner, P., Malisz, Z., & Kopp, S. (2014). Gesture and speech in interaction: An overview. *Speech Communication*, 57, 209–232. <https://doi.org/10.1016/j.specom.2013.09.008>
- Wallot, S., O'Brien, B. A., Haussmann, A., Kloos, H., & Lyby, M. S. (2014). The role of reading time complexity and reading speed in text comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(6), 1745. <https://doi.org/10.1037/xlm0000030>
- Walton, A. E., Richardson, M. J., Langland-Hassan, P., & Chemero, A. (2015). Improvisation and the self-organization of multiple musical bodies. *Frontiers in Psychology*, 06. <https://doi.org/10.3389/fpsyg.2015.00313>
- Walton, A. E., Washburn, A., Langland-Hassan, P., Chemero, A., Kloos, H., & Richardson, M. J. (2017). Creating Time: Social Collaboration in Music Improvisation. *Topics in Cognitive Science*, 1(1), 95–119. <https://doi.org/10.1111/tops.12306>

References

- Warren, W. H. (2006). The Dynamics of Perception and Action. *Psychological Review*, 113(2), 358–389. <https://doi.org/10.1037/0033-295X.113.2.358>
- Webber Jr., C. L., & Zbilut, J. P. (2005). Recurrence quantification analysis of nonlinear dynamical systems. In M.A. Riley & G. Van Orden (Eds.), *Tutorials in Contemporary Nonlinear Methods for the Behavioral Sciences*. National Science Foundation. <http://www.nsf.gov/sbe/bcs/pac/nmbs/nmbs.jsp>
- Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education*, 46(1), 71–95. <https://doi.org/10.1016/j.compedu.2005.04.003>
- Weiss, E. J., & Flanders, M. (2004). Muscular and postural synergies of the human hand. *Journal of Neurophysiology*, 92(1), 523–535. <https://doi.org/10.1152/jn.01265.2003>
- West, B. J., Geneston, E. L., & Grigolini, P. (2008). Maximizing information exchange between complex networks. *Physics Reports*, 468(1–3), 1–99. <https://doi.org/10.1016/j.physrep.2008.06.003>
- Whitacre, J. M. (2010). Degeneracy: A link between evolvability, robustness and complexity in biological systems. *Theoretical Biology & Medical Modelling*, 7, 6. <https://doi.org/10.1186/1742-4682-7-6>
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag.
- Wijnants, M. L., Cox, R. F. A., Hasselman, F., Bosman, A. M. T., & Van Orden, G. (2012). A trade-off study revealing nested timescales of constraint. *Frontiers in Physiology*, 3, 116. <https://doi.org/10.3389/fphys.2012.00116>
- Wijnants, M. L., Hasselman, F., Cox, R. F. A., Bosman, A. M. T., & Van Orden, G. (2012). An interaction-dominant perspective on reading fluency and dyslexia. *Annals of Dyslexia*, 62(2), 100–119. <https://doi.org/10.1007/s11881-012-0067-3>
- Wijnants, M. L. (2014). A Review of Theoretical Perspectives in Cognitive Science on the Presence of 1/f Scaling in Coordinated Physiological and Cognitive Processes. *Journal of Nonlinear Dynamics*, 2014, 1–17. <https://doi.org/10.1155/2014/962043>
- Wijnants, M. L., Bosman, A. M. T., Hasselman, F., Cox, R. F. A., & Van Orden, G. C. (2009). 1/f Scaling in Movement Time Changes with Practice in Precision Aiming. *Nonlinear Dynamics, Psychology, and Life Sciences*, 13(1), 79–98.
- Wilke, C. O. (2020a). *Improved Text Rendering Support for “ggplot2”*(0.1.1) [Computer software].
- Wilke, C. O. (2020b). *Streamlined Plot Theme and Plot Annotations for “ggplot2”*(1.1.1) [Computer software].
- Wiltshire, T. J., Steffensen, S. V., & Fiore, S. M. (2019). Multiscale movement coordination dynamics in collaborative team problem solving. *Applied Ergonomics*, 79, 143–151. <https://doi.org/10.1016/j.apergo.2018.07.007>
- Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications. *ArXiv:1308.5499 [Cs]*. <http://arxiv.org/abs/1308.5499>

- Withagen, R., Araújo, D., & De Poel, H. J. (2017). Inviting affordances and agency. *New Ideas in Psychology*, 45, 11–18. <https://doi.org/10.1016/j.newideapsych.2016.12.002>
- Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: A Professional Framework for Multimodality Research. *Proceedings of LREC 2006*, 4.
- Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17(2), 89–100. <https://doi.org/10.1111/j.1469-7610.1976.tb00381.x>
- Wood, S. (2020). *Mixed GAM Computation Vehicle with Automatic Smoothness Estimation* (1.8-33) [Computer software].
- Xavier, J., Gauthier, S., Cohen, D., Zahoui, M., Chetouani, M., Villa, F., Berthoz, A., & Anzalone, S. (2018). Interpersonal Synchronization, Motor Coordination, and Control Are Impaired During a Dynamic Imitation Task in Children With Autism Spectrum Disorder. *Frontiers in Psychology*, 9. <https://doi.org/10.3389/fpsyg.2018.01467>
- Xiong, Y., & Quek, F. (2006). Hand Motion Gesture Frequency Properties and Multimodal Discourse Analysis. *International Journal of Computer Vision*, 69(3), 353–371. <https://doi.org/10.1007/s11263-006-8112-5>
- Xu, T. L., Barbaro, K. de, Abney, D., & Cox, R. (2019). *Finding Structure in Time: Visualizing and Analyzing Behavioral Time Series*. <https://doi.org/10.31234/osf.io/mpz9g>
- Yan, Z., & Fischer, K. (2002). Always under construction: Dynamic variations in adult cognitive microdevelopment. *Human Development*, 45(3), 141–160. <https://doi.org/10.1159/000057070>
- Yeo, A., & Alibali, M. W. (2018). Does visual salience of action affect gesture production? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(5), 826–832. <https://doi.org/10.1037/xlm0000458>
- Yliveronnen, V., Marjanen, P., & Seitamaa-Hakkarainen, P. (2018). Peer Collaboration of Six-Year Olds When Undertaking a Design Task. *Design and Technology Education*, 23(2), n2.
- Young, A. G., Alibali, M. W., & Kalish, C. W. (2019). Causal learning from joint action: Collaboration helps first graders but hinders kindergartners. *Journal of Experimental Child Psychology*, 177, 166–186. <https://doi.org/10.1016/j.jecp.2018.08.001>
- Zbilut, J. P., Giuliani, A., & Webber, C. L. (1998). Detecting deterministic signals in exceptionally noisy environments using cross-recurrence quantification. *Physics Letters A*, 246(1–2), 122–128. [https://doi.org/10.1016/s0375-9601\(98\)00457-5](https://doi.org/10.1016/s0375-9601(98)00457-5)
- Zhang, L. (2019). “Hands-on” plus “inquiry”? Effects of withholding answers coupled with physical manipulations on students’ learning of energy-related science concepts. *Learning and Instruction*, 60, 199–205. <https://doi.org/10.1016/j.learninstruc.2018.01.001>



