University of Groningen

How hand movements and speech tip the balance in cognitive development

de Jonge-Hoekstra, Lisette

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

4

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

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 (Boncoddo 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

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, Boncoddo, 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

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
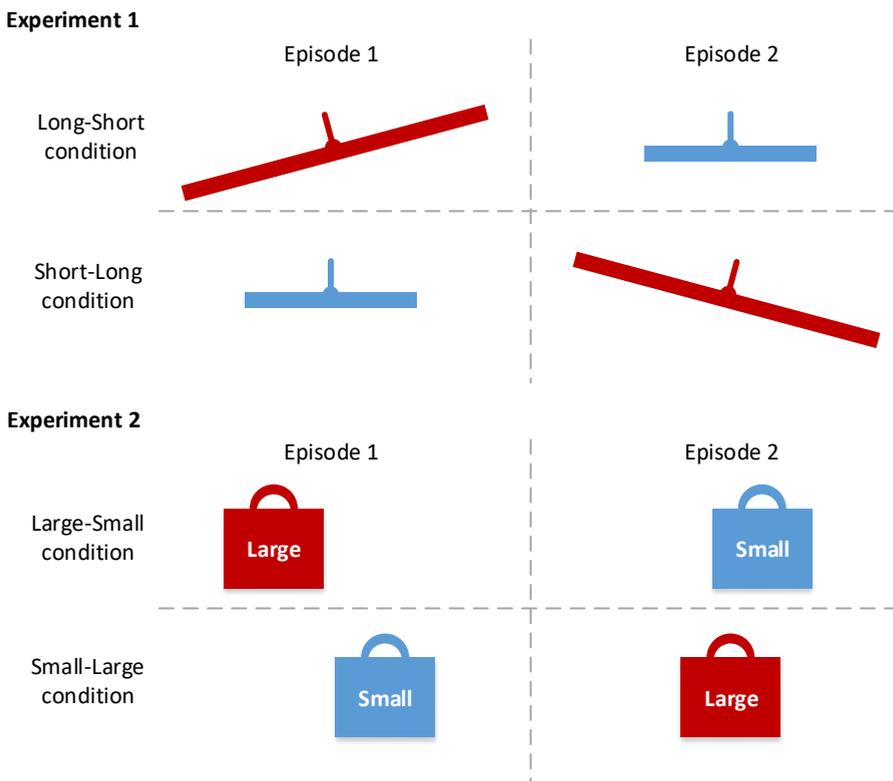

Figure 1. Design of Experiment 1 and Experiment 2.

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 explored[1] 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.

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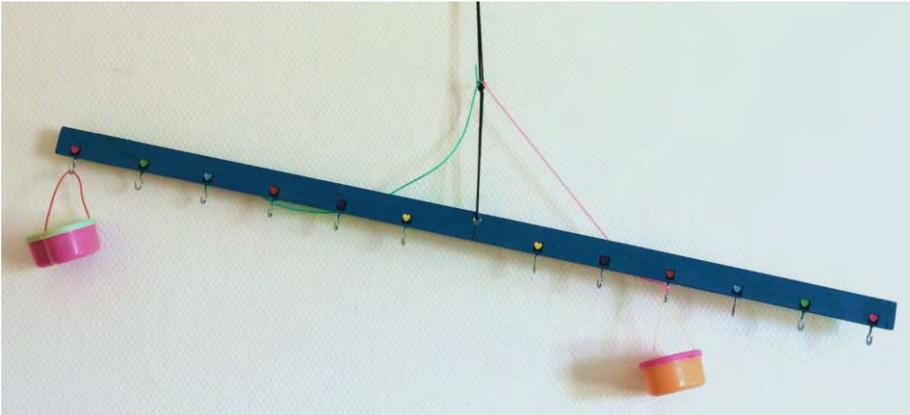
Figure 2. The long balance scale with two weights attached to it.