Nederlandse samenvatting

[Dutch summary]

Nederlandse Samenvatting

Het doel van mijn proefschrift is beter begrijpen hoe cognitieve ontwikkeling gerelateerd is aan hoe kinderen hun *handen bewegen* en hoe ze *spreken* tijdens cognitieve taken. Als volwassenen en kinderen praten dan bewegen ze ook vaak hun handen. Spraak en handbewegingen zijn dus gekoppeld. Een interessante bevinding uit eerdere studies is dat handbewegingen van kinderen de cognitieve ontwikkeling lijken te leiden, en daarmee vooruit lijken te lopen op spraak. Zo gebruiken kinderen bijvoorbeeld hun handen om te verkennen en te exploreren, en laten ze begrip in gebaren zien, nog voordat ze dit begrip onder woorden kunnen brengen.

Eerdere verklaringen voor de leidende rol van de handbewegingen van kinderen in cognitieve ontwikkeling gingen er vanuit dat begrip iets is wat “in het hoofd” plaatsvindt. Een duidelijke metafoor hiervoor is dat de menselijke cognitie vergelijkbaar is met hoe computers werken. Een computer gebruikt input die wordt verwerkt in de processor. Dit zou dan vergelijkbaar zijn met hoe visuele informatie of een rekensom bij mensen door het brein wordt verwerkt. Na een verwerking geeft een computer een uitkomst of een mooie visualisatie. Volgens de computer-metafoor geeft het brein na verwerking ook output, in de vorm van spraak en gebaren of iets anders, waarmee het ‘snappen’ van het probleem wordt getoond. We weten echter dat menselijk gedrag niet zo begrepen kan worden als we computers begrijpen. De uitdaging voor de Psychologie is dan ook: Laten zien hoe menselijk gedrag en begrip van de wereld dan wél tot stand komt.

Het perspectief waar ik in dit proefschrift vanuit ben gegaan, is dat kinderen *complexe dynamische systemen* zijn. Complex betekent dat ze uit veel verschillende onderdelen bestaan op verschillende niveaus, zoals cellen, spieren, bloedvaten, het centrale zenuwstelsel, het skelet, etc. Tevens zijn kinderen zelf ook weer onderdeel van systemen, zoals hun gezin, klas, etc. Al deze verschillende onderdelen op al die verschillende niveaus interacteren met elkaar. Daardoor *organiseren zij zichzelf* tot patronen, zoals lopen, praten, of handbewegingen maken. Dynamisch betekent dat kinderen veranderen. Hierbij zijn de veranderingen op verschillende tijdschalen, bijvoorbeeld de veranderingen van seconde tot seconde, van minuut tot minuut, van uur tot uur, en van jaar tot jaar, aan elkaar gerelateerd. Hoe veranderingen op die verschillende tijdschalen aan elkaar gerelateerd zijn, is te zien aan de variabiliteit van een systeem. Grote veranderingen (ook wel transities genoemd), zoals leren lopen, een verhuizing, of een cognitief inzicht, gaan vaak samen met een grote toename in variabiliteit op allerlei tijdschalen en gebieden.

Verder zijn kinderen voortdurend in interactie met hun omgeving. Deze omgeving beslaat zowel fysieke aspecten, zoals objecten en taakeigenschappen, als sociale aspecten, zoals

Dutch summary

andere mensen, kinderen, of cultuur. Kinderen hun handbewegingen, spraak, maar ook cognitieve ontwikkeling, ontstaan uit deze voortdurende interactie met hun fysieke en sociale omgeving.

Gebaseerd op dit perspectief heb ik daarom onderzocht hoe de handbewegingen, gebaren en spraak van kinderen zich verhouden tot elkaar en tot de fysieke en sociale omgeving, wanneer kinderen zich bezighouden met wetenschap-en-techniek-taken. Op die manier hoopte ik mijn doel, namelijk beter begrijpen hoe cognitieve ontwikkeling gerelateerd is aan hoe kinderen hun *handen bewegen* en hoe ze *spreken* tijdens cognitieve taken, te bereiken. Hierna volgt een samenvatting van de studies die ik heb uitgevoerd. Tenslotte beschrijf ik kort de conclusies en implicaties van de bevindingen uit het proefschrift.

Studie 1 – Asymmetrische afstemming tussen kinderen hun spraak en gebaren (Hoofdstuk 2)

In de eerste studie (Hoofdstuk 2) heb ik onderzocht of de leidende rol van handbewegingen en gebaren (hierna: gebaren) in de cognitieve ontwikkeling van kinderen ook *binnen* een taak zelf aanwezig is. Eerdere studies hebben namelijk aangetoond dat de leidende rol van gebaren met name naar voren komt *tussen* taken die enige tijd na elkaar afgenomen worden. Ik heb onderzocht hoe 12 kinderen samen met een volwassene een wetenschap-en-techniek-taak uitvoerden. Ik codeerde de gebaren en spraak van kinderen en kende begripsniveaus toe aan zowel hun gebaren als spraak.

Vervolgens heb ik de koppeling tussen het begripsniveau in gebaren en het begripsniveau in spraak geanalyseerd. Hiermee kon ik onderzoeken of begrip in één van de modaliteiten (gebaren of spraak) voorliep in de tijd. Een manier om dit voorlopen voor te kunnen stellen is als twee paarden die, achter elkaar, voor een wagen gespannen zijn. Het paard dat vooraan loopt, loopt ook voor in de tijd en is (iets) eerder op de plaats van bestemming. Tevens kon ik op basis van de koppeling tussen gebaren en spraak onderzoeken of het begrip in één modaliteit het begrip in de andere modaliteit sterker aantrok dan andersom. Zo kan een groot en sterk paard een kleiner en zwakker paard makkelijker meetrekken dan andersom, ongeacht positie in de span (vooraan of achteraan).

De resultaten toonden aan dat voor kleuters het begrip in gebaren gemiddeld 18 seconden voorliep op het begrip in spraak. Voor kinderen uit groep 3 was het begrip in gebaren en spraak meer gesynchroniseerd, waarbij begrip in spraak enigszins voorliep op gebaren (gemiddeld met 0.71 seconden). Ik vond dus enkel bij jongere kinderen dat begrip in gebaren voorliep op spraak. Verder vond ik voor alle kinderen, ongeacht leeftijd, dat begrip in spraak het begrip in gebaren meer aantrok dan andersom. Dit betekent dat het begrip in spraak meer bepalend is

voor het begrip in gebaren, dan omgekeerd. Opvallend was dat deze asymmetrie in aantrekking tussen spraak en gebaren meer uitgesproken was voor kinderen die hoger scoorden op een gestandaardiseerde (CITO) taaltest. Bij kinderen die hoger scoorden op een gestandaardiseerde (CITO) rekentest, of op voorgaande wetenschap-en-techniek-taken, was deze asymmetrie minder uitgesproken, en waren spraak en gebaren dus meer in balans. Deze laatste bevinding zou mogelijk kunnen betekenen dat spraak gebaren relatief minder bepaalt bij deze kinderen, en dat kinderen hiervan profiteren tijdens wiskunde- of wetenschap-en-techniek-taken.

Studie 2 – Hoe de moeilijkheid van een taak beïnvloedt hoe gebaren en spraak op elkaar afstemmen (Hoofdstuk 3)

In de tweede studie (Hoofdstuk 3) heb ik onderzocht of de moeilijkheid van een taak de afstemming tussen gebaren en spraak beïnvloedt. Ik koos hierbij voor bachelorstudenten, omdat zij het vol zouden kunnen houden om ergens 1100 keer naar te wijzen en daarbij woorden te zeggen, terwijl dit voor jonge kinderen lastig zou zijn. Elfhonderd was het minimaal benodigde aantal voor het kunnen uitvoeren van de specifieke analyses.

In het experiment werd de moeilijkheid van de taak gemanipuleerd met behulp van een tablet-taak (zie ook Figuur 4, Hoofdstuk 3, p. 106). In de makkelijke conditie moesten de deelnemers staafjes en ringen van dezelfde kleur in een regelmatige volgorde matchen, door naar de locaties (links, midden of rechts) van de staafjes en ringen op een tablet-scherm te wijzen en het woord van de locaties ("links", "midden", of "rechts") uit te spreken. In de moeilijke conditie moesten de deelnemers op eenzelfde wijze staafjes en ringen van dezelfde kleur matchen, maar nu was de volgorde willekeurig in plaats van regelmatig.

Ik heb de afstemming tussen gebaren en spraak op drie manieren geanalyseerd. De eerste maat betrof de *tijd* tussen het wijzen naar een staafje of ring en het uitspreken van het woord. De tweede maat had betrekking op de *semantische overeenkomst* tussen de locatie waarnaar gewezen werd en het woord dat gezegd werd. Zo is het wijzen naar links en het zeggen van "links" een semantische overeenkomst, terwijl het wijzen naar links en het zeggen van "midden" een semantisch verschil is. De derde maat van afstemming had tenslotte betrekking op de organisatie op verschillende tijdschalen van gebaren en spraak. Als de organisatie van een systeem, zoals gebaren en spraak, over veel tijdschalen samenhangt, dan noemen we het systeem 'complex'. Indien de mate van complexiteit van gebaren en spraak overeenkomt, en de tijdschalen waarop beide systemen zich organiseren dus op elkaar lijken, dan interpreteren we dit als een hoge *complexiteits-afstemming*. Als de mate van complexiteit van gebaren en spraak niet overeenkomt, dan zien we dit als een lage complexiteits-afstemming.

Dutch summary

Ik vond dat de moeilijkheid van een taak inderdaad de afstemming tussen gebaren en spraak beïnvloedt. Verder vond ik een groter verschil in tijd tussen het wijzen en uitspreken van het woord, minder semantische overeenkomsten en meer complexiteits-afstemming tussen gebaren en spraak in de moeilijke conditie, in vergelijking met de makkelijke conditie. Deze laatste bevinding van meer complexiteits-afstemming in de moeilijke conditie suggereert dat gebaren en spraak niet enkel afstemden, maar als-het-ware samensmolten tot één geheel in de moeilijke conditie.

Studie 3 – De relatie tussen taakeigenschappen en de variabiliteit van kinderen hun handbewegingen en spraak (Hoofdstuk 4)

Eerdere studies hebben aangetoond dat een overgang van 'oud' naar 'nieuw' begrip een reorganisatie van het systeem is, waarbij het systeem zichzelf dus opnieuw organiseert. Zo'n reorganisatie gaat samen met een toename van variabiliteit. Daarnaast heeft de leidende rol van handbewegingen in de cognitieve ontwikkeling mogelijk te maken met een sterke koppeling van handbewegingen met ruimtelijke taakeigenschappen. Spraak is dan weer sterker gekoppeld met geluid. Als dit zo is, dan zou men een verschil verwachten in de variabiliteit van de handbewegingen als de relevante ruimtelijke taakeigenschappen veranderen, maar geen of minder verschil in de variabiliteit van spraak. In de derde studie (Hoofdstuk 4) heb ik dit onderzocht bij kinderen tussen 4 en 7 jaar oud. Dit is de leeftijd waarop ze vaak beginnen te begrijpen dat niet alleen de massa van de gewichten relevant is voor het oplossen van balanstak-problemen, maar ook de afstand van de gewichten tot het midden van de balansschaal.