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

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, g*esturing* 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

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
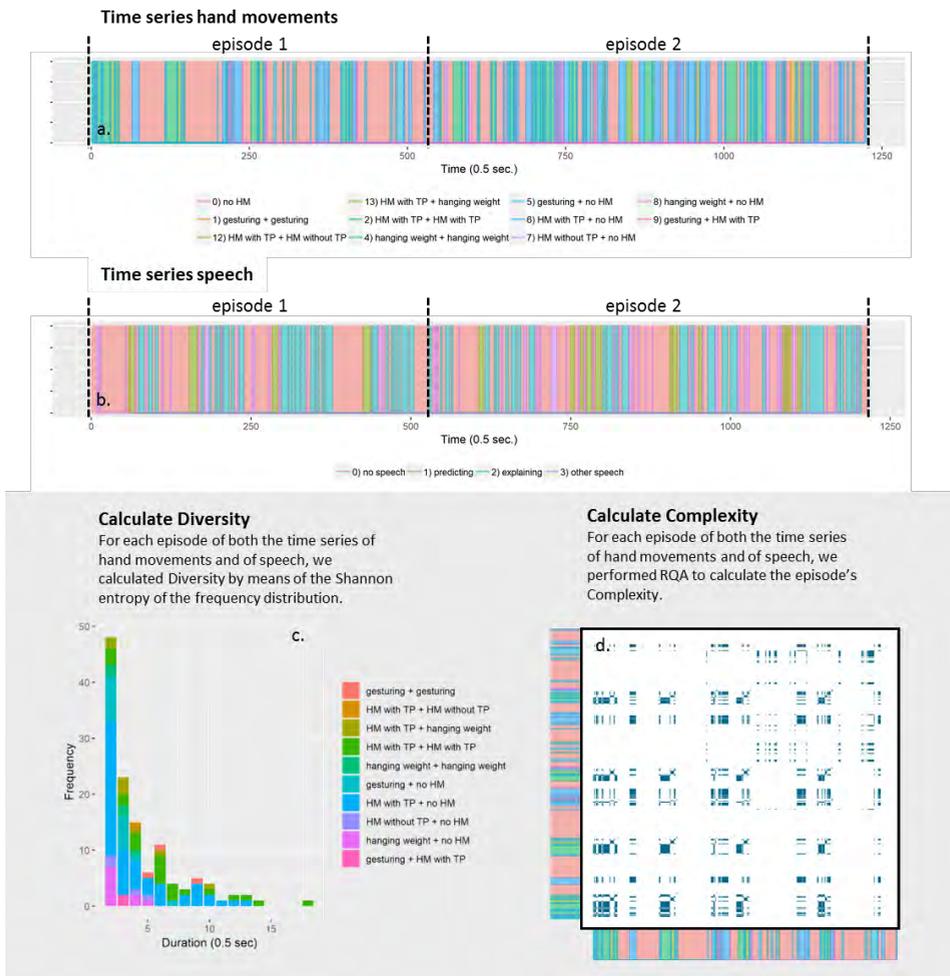
Figure 3. Overview of data and analyses. *Panel a* shows the categories of hand movements (HM = Hand movements, TP = Task Property) over time for one participant in the Long-Short condition, and *panel b* shows the categories of speech over time for the same participant. The dotted line in the middle 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.

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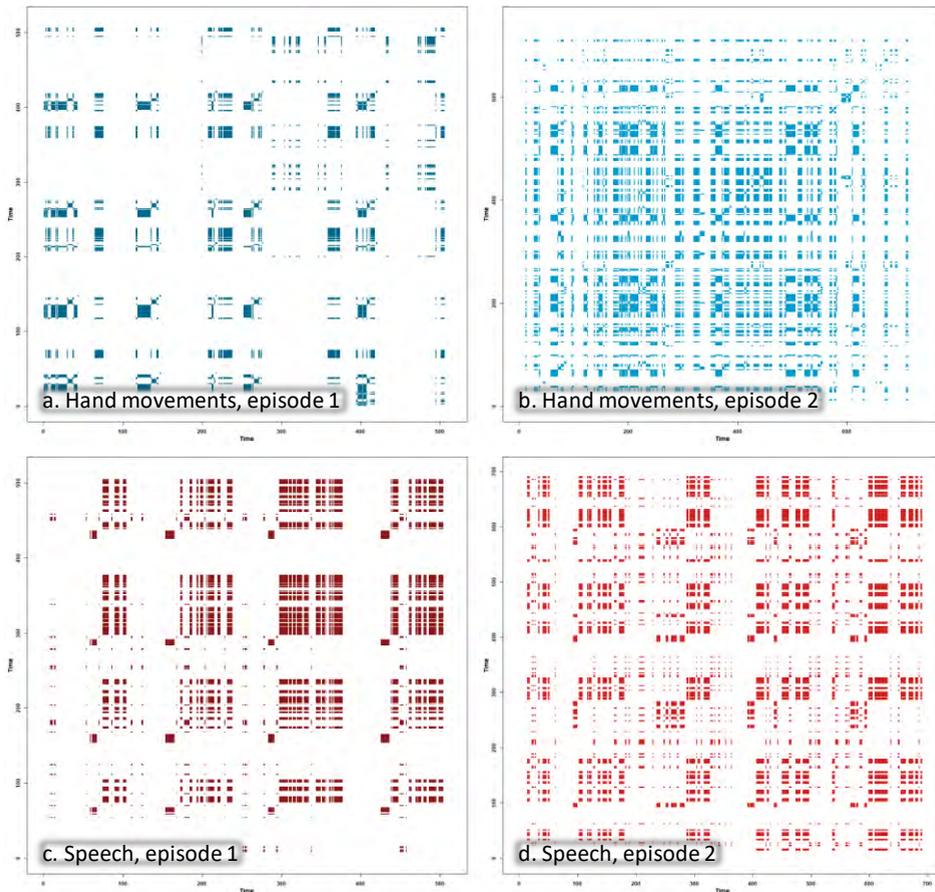
Figure 4. Recurrence plots for the first episode (left) and the second episode (right) for hand movements (a and b) and speech (c and d) of one participant in the Long-Short condition.