Kinderen werden door een experimentleider gevraagd om de uitkomsten van balanstak-problemen te voorspellen en uit te leggen. Kinderen namen hierbij deel aan één van de twee experimenten, met elk twee condities. In het eerste experiment werkten kinderen met een lange balansschaal in de eerste helft van de taak en met een korte balansschaal in de tweede helft (Lang-Kort-conditie), of omgekeerd (Kort-Lang-conditie). Het verschil in lengte van de balansschaal was gerelateerd aan de afstand tot het midden van de balansschaal, wat een relevante dimensie is voor het oplossen van balanstak-problemen. In het tweede experiment werkten kinderen in de eerste helft van de taak met een groot verschil in massa tussen de gewichten en met een klein verschil in massa in de tweede helft van de taak (Groot-Klein-conditie), of omgekeerd (Klein-Groot-conditie). Het verschil in massa tussen de gewichten was gerelateerd aan de gewichts-dimensie, die tevens relevant is voor het oplossen van balanstak-problemen. Ik heb op twee manieren naar variabiliteit gekeken: Diversiteit betrof de verscheidenheid van spraak of handbewegingen, en Complexiteit had betrekking op hoe spraak of handbewegingen over de tijd georganiseerd waren. Voor zowel handbewegingen als

spraak vond ik een verschil in Complexiteit na een verandering in taakeigenschappen (bijvoorbeeld van een lange naar een korte balans, en van een groot naar een klein verschil in massa), behalve in de Lang-Kort-conditie. Ik vond geen verschillen in Diversiteit. Verder was de grootte van de verandering in Complexiteit (en Diversiteit) van handbewegingen en spraak aan elkaar gerelateerd in de Groot-Klein conditie. Samengevat beïnvloeden veranderingen in relevante ruimtelijke taakeigenschappen de flexibiliteit van zowel handbewegingen als spraak, maar niet het aanpassingsvermogen.

Studie 4 – Samenwerkende kinderen stemmen hun spraak, handbewegingen en hoofdbewegingen op elkaar af (Hoofdstuk 5)

In de laatste studie onderzocht ik hoe kinderen hun spraak, handbewegingen en hoofdbewegingen op elkaar afstemmen, wanneer ze samen een reeks van balanstak-problemen oplossen. Ik verzamelde data bij 25 tweetallen tussen de 6 en 10 jaar oud. De data werden verzameld in het Connecticut Science Center in Hartford in de Verenigde Staten.

Bij elk balanstak-probleem werd de kinderen eerst gevraagd om individueel te voorspellen naar welke kant (links, evenwicht, rechts) een balans, met gewichtjes op bepaalde afstanden tot het midden, zou vallen als deze los werd gelaten. Kinderen deden deze voorspelling door op een knop van een gamecontroller te drukken. Als de kinderen een verschillende en/of verkeerde voorspelling gaven, moesten de kinderen met elkaar bespreken wat hun individuele voorspelling was en waarom ze die hadden gedaan. Na deze bespreking over het balanstak-probleem werd de kinderen gevraagd om een tweede, individuele voorspelling te doen over de uitkomst van hetzelfde probleem. Daarna kregen ze de uitkomst te zien en gingen ze door naar het volgende balanstak-probleem.

Ik heb de hand- en hoofdbewegingen van kinderen tijdens besprekingen geregistreerd en hun spraak opgenomen. Belangrijk om hierbij op te merken is dat bewegingen, maar ook spraak, gezien kunnen worden als een lange golfbeweging. Zo registreerde ik voor elke bespreking over een balansprobleem een golfbeweging van spraak, een golfbeweging van handbewegingen, en een golfbeweging van hoofdbewegingen van elk kind. Elk van deze golfbewegingen bestaat weer uit allerlei sub-golven [~] met verschillende frequenties. De frequentie geeft aan hoeveel golven, of oscillaties, er in één seconde passen, uitgedrukt in Hertz (Hz).

Om de afstemming tussen de kinderen te onderzoeken, heb ik onderzocht in hoeverre de frequenties van hun hand- en hoofdbewegingen en spraak gedurende een bespreking overeenkwamen. Hierbij heb ik een onderscheid gemaakt tussen afstemming op drie tijdschalen: Een snelle tijdschaal (8 - 2 Hz); een gemiddelde tijdschaal (2 - 0.5 Hz), en een langzame tijdschaal (0.5 - 0.25 Hz). Uit eerder onderzoek blijkt dat we gemiddeld 2 tot 8

Dutch summary

lettergrepen per seconde uitspreken, dus met 8 tot 2 Hz. Zinnen duren vaak 2 tot 6 secondes, wat correspondeert met 0.5 tot 16 Hz. Verder duren gebaren vaak 0.5 tot 2 secondes, wat correspondeert met 2 tot 0.5 Hz.

Als de frequenties van oscillaties overeenkomen, kan vervolgens ook geanalyseerd worden of de oscillaties in-fase zijn, wat betekent dat ze dezelfde kant op bewegen. Daarnaast kunnen de oscillaties ook anti-fase zijn, wat betekent dat ze de tegengestelde kant op bewegen. Verder heb ik onderzocht of de mate van afstemming en de fase-relatie (in-fase of anti-fase) tussen de hand- en hoofdbewegingen en spraak van kinderen gerelateerd waren aan hoe goed ze samen de balanstak-problemen konden voorspellen.

Ik ontdekte dat kinderen hun spraak meestal op elkaar afstemden op de snelle tijdschaal ($8 - 2 \sim$ per seconde), terwijl zij hun hand- en hoofdbewegingen meestal op elkaar afstemden op de gemiddelde en langzame tijdschaal van ($2 - 0.25 \sim$ per seconde). Deze resultaten zijn vergelijkbaar met bevindingen over de afstemming tussen volwassenen, zij het minder uitgesproken. Verder ontdekte ik dat de spraak van kinderen op de snelle en gemiddelde tijdschaal ($8 - 0.5 \sim$ per seconde) meestal in een in-fase relatie is. Dat wil zeggen dat de oscillaties op die tijdschalen elkaar dus spiegelen, zij het soms met enige vertraging. Tenslotte vond ik geen verband tussen hoe goed kinderen samen de balanstak-problemen konden voorspellen en de mate van afstemming, dan wel fase-relaties, tussen de hand- en hoofdbewegingen en spraak.

Conclusies en implicaties (Hoofdstuk 6)

Een centrale bevinding in alle studies is dat er een duidelijke koppeling en coördinatie te zien tussen kinderen hun gebaren/handbewegingen en spraak, wanneer ze gevraagd worden om over wetenschap-en-techniek-taken uit te leggen. In geen van de studies heb ik een duidelijke leidende rol van handbewegingen in de cognitieve ontwikkeling gezien. De rol van handbewegingen in cognitieve ontwikkeling lijkt eerder genuanceerd, in tegenstelling tot wat de studies die aan het begin van de Nederlandse samenvatting besproken werden deden vermoeden. Een alternatief perspectief is dat zowel spraak, handbewegingen en de koppeling hiertussen zijn "genest" binnen cognitief begrip. Dat betekent dat cognitief begrip zowel (de koppeling tussen) spraak en handbewegingen begrenst als tot stand brengt. Bovendien wordt cognitief begrip andersom ook begrenst en tot stand gebracht door (de koppeling tussen) spraak en handbewegingen. Daarnaast zijn cognitieve inzichten, spraak, handbewegingen en hun koppeling ook ingebed in de sociale en fysieke aspecten van de omgeving. Met andere woorden: Ze hangen allemaal af van de eigenschappen van de taak en hoe daar vragen over worden gesteld.

Dit alles heeft belangrijke consequenties voor de rol van handbewegingen binnen de cognitieve ontwikkeling. Kinderen leren de fysieke structuur van objecten kennen door ze te manipuleren. Dit doen ze met hun handen, omdat deze nou eenmaal het meest geschikt zijn om objecten te exploreren. De fysieke, ruimtelijke eigenschappen van objecten, zoals de vorm, grootte of het materiaal, zijn lastiger tot niet te ontdekken met behulp van iemands eigen spraak. Ik zou daarom willen suggereren dat het vreemd is om te zeggen dat handbewegingen een “leidende rol” hebben bij het exploreren van (nieuwe) hands-on taken en hun eigenschappen. Door middel van onze eigen spraak kunnen we immers niet direct dezelfde informatie over de fysieke eigenschappen achterhalen of communiceren. Als kinderen echter wordt gevraagd om niet (enkel) te exploreren, maar om vragen te beantwoorden of een taak te beschrijven, dan zorgt dit er natuurlijk voor dat ze gaan spreken, maar het beïnvloedt ook hun handbewegingen. Handbewegingen worden als het ware meegetrokken met de spraak, in de vorm van gebaren. Die gebaren zijn dan zowel aangepast aan het ritme van spraak als aan de fysieke eigenschappen van de taak.

Tijdens het ontstaan van cognitieve inzichten worden handbewegingen en spraak van kinderen meer variabel. Hierbij worden de handbewegingen gestructureerd door de fysieke, ruimtelijke eigenschappen van de taak. Voor spraak is dit echter niet het geval. Hoewel het op dat moment zou kunnen lijken alsof het begrip in handbewegingen voorloopt op spraak, reflecteren zowel handbewegingen als spraak de variabiliteit van het systeem tijdens cognitieve inzichten. Met andere woorden: De cognitieve ontwikkeling wordt door zowel handbewegingen als spraak even sterk voortgestuwd. Cognitieve ontwikkeling komt voort uit de interactie van kinderen, inclusief hun handbewegingen en spraak en hun koppeling, met hun fysieke en sociale omgeving.

Hoofdstuk 6 eindigt met een paar bescheiden aanbevelingen voor de onderwijspraktijk. In de eerste plaats is het van belang om ons bewust te zijn van de koppeling tussen handbewegingen en spraak en om kinderen de gelegenheid te geven om op een multimodale manier te communiceren, en dus niet alleen met spraak. Hierbij is het belangrijk om te beseffen dat cognitieve ontwikkeling vaak onduidelijk is. Een toename in vaagheid, en de ambigue handbewegingen en spraak die daarbij horen, is juist een teken dat een cognitief inzicht ontstaat. Dit is iets waar leerkrachten op zouden kunnen letten, gebruik van zouden kunnen maken, en zelfs zouden kunnen uitlokken. Daarnaast heb ik een verschil gevonden tussen kinderen in de kleuterklas en kinderen in groep 3, wat op een potentieel leeftijdsverschil in multimodale communicatie zou kunnen duiden. Dit potentiële leeftijdsverschil zou mogelijk meegenomen kunnen worden in de lespraktijk. Tenslotte beïnvloeden de specifieke eigenschappen van een taak, zoals de fysieke of sociale structuur, de handbewegingen, gebaren en spraak van kinderen en daarbij hun cognitieve ontwikkeling.



Dankwoord

[Acknowledgements]

Dankwoord

Leren, leven, telkens doorgaan
Soms met vallen en weer opstaan
Maar altijd samen, nooit alleen
Veel lieve dank aan iedereen

De afgelopen jaren hebben in het teken gestaan van het, samen met Bram, opbouwen van ons leven (inclusief bruiloft, twee prachtige kinderen, vier verhuizingen, een verbouwing, en vertrek voor een half jaar naar de Verenigde Staten) en het altijd aanwezige werken aan dit proefschrift. Het is een intense, bij tijden chaotische, maar bovenal gelukkige periode geweest. "I wouldn't have it any other way" – Ik zou het allemaal zo nog een keer doen. Gelukkig heb ik, en hebben we, het niet alleen hoeven doen, maar ben ik ondersteund en geholpen op allerlei vlakken. Zonder jullie was het allemaal niet mogelijk geweest en daarvoor wil ik jullie graag bedanken.

Ik begin hierbij met mijn geweldige team van begeleiders: Ralf, Steffie en Paul. Bedankt voor jullie begeleiding, waarmee ik dingen kon onderzoeken en leren begrijpen die ikzelf belangrijk vond. Het is voor jullie niet altijd makkelijk geweest om een promotieproject, slechts voorzien van beperkte middelen en beperkte tijd, in goede banen te kunnen leiden. Ik zou liegen als ik zou zeggen dat ik niet blij zou zijn geweest met meer financiering, maar tegelijkertijd waardeer ik de vrijheid die dit bood, en die jullie me ook gaven. Ik heb altijd met plezier aan m'n proefschrift gewerkt. Verder wil ik benadrukken hoe blij ik ben dat jullie samen mijn begeleidingsteam vormden. Jullie vullen elkaar goed aan, waardoor ik veel van jullie heb geleerd.