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}

= -.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 a 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.

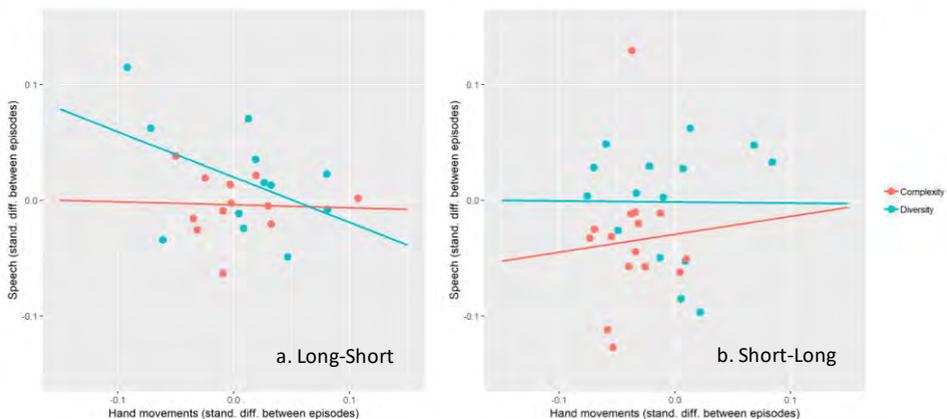*Figure 5.* Relation between hand movements and speech with regard to standardized difference between episodes of Diversity (blue) and Complexity (red) in the Long-Short condition (panel a, $r_{Diversity}$ = -.46, $p_{Diversity}$ = .06, $r_{Complexity}$ = -.04, $p_{Complexity}$ = .47) and the Short-Long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

(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.
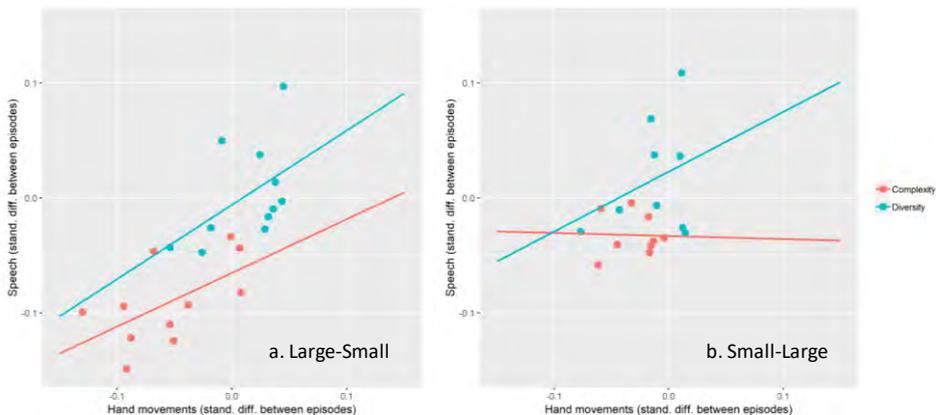

Figure 6. Relation between hand movements and speech with regard to standardized difference between episodes of Diversity (blue) and Complexity (red) in the Large-Small condition (panel a, $r_{Diversity}$ = .49, $p_{Diversity}$ = .05, $r_{Complexity}$ = .58, $p_{Complexity}$ = .04) and the Small-Large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

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, Boncoddo, 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