Steffie, onze samenwerking begon bijna 10 jaar geleden! In september 2011 startte ik met het schrijven van mijn masterthese, onder jouw begeleiding. Jouw enthousiasme en vertrouwen, tezamen met een heel mooi onderzoeksproject, maakten dat ik al snel besloot om een 30 EC these te schrijven, in plaats van de 11 EC these die ik in eerste instantie van plan was. Tevens was jij het die opperde dat gebaren van kinderen interessant zouden kunnen zijn om te onderzoeken – een bepalend duwtje in de goede richting. Dat masterthese-project vormt op verschillende manieren de basis van dit proefschrift. En niet in de laatste plaats omdat ik toen, onlosmakelijk verbonden aan jouw begeleiding, mijn liefde voor onderzoek ontdekt en ontwikkeld heb. Na mijn masterthese bleven we samenwerken aan wat uiteindelijk een promotieproject werd met jou als copromotor. Heel erg bedankt voor je steun gedurende alle jaren, en voor je aanmoediging, feedback, begrenzingen en structurering (heel belangrijk!), gezelligheid, en dat je altijd voor me klaar stond en me het gevoel gaf dat het allemaal zou lukken. Dit beperkte zich niet enkel tot het proefschrift. Je bent een voorbeeld voor me!

Acknowledgements

Ralf, onze samenwerking begon iets later, in maart 2013. Paul gaf mij de tip om jou te benaderen, omdat ik aan een promotieproject over “embodied cognition” zat te denken. Dit mondde uiteindelijk uit in een eerste, niet gehonoreerde, subsidieaanvraag. Toen ik na dat teleurstellende bericht de hoop op mijn promotieproject had opgegeven, mailde jij dat we vooral door moesten gaan met het zoeken naar financiering. Ik ben je nog altijd dankbaar voor dat bericht en je onvermoeibare doorzettingsvermogen, want anders was dit proefschrift er nooit gekomen. Maar daar stopt mijn dank natuurlijk niet! Gedurende de jaren, en onder jouw begeleiding, raakte ik steeds verder verstrikt in en verslingerd aan *complexity*, *coordination dynamics* en *affordances* – de theoretische bouwstenen van dit proefschrift. Onze eindeloze discussies en overdenkingen (zie ook stelling 7), toegestuurde artikelen en tips voor interessante conferenties hebben mijn visie op ontwikkeling en leren voor een belangrijk deel gevormd. Jouw aansporingen om dingen nog beter te overdenken en op te schrijven hebben een belangrijke bijdrage geleverd aan de diepgang in mijn proefschrift. Als mijn inspanningen dan werden opgevolgd door één van je spaarzame complimenten, dan was dat een duidelijke indicatie dat er daadwerkelijk wat moois op papier stond. Ik heb het als een voorrecht ervaren dat je mijn co/semi-promotor was, bedankt voor alles!

Paul, je hebt mij eens verteld hoeveel impact jouw eerste kennismaking met *complex dynamical systems theory* op jou had, tijdens een lezing op een avond in Groningen (als ik het mij goed herinner). En ik herken dit gevoel van het eerste college dat jij erover gaf. Ik vond het fantastisch en intrigerend dat er een raamwerk was dat zowel het weer, mierenkolonies en menselijk gedrag en ontwikkeling (etc.) kon vangen. Voor mij, iemand die eigenlijk niet wilde kiezen tussen vakken als natuurkunde en wiskunde enerzijds en Psychologie anderzijds (of eigenlijk gewoon nooit wil kiezen tussen interessante dingen), voelde het alsof dingen samenkwamen en op hun plek vielen. Daarnaast sprak de individuele procesbenadering in jouw onderzoek mij erg aan. Ik had het geluk om hier vervolgens verder mee aan de slag te gaan tijdens mijn masterthese-project bij Steffie. Door dit project ontstond bij mij het idee om te promoveren. Na een (voor mij heel spannend!) overleg met jou over promotiemogelijkheden, heb je die plannen vervolgens altijd gefaciliteerd. Bedankt voor het aandragen van mijn eerste echte baan bij Ontwikkelingspsychologie als onderzoeksassistent, voor het contact leggen met Ralf, voor het meewerken aan verschillende subsidieaanvragen en bovenal voor je bereidheid om mijn promotor te zijn, zelfs al reikte de afronding van mijn proefschrift tot ver in je emeritaat. Verder wil ik je bedanken voor je verrijkende steun en feedback tijdens de eindfase van mijn proefschrift, en voor je vertrouwen in Ralf, Steffie en mij tijdens de eerdere fases.

When I think about who or what have been essential to my development as a researcher, nothing has taught me so much in six months as my research visit to the United States in general, and my time at CESPA (UConn) in specific. J, many thanks for your support and warm

welcome at CESPA. Your feedback at different stages of different research projects brought them to a level which would not have been possible without you.

Wim, bedankt dat ik met je samen mocht werken en dat je me op sleeptouw nam in veel van jouw onderzoeksprojecten. Bedankt voor onze discussies over hoe dat nu eigenlijk zit tussen gebaren, spraak en communicatie en hoe we dit zouden kunnen onderzoeken in allerlei ingenieuze studies. Deze discussies hebben een onmiskenbare stempel achtergelaten op mijn proefschrift. Ik heb heel veel van je geleerd, waar ik nog elke dag profijt van heb!

Furthermore, I'd like to thank Martin for the nice conversations, for teaching me essential programming skills, and for beating Bram in racket ball (1 out of 10). Also many thanks to Alex, Adrian, Ashley, Ben, Gaby, Megan, Steve, Claire, Bert, Bruce, Ed, Claudia, Michael and Kerry – I had a great time and learned a lot thanks to all of you. Lastly I'd like to thank Karen Adolph for inviting me to the NYU infant action lab, and Drew Abney for inviting me to the IU cognitive development lab. Both visits, thanks to the lovely and inspiring people I met, have made an indispensable impression on me and on my further development as a researcher.

Dear members of the reading committee, Asli Özyürek, Joanna Rączaszek-Leonardi, and Marijn van Dijk, thank you for your time and effort to read and evaluate my thesis. It's an absolute honor to have you in my thesis' reading committee. I cannot feel any different than being proud that you, as renowned experts in the fields of gestures and speech, child development, complexity, coordination dynamics, and affordances, have approved of my thesis.

Tijdens het schrijven van mijn proefschrift heb ik veel tijd gependend met mijn collega's van (Ontwikkelings-)Psychologie. Ik heb veel gehad aan mijn mede-promovendi van "in het begin", die hun promovendi-jaren inmiddels al lang en breed achter zich hebben gelaten: Mandy, Naomi, Sabine, Astrid en Carla, bedankt! Tevens wil ik de collega's met wie ik een kamer mocht delen bedanken voor de gezelligheid, steun en fijne gesprekken: Frank Assies (in het bijzonder dank voor de gesprekken over verbouwen en kleine kinderen!), Heidi, Bertus, Elske, Niklas en Ymkje-Anna. Marijn, bedankt voor je betrokkenheid en waardering, en voor het "uit de wind houden" als het nodig was. Peter, bedankt voor het meedenken met mijn project en mede mogelijk maken van mijn tijd in de Verenigde Staten. Ruud, bedankt voor je enthousiasme en welkome ideeën, en dat je deur altijd voor me openstond. Verder dank aan al mijn andere collega's - voor jullie betrokkenheid, steun, inzet en het plezier waarmee ik met jullie heb mogen samenwerken: Henderien, Ole, Lucia, Laura, Annemie, Saskia, Niek, Vera, Annelieke, Hanneke, Anne-Margit, Marieke, Ando, Barbera, Jannis, Nicol, en Mark Span (excuus als ik iemand vergeet!). Tevens wil ik José Heesink bedanken voor de kans om aan de slag te gaan als docent bij Psychologie en voor het faciliteren van mijn promotieproject binnen die docent-aanstelling.

Acknowledgements

I'd like to thank the Mindwise-board for the opportunity to be part of the great science communication initiative that is Mindwise. Tassos, Jeremy, Marije, Namkje, Eric, Lara, Yvonne and Stephan, thanks for the fun times, and for teaching me import skills. The experience with finding the right tone to address your audience has definitely helped me with writing my thesis.

Naast mijn collega's bij Psychologie heb ik ook samengewerkt met mensen daarbuiten. Marleen, sinds onze eerste ontmoeting werk ik met veel plezier met je samen. Ik waardeer het zeer hoe je altijd het beste met me voorhebt en je enthousiasme uitspreekt! Onze samenwerking heeft ertoe geleid dat ik een grote interesse heb ontwikkeld voor onderzoek naar communicatie en interactie bij mensen waarvoor dit niet vanzelfsprekend of moeizaam verloopt. Dit laatste heeft geleid tot mijn tijd bij Kentalis. Lieve oud-collega's: Madelon, Peia, Ulrika, Saskia, Nina, Margje, Eline, Esther, Iris, Ilse, Maartje, Rita, Vera, Thea, Trees, Hanneke, Annet, Danielle, Melanie, Marian, Jolanda, en Hans, heel erg bedankt voor de mooie tijd en alles wat ik van jullie geleerd heb, als ook voor jullie steun en betrokkenheid tijdens de laatste loodjes van mijn proefschrift.

Thanks to Riccardo Fusaroli for writing a Marie Curie proposal together with me. This collaboration has been a shaping force behind writing up Chapter 5 of this thesis.

Furthermore, I'd like to thank all my extended colleagues on twitter. You've been very helpful for thinking and reasoning about concepts like a synergy (<https://bit.ly/3hv7daH>) or cognitive understanding (<https://bit.ly/2SKVlHb>). These fruitful discussions have helped shape my thesis' General Introduction and General Discussion. The exchanges have been exceptionally helpful given the world shut down during COVID-times, and many other research-related conversations thereby came to a halt. I look forward to future inspiring online exchanges!

Dear students / Beste studenten: Geanne, Sandra, Emilie, Angela, Rian, Dianne, Pieter-Bas, Wendy, Tessa, Yvonne, Sanne, Vera, Marie, Ursula, Naomi, Mirre, Helen, Wouter, Lydia, Miriam, Hannah, Sven, Lotte, Abigail, Dorien, David, Panagiotis, Erdem, Elise, Joyce, Klodia, Tabitha, Hanne, Esther, Aylinn, Sytske, Nikki, Anne-Laure, Elianne, Hannah, Lysanne, Yaela, Anouk, Jitske, Rick, Suzie, Daniëlle, Levi, Emmy, Nicolien, Eva, Tijmen, Cas, Alexander, Keri en Frederieke. Thanks for your effort and collaboration, this thesis wouldn't have been possible without you! / Bedankt voor je inzet en samenwerking, dit proefschrift zou zonder jou niet mogelijk zijn geweest!

Yannick en Mireille, bedankt dat jullie paranimf willen zijn! Yannick, we have been pursuing our PhD around the same time, and your research, albeit being in sports instead of child development, always felt very close to mine. I loved our conversations about research stuff and beyond, and I will genuinely miss you as you're almost leaving Groningen. Thanks for everything! Mireille, Miek, allerliefste zus, bijzonder dat je als paranimf zo'n directe rol hebt bij het hele promotie-gebeuren. Het is een mooie weerspiegeling van de stabiele, lieve en mooie factor die

jij altijd bent, en ook bent geweest gedurende het schrijven van dit proefschrift. Als we elkaar zien is het altijd goed en kan ik er daarna altijd weer tegenaan. Dankjewel lieve Miek.

Zoals in stelling 9 beschreven, zou het nooit gelukt zijn om een proefschrift te schrijven en twee kinderen op te laten groeien zonder een “village” om ons heen. Ingeborg en Hans, toen de wereld tijdens de eerste corona-golf op slot ging heb ik het erg gewaardeerd dat wij en de kinderen samen op konden trekken! Ook bedankt dat Thom en Lieke bij jullie terecht konden toen wij in die periode verhuisden naar de Knipe. Sander en Aaltsje, de eindsprint die nodig was om dit proefschrift af te schrijven, terwijl tegelijkertijd school en opvang dicht waren tijdens de tweede coronagolf, had ik nooit kunnen volbrengen zonder jullie, bedankt!

Lieve (schoon)familie de Jonge, wat ben ik blij en dankbaar dat ik deel van jullie uitmaak! Bedankt voor alles, waaronder dat jullie altijd interesse hebben getoond in mijn proefschrift, me hebben gestimuleerd, hebben meegeleefd met teleurstellingen en blij momenten, hebben geholpen bij het knutselen van balansschalen en gewichtjes, en mee-knutselende studenten op koffie en appelgebak trakteerden. Maar ik wil jullie, en in het bijzonder Bert en Liesbeth, bovenal bedanken voor de fantastische tijd die Thom en Lieke bij jullie hebben gehad als ik aan het proefschrift moest werken, we verhuisden en klusten, of wanneer ik weer eens met m'n handen in het haar zat omdat ik niet wist wanneer ik anders aan een revisie moest te werken.

Lief gezin waar ik ooit ben opgegroeid, en wat de laatste jaren veranderd is en uitgebreid met nog meer lieve mensen. Mem, heit, bedankt dat jullie me nooit gepusht hebben, maar juist de vrijheid en het vertrouwen hebben gegeven om te doen en ontdekken wat ik zelf leuk en interessant vind. Jullie zijn er waarschijnlijk niet altijd van overtuigd geweest dat die benadering goed zou uitpakken, zoals toen ik met hakken over de sloot het VWO haalde. Maar ik denk dat het mijn eigen motivatie en passie voor het schrijven van o.a. dit proefschrift voor een belangrijk deel bepaald heeft. Heit & Gerda en mem & Dirk, bedankt voor jullie steun en meeleven de afgelopen jaren, en ook voor de fijne tijd die Thom en Lieke bij jullie hadden als ik aan dit boekje liep te zwoegen. Lieve Jelle, Miek & Jitze, lieve broer, zus & sweager, dank voor de interesse en het meeleven, en voor alle lol die Thom en Lieke met jullie hebben gehad. Jelle, jouw opmerking toen je hoorde over mijn eerste baan als onderzoeker draag ik altijd met liefde met mij mee: “Cool, nu ga je nieuwe dingen ontdekken!”.

Liefste Bram, Thom en Lieke, bedankt voor jullie geduld, en voor alle liefde waar jullie mij de afgelopen jaren mee omringd hebben. Bram, bedankt voor de ruimte die je me gegeven hebt om dit proefschrift af te maken, zoals de zes maanden dat je de zorg voor Thom en Lieke op je nam toen we in de VS woonden. Ik waardeer de liefde en humor waarmee je dingen relateert en me weer met beide benen op de grond zet heel erg. Thom en Lieke, mama's boekje is eindelijk af. Ik hou heel veel van jullie allemaal, niets is zo mooi als een leven met jullie. ♥

Acknowledgements

Credits for photos in this dissertation

Cover background: Dan-Cristian Pădureț (<https://unsplash.com/@dancristianp>)

Title page Chapter 1 – p. 5: David Boca (<https://unsplash.com/@davidboca>)

Title page Chapter 2 – p. 26: Jayden Seah (<https://unsplash.com/@jsyz>)

Title page Chapter 3 – p. 60: cottonbro (<https://www.pexels.com/@cottonbro>)

Title page Chapter 4 – p. 94: Monstera (<https://www.pexels.com/@gabby-k>)

Title page Chapter 5 – p. 122: Amina Filkins (<https://www.pexels.com/@amina-filkins>)

Title page Chapter 6 – p. 156: Sharon McCutcheon (<https://www.pexels.com/@mccutcheon>)

Title page Appendices – p. 176: Tatiana Syrikova (<https://www.pexels.com/@tatianasyrikova>)

Title page References – p. 194: cottonbro (<https://www.pexels.com/@cottonbro>)

Title page Nederlandse samenvatting – p. 224: Dirk Faber

Title page Dankwoord – p. 234: Yunus Tuğ (<https://www.pexels.com/@yunustug/>)

Curriculum vitae

Lisette de Jonge-Hoekstra was born in Dordrecht (the Netherlands) on August 21, 1990, and moved to Beetsterzwaag (Friesland, the Netherlands) when she was nine years old. After finishing high school in Drachten, she started with a bachelor in Psychology at the University of Groningen in 2008. Throughout her bachelor she developed a keen interest in child development (and philosophy of science), which made her pursue a master in Developmental Psychology at the University of Groningen in 2011. She finished her master cum laude, and her master thesis was regarded as one of the three best written master theses in Psychology at the University of Groningen (academic year 2011-2012). During her master, Lisette truly discovered her passion for science and research, and thus decided to do a PhD.

After having worked as a research assistant at the department of Developmental Psychology of the University of Groningen in 2012 and 2013, in 2014 Lisette joined as teaching staff while simultaneously working on her PhD project about children's hand movements, speech, and cognitive development. She set-up and taught courses about research methods, science communication, academic skills, and complex dynamical systems in child development, supervised numerous students during their bachelor or master thesis, and was also an editor for [Mindwise](#) (the science communication and outreach platform of the Department of Psychology). Lisette obtained funding for her PhD project from the Scholten-Cordes Fonds and from the Prins Bernhard Cultuurfonds. These last funds enabled her to move to the United States for six months to conduct one of her thesis' studies at the Center for the Ecological Study of Perception and Action (CESPA) at the University of Connecticut. During the last year of her PhD, she worked at Royal Dutch Kentalis, where she was involved in applied research to facilitate the communication of people with (dual) sensory impairments.

Lisette is currently working as a postdoc at the Department of Orthopedagogy & Clinical Educational Science at the University of Groningen. She works on two projects: One about high school teacher's differentiation practices and inclusive education, and one about the development of a Dynamic Assessment instrument for young children with a visual impairment.



Publications

- De Jonge-Hoekstra, L.**, Cox, R. F. A., Van der Steen, S., & Dixon, J. A. (2021). Easier said than done? Task difficulty's influence on temporal alignment, semantic similarity, and complexity matching between gestures and speech. *Cognitive Science*.
<https://doi.org/10.1111/cogs.12989>
- De Jonge-Hoekstra, L.**, Van der Steen, S., & Cox, R. F. A. (2020). Movers and shakers of cognition: Hand movements, speech, task properties, and variability. *Acta Psychologica*, 211. <https://doi.org/10.1016/j.actpsy.2020.103187>
- Pouw, W., **De Jonge-Hoekstra, L.**, Harrison, S. J., Paxton, A., & Dixon, J. A. (2020). Gesture-speech physics in fluent speech and rhythmic upper limb movements. *Annals of the New York Academy of Sciences*, 1491(1), 89-105. <https://doi.org/10.1111/nyas.14532>
- De Jonge-Hoekstra, L.**, Van der Steen, S., Van Geert, P., & Cox, R. F. A. (2016). Asymmetric Dynamic Attunement of Speech and Gestures in the Construction of Children's Understanding. *Frontiers in Psychology*, 7. <https://doi.org/10.3389/fpsyg.2016.00473>
- Cox, R. F. A., Van der Steen, S., Guevara, M., **De Jonge-Hoekstra, L.**, & Van Dijk, M. (2016). Chromatic and Anisotropic Cross-Recurrence Quantification Analysis of Interpersonal Behavior. In *Springer Proceedings in Physics* (pp. 209–225). Springer International Publishing. https://doi.org/10.1007/978-3-319-29922-8_11